

**PREDIKSI PENJUALAN PRODUK PADA TOKO BASMALAH DI  
WILAYAH KABUPATEN/KOTA MALANG MENGGUNAKAN  
ARTIFICIAL NEURAL NETWORK DAN EXTREME  
GRADIENT BOOSTING**

**TESIS**

**Oleh:**  
**MUHAMMAD IQBAL AKKAD**  
**NIM. 220605210004**



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

**PREDIKSI PENJUALAN PRODUK PADA TOKO BASMALAH DI  
WILAYAH KABUPATEN/KOTA MALANG MENGGUNAKAN  
ARTIFICIAL NEURAL NETWORK DAN EXTREME  
GRADIENT BOOSTING**

**TESIS**

**Diajukan kepada:**  
**Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang**  
**Untuk Memenuhi Salah Satu Persyaratan Dalam**  
**Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:**  
**MUHAMMAD IQBAL AKKAD**  
**NIM. 220605210004**

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

**PREDIKSI PENJUALAN PRODUK PADA TOKO BASMALAH DI WILAYAH  
KABUPATEN/KOTA MALANG MENGGUNAKAN ARTIFICIAL NEURAL  
NETWORK DAN EXTREME GRADIENT BOOSTING**

**TESIS**

**Oleh:**  
**MUHAMMAD IQBAL AKKAD**  
**NIM. 220605210004**

**Telah Diperiksa dan Disetujui untuk Diuji:**  
**Tanggal: 28 November 2025**

Pembimbing I,



Dr. Ir. Mokhamad Amin Hariyadi, M.T  
NIP 19670018 200501 1 001

Pembimbing II,



Dr. Agung Teguh Wibowo Almais, S.Kom.  
NIP 19860301 2023321 1 016

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



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

**PREDIKSI PENJUALAN PRODUK PADA TOKO BASMALAH DI WILAYAH  
KABUPATEN/KOTA MALANG MENGGUNAKAN ARTIFICIAL NEURAL  
NETWORK DAN EXTREME GRADIENT BOOSTING**

**TESIS**

**Oleh:**  
**MUHAMMAD IQBAL AKKAD**  
**NIM. 220605210004**

Telah Dipertahankan di Depan Dewan Penguji Thesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Magister Komputer (M. Kom)  
Tanggal 03 Desember 2025

Penguji I : Dr. M. Ainul Yaqin, M.Kom  
NIP. 19761013 200604 1 004

Penguji II : Dr. Usman Pagalay, M.Si  
NIP. 19650414 200312 1 001

Pembimbing I : Dr. Ir. Mokhamad Amin Hariyadi, M.T  
NIP 19670018 200501 1 001

Pembimbing II : Dr. Agung Teguh Wibowo Almais, S.Kom, M.T.  
NIP 19860301 2023321 1 016

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



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



## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Muhammad Iqbal Akkad  
NIM : 220605210004  
Program Studi : Magister Informatika  
Fakultas : Sains dan Teknologi  
Judul Thesis : "Prediksi Penjualan Produk pada Toko Basmalah di Wilayah Kabupaten/Kota Malang Menggunakan *Artificial Neural Network* dan *Extreme Gradient Boosting*"

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-benar 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 jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 24 Desember 2025

Menbuat pernyataan,



Muhammad Iqbal Akkad  
NIM. 220605210004

## **HALAMAN MOTTO**

“Ilmu tanpa adab, seperti api tanpa kayu bakar,  
dan adab tanpa ilmu seperti jasad tanpa ruh”

## HALAMAN PERSEMBAHAN

الْحَمْدُ لِلّٰهِ رَبِّ الْعَالَمِينَ

Puji syukur atas kehadirat Allah ﷺ

Shalawat serta salam kepada Rasulullah ﷺ

Dari lubuk hati, penulis mempersembahkan karya tesis ini kepada:

Pengasuh Pondok Pesantren Sidogiri KH. Fuad Noerhasan, Katib Majelis Keluarga Mas d. Nawawy Sadoellah, dan Anggota Majelis Keluarga KH. Bahruddin Thoyib. Almarhum Abi, sosok yang paling semangat saat tahu putranya akan sekolah lagi. Juga untuk Umi yang tidak pernah lelah untuk terus mendoakan putra-putranya.

Keluarga kecil saya, istri tercinta: Hayati. Putri tersayang Aisyah Humaira dan Putra terlucu, Mohammad Hanif. Yang sudah sudi mengorbankan waktu kebersamaan bersama saya. Sehingga agenda liburan pun sering ditunda.

Ketua Pengurus Pusat Ikatan Alumni Santri Sidogiri, Mas Achmad Sa'dulloh, dan Wakil Bidang Pendidikan PP IASS, Pengurus Kopontren Sidogiri, Pengurus LAZ Sidogiri yang telah memberikan kepercayaan kepada saya untuk melanjutkan pendidikan.

Dosen pembimbing Dr. Ir. Mokhamad Amin Hariyadi, M.T dan Dr. Agung Teguh Wibowo Almais, M.T yang telah membimbing penelitian ini dengan memberikan banyak pengarahan dan pengalaman yang berharga.

Segenap sivitas akademika Program Studi Magister Informatika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya.

Seluruh rekan-rekan mahasiswa Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang

Penulis ucapkan “*jazakumullah ahsanal jaza*”. Semoga selalu diridhoi Allah SWT. *Aamiin Ya Rabbal 'Alamiin.*

## KATA PENGANTAR

*Assalamu 'alaikum Warahmatullahi Wabarakatuh*

Syukur alhamdulillah penulis hantarkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik. Shalawat dan salam semoga senantiasa tercurahkan kepada junjungan Nabi Muhammad SAW yang senantiasa menjadi sumber inspirasi dan teladan terbaik begitu juga keluarga, para sahabat dan para pengikutnya seluruh umat Islam.

Penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaiannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Ir. Mokhamad Amin Hariyadi, M.T dan Bapak Dr. Agung Teguh Wibowo Almais, M.T selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
2. Bapak Dr. M. Ainul Yaqin, M. Kom dan Dr. Usman Pagalay, M. Si dan selaku dosen penguji Thesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
3. Keluarga tercinta yang telah banyak memberikan doa dan dukungan kepada penulis secara moril maupun materil hingga Thesis ini dapat terselesaikan.
4. Segenap Civitas Akademika Program Studi Magister Informatika, terutama seluruh Bapak dan Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya
5. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini baik berupa materiil maupun moril yang tidak bisa penulis sebutkan satu persatu tanpa mengurangi rasa hormat dan terimakasih.

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

*Wassalamu 'alaikum Warahmatullahi Wabarakatuh*

Malang, 12 November 2025

Penulis

## DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN .....	iv
PERNYATAAN KEASLIAN TULISAN .....	v
HALAMAN MOTTO .....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xv
ABSTRAK .....	xviii
ABSTRACT .....	xix
مختصر البحث .....	xx
BAB I PENDAHULUAN .....	1
1. 1 LATAR BELAKANG .....	1
1. 2 Pernyataan masalah .....	5
1. 3 Tujuan penelitian .....	6
1. 4 Batasan masalah .....	6
1. 5 Manfaat Penelitian .....	6
BAB II STUDI PUSTAKA .....	7

2. 1	Prediksi Penjualan Produk Pada Toko Ritel .....	7
2. 2	Kerangka Teoritis.....	11
BAB III DESAIN PENELITIAN .....		16
3. 1	Prosedur Penelitian .....	16
3.1. 1	Pengumpulan Data.....	17
3.1. 2	Rekayasa Data.....	18
3.1. 3	Desain sistem.....	19
3.1.3. 1	Ekstraksi Fitur .....	25
3.1. 4	Implementasi Sistem.....	25
3.1. 5	Eksperimen .....	28
3. 2	Research Instrument.....	32
BAB IV PREDIKSI PENJUALAN DENGAN ARTIFICIAL NEURAL NETWORK (ANN) .....		34
4. 1	Desain Jaringan .....	34
4. 2	Implementasi Model ANN.....	35
4. 3	Pelatihan Model ANN.....	36
4.3.1	Pelatihan Model ANN-A.....	37
4.3.2	Pelatihan Model ANN-B .....	38
4.3.3	Pelatihan Model ANN-C .....	40
4.3.4	Pelatihan Model ANN-D .....	41
4.3.5	Pelatihan Model ANN-E .....	43
4.3.6	Pelatihan Model ANN-F .....	44
4.3.7	Pelatihan Model ANN-G .....	46
4.3.8	Pelatihan Model ANN-H.....	48

4.3.9	Pelatihan Model ANN-I .....	49
4.3.10	Pelatihan Model ANN-J .....	51
4.3.11	Pelatihan Model ANN-K.....	52
4.3.12	Pelatihan Model ANN-L .....	54
4.3.13	Pelatihan Model ANN-M .....	56
4.3.14	Pelatihan Model ANN-N.....	58
4.3.15	Pelatihan Model ANN-O.....	60
4.3.16	Pelatihan Model ANN-P .....	62
4.4	Pengujian Model ANN.....	64
4.5	Kesimpulan .....	66
<b>BAB V</b>	<b>PREDIKSI PENJUALAN DENGAN MODEL XGBoost.....</b>	<b>66</b>
5.1	Implementasi Model XGBoost .....	66
5.2	Pelatihan Model XGBoost .....	68
5.2.1	Pelatihan Model XGB-A .....	68
5.2.2	Pelatihan Model XGB-B .....	70
5.2.3	Pelatihan Model XGB-C .....	72
5.2.4	Pelatihan Model XGB-D .....	74
5.2.5	Pelatihan Model XGB-E .....	77
5.2.6	Pelatihan Model XGB-F .....	79
5.2.7	Pelatihan Model XGB-G .....	81
5.2.8	Pelatihan Model XGB-H .....	84
5.2.9	Pelatihan Model XGB-I.....	86
5.2.10	Pelatihan Model XGB-J .....	89
5.2.11	Pelatihan Model XGB-K .....	91

5.2.12	Pelatihan Model XGB-L .....	93
5.2.13	Pelatihan Model XGB-M .....	95
5.2.14	Pelatihan Model XGB-N .....	97
5.2.15	Pelatihan Model XGB-O .....	99
5.2.16	Pelatihan Model XGB-P .....	101
5.3	Pengujian Model XGBoost .....	103
5.4	Kesimpulan .....	104
BAB VI	DISKUSI DAN PEMBAHASAN .....	106
BAB VII	KESIMPULAN .....	122
7.1	Kesimpulan .....	122
7.2	Implikasi dan Rekomendasi .....	123
REFERENSI .....		124

## DAFTAR TABEL

Tabel 2. 1 Performa algoritma dalam memprediksi penjualan produk toko ritel.....	13
Tabel 3. 1 Data penjualan produk outlet Toko Basmalah di Wilayah Kabupaten/Kota Malang Jawa Timur .....	17
Tabel 3. 2 Hasil normalisasi data .....	18
Tabel 3. 3 Detail data penjualan produk outlet Toko Basmalah .....	19
Tabel 3. 4 Ekstraksi fitur .....	25
Tabel 3. 5 Variasi skenario uji coba .....	31
Tabel 3. 6 Variabel penelitian .....	32
Tabel 4. 1 Parameter model ANN .....	36
Tabel 4. 2 Proses Pembelajaran Model ANN-A .....	37
Tabel 4. 3 Proses Pembelajaran Model ANN-B.....	38
Tabel 4. 4 Proses Pembelajaran Model ANN-C.....	40
Tabel 4. 5 Proses Pembelajaran Model ANN-D .....	42
Tabel 4. 6 Proses Pembelajaran Model ANN-E .....	43
Tabel 4. 7 Proses Pembelajaran Model ANN-F (50:50, K-Fold k = 5).....	45
Tabel 4. 8 Proses Pembelajaran Model ANN-G (60:40, K-Fold k = 5).....	46
Tabel 4. 9 Proses Pembelajaran Model ANN-H (70:30, K-Fold k = 5).....	48
Tabel 4. 10 Proses Pembelajaran Model ANN-I (80:20, K-Fold k = 5) .....	49
Tabel 4. 11 Proses Pembelajaran Model ANN-J (90:10, K-Fold k = 5) .....	51
Tabel 4. 12 Proses Pembelajaran Model ANN-K (90:10, Tren Penjualan Turun).....	52
Tabel 4. 13 Proses Pembelajaran Model ANN-L (90:10, Tren Penjualan Stabil).....	54
Tabel 4. 14 Proses Pembelajaran Model ANN-M (90:10, Tren Penjualan Naik) .....	56
Tabel 4. 15 Proses Pembelajaran Model ANN-N (90:10, Kompetitor Rendah) .....	58
Tabel 4. 16 Proses Pembelajaran Model ANN-O (90:10, Kompetitor Sedang) .....	60
Tabel 4. 17 Proses Pembelajaran Model ANN-P (90:10, Kompetitor Tinggi) .....	62
Tabel 4. 18 Hasil Pengujian Model ANN .....	65
Tabel 5. 1 Parameter Model XGBoost .....	67

Tabel 5. 2 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-A .....	68
Tabel 5. 3 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-B.....	70
Tabel 5. 4 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-C.....	72
Tabel 5. 5 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-D .....	75
Tabel 5. 6 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-E .....	77
Tabel 5. 7 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-F (50:50, K-Fold k = 5).....	79
Tabel 5. 8 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-G (60:40, K-Fold k = 5).....	82
Tabel 5. 9 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-H (70:30, K-Fold k = 5).....	84
Tabel 5. 10 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-I (80:20, K-Fold k = 5) .....	87
Tabel 5. 11 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-J (90:10, K-Fold k = 5) .....	89
Tabel 5. 12 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-K (90:10).....	92
Tabel 5. 13 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-L (90:10) .....	93
Tabel 5. 14 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-M (90:10) .....	95
Tabel 5. 15 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-N (90:10).....	97
Tabel 5. 16 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-O (90:10).....	99
Tabel 5. 17 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-P (90:10) .....	101
Tabel 5. 18 Hasil Pengujian Model XGBoost .....	103
Tabel 6. 1 Perbandingan Hasil Pelatihan dan Pengujian Model ANN dan XGBoost .....	112

## DAFTAR GAMBAR

Gambar 2. 1 Kerangka teoritis prediksi penjualan produk Toko Ritel.....	12
Gambar 3. 1 Prosedur penelitian .....	17
Gambar 3. 2 Desain sistem.....	19
Gambar 3. 3 Flowchart proses pelatihan ANN Backpropagation .....	26
Gambar 3. 4 Flowchart proses pelatihan XGBoost .....	28
Gambar 4. 1 Desain arsitektur jaringan model ANN .....	34
Gambar 4. 2 Tingkat Kesalahan Pelatihan AANN-A .....	37
Gambar 4. 3 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-A.....	38
Gambar 4. 4 Tingkat Kesalahan Pelatihan AANN-B.....	39
Gambar 4. 5 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-B	
Gambar 4. 6 Tingkat Kesalahan Pelatihan AANN-C.....	41
Gambar 4. 7 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-C.....	41
Gambar 4. 8 Tingkat Kesalahan Pelatihan AANN-D .....	42
Gambar 4. 9 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-D.....	43
Gambar 4. 10 Tingkat Kesalahan Pelatihan AANN-E .....	44
Gambar 4. 11 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-E .....	44
Gambar 4. 12 Tingkat Kesalahan Pelatihan AANN-F .....	45
Gambar 4. 13 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-F .....	46
Gambar 4. 14 Tingkat Kesalahan Pelatihan AANN-G .....	47
Gambar 4. 15 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-G	
Gambar 4. 16 Tingkat Kesalahan Pelatihan AANN-H .....	48
Gambar 4. 17 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-H.....	49
Gambar 4. 18 Tingkat Kesalahan Pelatihan AANN-I .....	50
Gambar 4. 19 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-I	
Gambar 4. 20 Tingkat Kesalahan Pelatihan AANN-J.....	51
Gambar 4. 21 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-J.....	52
Gambar 4. 22 Tingkat Kesalahan Pelatihan AANN-K .....	53
Gambar 4. 23 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-K.....	54
Gambar 4. 24 Tingkat Kesalahan Pelatihan AANN-L .....	55

Gambar 4. 25 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-L .....	56
Gambar 4. 26 Tingkat Kesalahan Pelatihan AANN-M .....	57
Gambar 4. 27 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-M .....	58
Gambar 4. 28 Tingkat Kesalahan Pelatihan AANN-N .....	59
Gambar 4. 29 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-N .....	60
Gambar 4. 30 Tingkat Kesalahan Pelatihan AANN-O .....	61
Gambar 4. 31 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-O .....	62
Gambar 4. 32 Tingkat Kesalahan Pelatihan AANN-P .....	63
Gambar 4. 33 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-P .....	64
Gambar 5. 1 Nilai Deviasi Data Penjualan Aktual dan Prediksi Pelatihan XGB-A .....	69
Gambar 5. 2 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-A .....	69
Gambar 5. 3 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-B .....	71
Gambar 5. 4 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-B .....	72
Gambar 5. 5 Nilai Deviasi Data Penjualan Aktual dan Prediksi Pelatihan XGB-C .....	73
Gambar 5. 6 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-C .....	74
Gambar 5. 7 Nilai Deviasi Data Penjualan Aktual dan Prediksi Pelatihan XGB-D .....	75
Gambar 5. 8 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-D .....	76
Gambar 5. 9 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-E .....	78
Gambar 5. 10 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-E .....	78
Gambar 5. 11 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-F .....	80
Gambar 5. 12 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-F .....	81
Gambar 5. 13 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-G .....	83
Gambar 5. 14 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-G .....	83
Gambar 5. 15 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-H .....	85

Gambar 5. 16 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-H.....	86
Gambar 5. 17 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-I .....	88
Gambar 5. 18 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-I .....	88
Gambar 5. 19 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-J.....	90
Gambar 5. 20 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-J .....	91
Gambar 5. 21 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-K.....	92
Gambar 5. 22 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-K.....	93
Gambar 5. 23 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-L .....	94
Gambar 5. 24 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-L .....	95
Gambar 5. 25 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-M .....	96
Gambar 5. 26 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-M .....	97
Gambar 5. 27 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-N.....	98
Gambar 5. 28 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-N.....	99
Gambar 5. 29 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-O.....	100
Gambar 5. 30 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-O.....	101
Gambar 5. 31 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-P .....	102
Gambar 5. 32 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-P .....	102
Gambar 6. 1 Waktu Konvergensi Proses Pelatihan Model ANN.....	111
Gambar 6. 2 Perbandingan Hasil Pengujian Model ANN dan XGBoost.....	113
Gambar 6. 3 Nilai Deviasi Proses Pelatihan Model XGBoost .....	114
Gambar 6. 4 Perbandingan Performa Terbaik antar Model ANN dan XGBoost .....	115

## ABSTRAK

Akkad, Muhammad Iqbal 2025. **Prediksi Penjualan Produk pada Toko Basmalah di Wilayah Kabupaten/Kota Malang Menggunakan Artificial Neural Network dan Extreme Gradient Boosting.** Tesis. Program Studi Magister Informatika Fakultas Sains Dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Ir. Mokhamad Amin Hariyadi, M.T.(II) Dr. Agung Teguh Wibowo Almais, M.T

Kata kunci: Prediksi Penjualan, Toko Basmalah, *Artifical Neural Network*, *Backpropagation*, XGBoost

Penelitian ini bertujuan memprediksi penjualan produk di Toko Basmalah wilayah Kabupaten/Kota Malang sebagai upaya meningkatkan efisiensi distribusi dan pengelolaan stok pada jaringan retail syariah yang terus berkembang. Data penjualan harian periode 1 Januari 2023–31 Desember 2024 diolah melalui proses normalisasi, ekstraksi fitur, serta pengujian dua algoritma machine learning, yaitu *Artificial Neural Network* (ANN) *Backpropagation* dan *Extreme Gradient Boosting* (XGBoost), dengan variasi komposisi data latih–uji 50:50 hingga 90:10. Hasil menunjukkan bahwa model ANN-E dengan arsitektur 7-6-3-1 dan komposisi data 90:10 memberikan performa terbaik dengan nilai *Mean Squared Error* (MSE) 34,02 persen, Mean Absolute Error (MAE) sebesar 27,88 persen, serta koefisien determinasi  $R^2$  sebesar 0,304. Sedangkan model XGB-E menghasilkan MSE sebesar 34,98 persen, MAE sebesar 27,63 persen, dan nilai  $R^2$  sebesar 0,329. ANN unggul dalam mengenali pola non-linear dan fluktuatif, sementara XGBoost lebih stabil dan mudah diinterpretasikan. Penelitian ini menegaskan pentingnya penerapan prediksi penjualan berbasis data bagi manajemen Toko Basmalah serta merekomendasikan pengembangan model hybrid ANN–XGBoost dengan variabel eksternal untuk meningkatkan akurasi dan mendukung pengambilan keputusan strategis.

## ABSTRACT

Akkad, Muhammad Iqbal. 2025. **Product Sales Prediction at Toko Basmalah in the Malang Regency/City Area Using Artificial Neural Network and Extreme Gradient Boosting.** Thesis. Master of Informatics Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Ir. Mokhamad Amin Hariyadi, M.T. (II) Dr. Agung Teguh Wibowo Almais, M.T.

Keywords: Sales Prediction, Toko Basmalah, Artificial Neural Network, Backpropagation, XGBoost

This study aims to predict product sales at Basmalah Stores in the Malang Regency and City areas, enhancing distribution efficiency and stock management within an expanding Sharia retail network. Daily sales data from January 1, 2023, to December 31, 2024, were processed through normalization, feature extraction, and the testing of two machine learning algorithms: Artificial Neural Network (ANN) Backpropagation and Extreme Gradient Boosting (XGBoost). The research utilized various training-to-testing data compositions ranging from 50:50 to 90:10. The results indicate that the ANN model with a 7-6-3-1 architecture and a 90:10 data composition delivered the best performance, yielding a Mean Squared Error (MSE) of 34.02%, a Mean Absolute Error (MAE) of 27.88%, and a coefficient of determination  $R^2$  of 0.304. Meanwhile, the XGBoost model produced an MSE of 34.98%, an MAE of 27.63%, and an  $R^2$  value of 0.329. ANN proved superior in recognizing non-linear and fluctuant patterns, whereas XGBoost demonstrated greater stability and ease of interpretation. This study confirms the importance of data-driven sales prediction for managing Basmalah Stores. It recommends the development of a hybrid ANN–XGBoost model that incorporates external variables to enhance accuracy and support informed strategic decision-making.

## مستخلص البحث

أكاد، محمد إقبال. 2025. التنبؤ بمبيعات المنتجات في Toko Basmalah بمنطقة مقاطعة ومدينة مالانج باستخدام Artificial Neural Network و Extreme Gradient Boosting. رسالة الماجستير. قسم المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف الأول: د. محمد أمين هريادي، الماجستير؛ المشرف الثاني: د. أغونج تيغوه وبيورو ألماس، الماجستير.

**الكلمات الرئيسية:** التنبؤ بمبيعات، توكو بسملة، الشبكة العصبية الاصطناعية (Artificial Neural Network – ANN)، Backpropagation، التعزيز التدرجى المتطرف (XGBoost).

تهدف هذه الدراسة إلى التنبؤ بمبيعات المنتجات في توكو بسملة بمنطقة مقاطعة ومدينة مالانج، وذلك من أجل تحسين كفاءة التوزيع وإدارة المخزون ضمن شبكة تجارة التجزئة الشرعية التي تشهد نمواً مستمراً. تمت معالجة بيانات المبيعات اليومية للفترة من 1 يناير 2023 إلى 31 ديسمبر 2024 من خلال عمليات التطبيع (normalization)، واستخراج الخصائص (feature extraction)، واختبار خوارزميتين من خوارزميات التعلم الآلي، وهما الشبكة العصبية الاصطناعية (Artificial Neural Network – ANN) وـXGBoost. ظهرت النتائج أن XGBoost – ANN يحقق أداءً أفضل بنسبة 10:90 مقارنة بـXGBoost، مع استخدام عدة نسب لبيانات التدريب والاختبار تتراوح بين 50:50 و90:10. تم تطبيق XGBoost على بيانات التنبؤ بمبيعات المنتجات في توكو بسملة، حيث بلغت قيمة Mean Squared Error (MSE) 34.02، وقيمة Mean Absolute Error (MAE) 27.88، وقيمة Mean Squared Error (MSE) 34.98، وقيمة Mean Absolute Error (MAE) 27.63، وقيمة R<sup>2</sup> 0.304، وقيمة R<sup>2</sup> 0.329. يتميز نموذج XGBoost بقدرته على التعرف على الأنماط غير الخطية والمتقلبة، في حين يتميز نموذج التعزيز التدرجى المتطرف (Extreme Gradient Boosting – XGBoost) بدرجة أعلى من الاستقرار وسهولة التفسير. تؤكد هذه الدراسة أهمية تطبيق التنبؤ بمبيعات المنتجات على البيانات في إدارة متجر بسملة، كما توصي بتطوير نموذج هجين يجمع بين الشبكة العصبية الاصطناعية (Artificial Neural Network – ANN) وـXGBoost. يهدف تحسين مستوى الدقة ودعم عملية اتخاذ القرار الاستراتيجي.

## **BAB I**

### **PENDAHULUAN**

#### **1. 1 LATAR BELAKANG**

Toko retail merupakan jenis usaha yang menjual produk secara langsung kepada konsumen akhir untuk kebutuhan sehari-hari. Di Indonesia, industri retail mengalami perkembangan pesat seiring dengan pertumbuhan ekonomi dan perubahan gaya hidup masyarakat (Har et al., 2022). Menurut data Asosiasi Pengusaha Ritel Indonesia (APRINDO), sektor retail di Indonesia terus mengalami pertumbuhan dengan kontribusi sekitar 12% terhadap Produk Domestik Bruto (PDB) nasional. Keberadaan toko retail modern seperti minimarket, supermarket, dan hypermarket telah mengubah pola belanja masyarakat dari yang sebelumnya mengandalkan pasar tradisional menjadi lebih mengutamakan kenyamanan dan aksesibilitas yang ditawarkan oleh toko retail modern (Tayibnapis et al., 2018).

Perkembangan pesat ini juga didukung oleh ekspansi jaringan retail besar seperti Indomaret dan Alfamart, yang masing-masing memiliki lebih dari 20.000 gerai yang tersebar di seluruh Indonesia. Selain itu, munculnya retail berbasis syariah seperti Toko Basmalah juga menambah keragaman pilihan bagi masyarakat dalam berbelanja. Dengan persaingan yang semakin ketat, toko retail harus memiliki strategi yang matang untuk mempertahankan eksistensinya (Saepuloh & Noviardiansyah, 2024).

Salah satu jaringan toko retail yang berkembang pesat adalah Toko Basmalah. Toko ini merupakan retail berbasis syariah yang bernaung di bawah PT Sidogiri Mitra Utama salah satu unit usaha milik Pondok Pesantren Sidogiri. Sejak didirikan, Toko Basmalah berhasil memperluas jangkauannya dengan membuka lebih dari 294 cabang yang tersebar di berbagai daerah di Indonesia. Dan dari data Laporan Pertanggung Jawaban tahun 2024 tercatat ada 17 cabang Toko Basmalah di wilayah Kabupaten/Kota Malang, kehadiran Toko Basmalah menjadi alternatif bagi masyarakat dalam memenuhi kebutuhan sehari-hari dengan konsep bisnis yang mengedepankan prinsip syariah. Dibandingkan dengan retail konvensional lainnya, Toko Basmalah memiliki

keunggulan dalam menawarkan produk halal dengan sistem pengelolaan yang sesuai dengan prinsip ekonomi Islam (Lijuan et al., n.d.).

Meskipun mengalami pertumbuhan signifikan, Toko Basmalah menghadapi berbagai tantangan dalam pengelolaannya. Salah satu fenomena utama yang perlu diperhatikan adalah keterbatasan modal yang harus disediakan oleh perusahaan untuk menjaga ketersediaan stok barang di setiap cabang. Pengelolaan modal yang tidak efisien dapat menyebabkan stok barang yang tidak seimbang, baik kelebihan maupun kekurangan, sehingga mempengaruhi tingkat penjualan dan keuntungan toko. Dalam bisnis retail, pengelolaan stok menjadi faktor krusial karena berhubungan langsung dengan tingkat likuiditas dan profitabilitas perusahaan.

Masalah penyediaan modal menjadi tantangan tersendiri bagi Toko Basmalah, karena berbeda dengan toko retail konvensional yang dapat mengakses modal perbankan, Toko Basmalah hanya mengandalkan modal dari anggota atau investor internal. Pengumpulan modal ini dilakukan dalam periode tertentu, yaitu setahun sekali, sehingga fleksibilitas dalam menyesuaikan kebutuhan operasional menjadi terbatas. Selain itu, karena dana yang diperoleh berasal dari anggota, manajemen harus memastikan bahwa modal yang digunakan mampu memberikan return yang tinggi dan berkelanjutan agar para investor tetap tertarik untuk berkontribusi dalam pengembangan usaha. Jika perputaran modal tidak dikelola dengan baik, maka risiko ketidakseimbangan keuangan dapat terjadi, yang berpotensi menghambat pertumbuhan bisnis dan operasional toko.

Selain keterbatasan modal, jalur distribusi juga menjadi tantangan bagi Toko Basmalah. Dengan jumlah cabang yang terus bertambah, efisiensi distribusi barang menjadi faktor krusial dalam menjamin ketersediaan produk di setiap toko. Permasalahan dalam rantai pasok, seperti keterlambatan pengiriman atau kesalahan distribusi barang, dapat berdampak pada kepuasan pelanggan dan penurunan penjualan. Berdasarkan data Kementerian Perdagangan, lebih dari 30% toko retail mengalami kendala dalam rantai distribusi akibat kurangnya infrastruktur logistik yang memadai, terutama di daerah yang jauh dari pusat distribusi utama.

Perubahan daya beli masyarakat juga menjadi faktor yang mempengaruhi kinerja penjualan Toko Basmalah. Dengan adanya fluktuasi ekonomi, tingkat inflasi, serta perubahan kebiasaan belanja masyarakat, toko retail harus mampu beradaptasi agar tetap kompetitif. Menurut

data Badan Pusat Statistik (BPS), daya beli masyarakat Indonesia sempat mengalami penurunan akibat pandemi COVID-19 dan kenaikan harga bahan pokok, yang berdampak langsung pada penjualan sektor retail. Jika tidak ada strategi yang tepat dalam mengantisipasi perubahan daya beli ini, maka risiko penurunan pendapatan dapat terjadi (Roggeveen & Sethuraman, 2020).

Di samping itu, meningkatnya jumlah bisnis serupa juga menjadi tantangan bagi Toko Basmalah. Persaingan dengan minimarket konvensional maupun toko retail berbasis syariah lainnya mengharuskan Toko Basmalah untuk terus melakukan inovasi dalam strategi pemasaran dan operasionalnya agar dapat menarik lebih banyak pelanggan. Dengan semakin banyaknya pemain di industri retail, Toko Basmalah perlu mencari strategi diferensiasi yang dapat membuatnya unggul di pasar.

Untuk mengatasi berbagai tantangan tersebut, penelitian mengenai prediksi penjualan menjadi sangat penting. Dengan menggunakan pendekatan berbasis data, prediksi penjualan dapat membantu manajemen dalam merencanakan kebutuhan stok barang secara lebih akurat, sehingga keterbatasan modal dapat dikelola dengan lebih efisien. Dengan adanya prediksi yang tepat, perusahaan dapat mengoptimalkan anggaran untuk produk-produk yang memiliki tingkat permintaan tinggi. Studi dari Harvard Business Review menunjukkan bahwa penggunaan analisis prediktif dalam retail dapat meningkatkan efisiensi pengelolaan stok hingga 25% dan mengurangi risiko kekurangan barang di rak (Bresnahan, 2024).

Selain itu, prediksi penjualan juga dapat membantu dalam perencanaan distribusi barang yang lebih efektif. Dengan mengetahui pola permintaan berdasarkan data historis, manajemen dapat merancang strategi pengadaan dan distribusi yang lebih optimal, sehingga risiko kekurangan stok atau overstock dapat diminimalisir. Beberapa perusahaan retail besar telah memanfaatkan teknologi big data dan machine learning untuk meningkatkan akurasi prediksi penjualan dan distribusi barang mereka.

Lebih lanjut, pemanfaatan prediksi penjualan dapat membantu Toko Basmalah dalam menghadapi perubahan daya beli masyarakat. Dengan memahami tren penjualan, manajemen dapat menyesuaikan strategi harga, promosi, dan penawaran produk yang sesuai dengan kondisi pasar. Hal ini akan memberikan keunggulan kompetitif bagi toko dalam menghadapi persaingan

dengan bisnis serupa. Sebuah studi dari McKinsey & Company mengungkapkan bahwa bisnis retail yang menggunakan strategi prediksi berbasis data mengalami peningkatan pendapatan hingga 10% dibandingkan dengan bisnis yang masih mengandalkan metode tradisional.

Dalam konteks bisnis, khususnya dalam prediksi penjualan, ayat **QS. Al-Hasyr (59:18)** mengajarkan pentingnya memperhatikan dan merencanakan masa depan dengan penuh kesadaran dan tanggung jawab.

١٨ يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلَا تَنْتَظِرْ نَفْسَنَّ مَا قَدَّمْتُ لِغَدِّ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَيْرٌ بِمَا تَعْمَلُونَ

Artinya: *Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan.*

Sebagaimana ayat tersebut menekankan bahwa setiap individu harus mempersiapkan sesuatu untuk hari esok, begitu pula dalam dunia bisnis, perencanaan yang matang menjadi kunci keberlanjutan usaha. Prediksi penjualan berfungsi sebagai bentuk ikhtiar dalam menghadapi ketidakpastian pasar dengan melakukan analisis berbasis data terhadap tren permintaan, stok barang, serta strategi pemasaran yang efektif. Dengan memahami pola penjualan di masa lalu dan sekarang, manajemen dapat membuat keputusan yang lebih akurat dalam mengalokasikan modal dan sumber daya, sehingga mengurangi risiko kerugian akibat salah perhitungan dalam stok atau investasi. Prinsip ini sejalan dengan ajaran Islam yang mengajak umatnya untuk bersikap visioner, tidak hanya bertumpu pada kondisi saat ini, tetapi juga mempertimbangkan dampak jangka panjang dari setiap keputusan yang diambil.

Nilai yang terkandung dalam ayat ini sangat relevan dengan praktik manajemen modern, terutama dalam pengambilan keputusan berbasis data. Prediksi penjualan dapat dipandang sebagai bentuk ikhtiar rasional untuk mempersiapkan masa depan usaha dengan lebih terukur. Melalui analisis data historis, pola permintaan, dan faktor eksternal yang memengaruhi pasar, perusahaan berupaya mengantisipasi perubahan serta meminimalkan ketidakpastian. Langkah ini bukan hanya bertujuan untuk meningkatkan keuntungan, tetapi juga untuk menjaga keberlanjutan usaha agar tetap stabil dan bertanggung jawab.

Lebih lanjut, penggunaan model prediksi seperti Artificial Neural Network dan XGBoost mencerminkan upaya manusia dalam memanfaatkan ilmu pengetahuan dan teknologi secara optimal, tanpa mengesampingkan nilai nilai spiritual. Dalam perspektif Islam, usaha yang dilakukan secara sungguh sungguh dan terencana merupakan bagian dari amanah, sementara hasil akhirnya tetap diserahkan kepada ketentuan Allah. Dengan demikian, prediksi penjualan tidak dimaknai sebagai upaya menafikan takdir, melainkan sebagai sarana untuk menjalankan prinsip kehati hatian, efisiensi, dan tanggung jawab dalam pengelolaan sumber daya. Oleh karena itu, integrasi antara pendekatan analitis berbasis data dan nilai nilai keislaman memberikan landasan yang kuat bagi pengambilan keputusan bisnis. Perencanaan yang matang, didukung oleh prediksi yang akurat, sejalan dengan ajaran Islam yang mendorong umatnya untuk bersikap visioner, memperhitungkan konsekuensi jangka panjang, serta menjalankan setiap aktivitas ekonomi secara etis dan berkelanjutan.

Dengan demikian, penelitian ini diharapkan dapat memberikan manfaat bagi Toko Basmalah dalam mengembangkan bisnisnya secara lebih strategis. Dengan menggunakan prediksi penjualan sebagai alat bantu pengambilan keputusan, manajemen dapat meningkatkan efisiensi operasional, mengoptimalkan modal, serta meningkatkan kepuasan pelanggan, sehingga keberlanjutan dan pertumbuhan bisnis dapat terus terjaga di masa depan. Implementasi strategi berbasis data tidak hanya membantu dalam perencanaan jangka pendek tetapi juga memberikan dampak positif dalam pengembangan bisnis jangka panjang.

## 1.2 Pernyataan masalah

1. Dengan banyaknya cabang, manajemen harus memastikan distribusi barang lancar agar tidak terjadi kekurangan atau kelebihan stok.
2. Faktor ekonomi seperti inflasi dan tren belanja dapat memengaruhi permintaan produk di toko.
3. Bagaimana pengaruh algoritma ANN dan XG Boost dalam memprediksi penjualan produk outlet Toko Basmalah di Wilayah Kabupaten/Kota Malang Jawa Timur?
4. Penentuan metode terbaik dalam peramalan penjualan produk di Toko Basmalah dilakukan melalui evaluasi nilai Mean Squared Error (MSE) yang dihasilkan oleh masing-masing model.

### **1.3 Tujuan penelitian**

1. Menganalisis pengaruh distribusi barang terhadap ketersediaan stok di seluruh cabang Toko Basmalah di wilayah Kabupaten/Kota Malang, Jawa Timur.
2. Mengkaji dampak faktor ekonomi, seperti inflasi dan tren belanja konsumen, terhadap permintaan produk di Toko Basmalah.
3. Mengevaluasi kinerja algoritma Artificial Neural Network (ANN) dan Extreme Gradient Boosting (XGBoost) dalam memprediksi penjualan produk di outlet Toko Basmalah.
4. Menentukan metode prediksi penjualan yang paling optimal antara ANN dan XGBoost berdasarkan nilai Mean Squared Error (MSE).

### **1.4 Batasan masalah**

1. Penelitian ini difokuskan pada outlet-outlet Toko Basmalah yang berada di wilayah Kabupaten dan Kota Malang, Jawa Timur.
2. Data penjualan outlet Toko Basmalah di Wilayah Kabupaten/Kota Malang Jawa Timur pada periode 01 Januari 2023 – 31 Desember 2024.
3. Prediksi penjualan hanya dilakukan pada produk-produk tertentu yang memiliki volume penjualan tinggi dan data yang lengkap.
4. Algoritma yang dibandingkan terbatas pada Artificial Neural Network (ANN) dan Extreme Gradient Boosting (XGBoost).
5. Evaluasi performa prediksi dibatasi pada dua metrik MSE, MAE, dan  $R^2$ .
6. Faktor eksternal seperti promosi, kompetitor, dan perubahan regulasi tidak dianalisis secara mendalam dalam penelitian ini.

### **1.5 Manfaat Penelitian**

Hasil penelitian dapat dijadikan acuan oleh managemen Toko Basmalah di Wilayah Kabupaten/Kota Malang Jawa Timur dalam dipergunakan dalam menentukan strategi yang tepat saat prediksi penjualan produk menurun untuk meningkat trend penjualan pada Toko tersebut. Bagi praktisi akademisi dapat digunakan sumber referensi untuk penelitian berikutnya.

## **BAB II**

### **STUDI PUSTAKA**

#### **2. 1 Prediksi Penjualan Produk Pada Toko Ritel**

Pada studi yang dilakukan oleh Jain et al., (2015) menyajikan kasus penggunaan data mining untuk peramalan penjualan retail demand dan prediksi penjualan. Secara khusus, algoritma *Extreme Gradient Boosting (XGBoost)* digunakan untuk merancang model prediksi guna memperkirakan kemungkinan penjualan secara akurat untuk sebuah gerai ritel perusahaan ritel Farmasi Eropa besar. Prediksi potensi penjualan didasarkan pada campuran fitur temporal dan ekonomis termasuk data penjualan sebelumnya, promosi toko, pesaing ritel, liburan sekolah dan negara bagian, lokasi dan aksesibilitas toko serta waktu dalam setahun. Proses pembuatan model dipandu oleh penalaran akal sehat dan oleh pengetahuan analitik yang ditemukan selama analisis data dan kesimpulan definitif ditarik. Performa prediktor XGBoost dibandingkan dengan algoritme regresi yang lebih tradisional seperti Linear Regression and Random Forest Regression dengan nilai presisi sekitar 0.10531. Selain itu, temuan dalam penelitian ini menunjukkan bahwa algoritma XGBoost mengungguli pendekatan pemodelan tradisional dalam hal akurasi prediksi, tetapi juga mengungkap pengetahuan baru yang tersembunyi dalam data yang membantu dalam membangun kumpulan fitur yang lebih kuat dan memperkuat model prediksi penjualan.

Wu et al., (2018) melakukan penelitian terhadap pengaruh program diskon black Friday berdasarkan data penjualan. Model multiple regression merupakan Machine learning yang menjadi metode paling tepat untuk memprediksi penjualan selama program promo. Hal ini perlu dilakukan agar toko bisa menyiapkan staf dan mengendalikan krumunan. Studi ini berfokus pada bidang model prediksi untuk mengembangkan algoritma yang akurat dan efisien untuk menganalisis pengeluaran pelanggan di masa lalu dan menampilkan pengeluaran pelanggan di masa depan dengan fitur yang sama. Dalam studi ini, teknik Machine learning yang berbeda seperti regression dan neural network untuk mengembangkan model prediksi diimplementasikan dan perbandingan dilakukan berdasarkan kinerja dan akurasi prediksi. Teknik-teknik ini diimplementasikan menggunakan algoritma yang berbeda dan pada platform yang berbeda untuk menemukan prediksi terbaik. Berdasarkan pengujian terhadap tujuh algoritma, yaitu Linear Regression, MLK

Classifier, model Deep Learning berbasis Keras, Decision Tree, Decision Tree dengan Bagging, serta XGBoost, diperoleh tingkat akurasi dengan rentang nilai RMSE sekitar 24% hingga 60%. Hasil tersebut mengindikasikan bahwa secara umum seluruh algoritma yang digunakan belum mampu mencapai performa prediksi yang optimal. Asumsi dari penulis hal ini berkemungkinan disebabkan oleh perancangan desain pada masing-masing ketujuh algoritma tersebut.

Pada studi yang dilakukan oleh Zhao & Keikhosrokiani, (2022) melakukan penelitian untuk mendapatkan rekomendasi dan prediksi penjualan toko dengan memanfaatkan data historis penjualan untuk mengeksplorasi dan membangun model prediksi dan rekomendasi. Beberapa metode analisis dan algoritma analitik digunakan dalam penelitiannya, seperti Recency, Frequency, and Monetary (RFM) yang digunakan untuk analisis berdasarkan perilaku pengguna. Metode analisis RFM digunakan untuk mensegmentasikan level pelanggan di perusahaan untuk mengidentifikasi pentingnya setiap level. Untuk model rekomendasi produk, teori aturan asosiasi dan algoritma Apriori digunakan untuk menyelesaikan analisis keranjang dan merekomendasikan produk sesuai dengan hasil. Secara keseluruhan, model XGBoost mencapai kinerja dan akurasi yang lebih baik daripada model lainnya dengan nilai akurasi sekitar 78.9%.

Di sisi lain, (Massaro et al., 2018) dibandingkan dengan algoritma Linear Regression dengan nilai akurasi sekitar 96% (Catal et al., 2019) melakukan pengujian algoritma prediktif untuk memprediksi/meramal penjualan. Peneliti melakukan uji coba dengan membandingkan beberapa algoritma seperti *Support Vector Machine (SVM)*, *k-Nearest Neighbor (KNN)*, *Gradient Boosted Trees (GBT)*, *Decision Trees*, dan *Deep Learning*. Hasil pengujian dalam penelitian ini menunjukkan bahwa model Artificial Neural Network (ANN) memiliki kinerja paling unggul dibandingkan model lainnya, dengan tingkat akurasi yang mencapai sekitar 97,4%. Hal ini menunjukkan bahwa model ANN dapat berkinerja dengan baik berdasarkan perbandingan tingkat korelasi antara nilai realita dan nilai prediksi. Selain itu, model ANN terbukti dapat mempelajari model mulai data historis penjualan dan dengan melakukan pra-pemrosesan data.

Studi oleh Sumitra & Sidqi, (2024) memprediksi penjualan produk di toko ritel menggunakan metode ARIMA, Analisis Tren, dan Single Exponential Smoothing. Dari ketiga metode tersebut, Analisis Tren menunjukkan akurasi tertinggi dengan MAPE sebesar 9,91%, sehingga dapat membantu dalam pengambilan keputusan terkait persediaan, terutama untuk produk dengan risiko kerusakan atau kedaluwarsa yang tinggi. Metode Analisis Tren terbukti paling akurat dalam

mendukung keputusan stok. Namun, studi ini hanya membandingkan metode peramalan tradisional tanpa mempertimbangkan pendekatan machine learning atau deep learning, yang berpotensi meningkatkan akurasi prediksi. Selain itu, model yang digunakan bersifat statis dan mungkin kurang responsif terhadap perubahan mendadak dalam pasar atau rantai pasokan. Oleh karena itu, diperlukan pengujian terhadap model lain, seperti XGBoost dan *Artificial Neural Network (ANN)*, yang lebih adaptif dalam menangkap pola kompleks dan perubahan tren. Selain itu, menguji efektivitas metode peramalan pada berbagai jenis produk dapat membantu menentukan model terbaik berdasarkan karakteristik permintaan masing-masing produk, sehingga meningkatkan keakuratan prediksi dan efisiensi manajemen persediaan.

Selanjutnya, pada studi yang dilakukan oleh Dankorpho, (2024) menunjukkan bahwa integrasi algoritma XGBoost secara signifikan meningkatkan akurasi prediksi penjualan di toko ritel, dengan berhasil menurunkan Mean Absolute Error (MAE) sebesar 29,23%. Selain itu, model ini memungkinkan peramalan harian, sehingga membantu mengoptimalkan manajemen inventaris dan strategi pemasaran di berbagai kategori produk. Peralihan dari peramalan bulanan ke harian memberikan efisiensi yang lebih optimal dalam pengelolaan bisnis. Namun, studi ini kemungkinan hanya menguji model pada dataset tertentu, sehingga efektivitasnya belum tentu sama jika diterapkan pada berbagai jenis ritel dengan pola penjualan yang berbeda. Akurasi model XGBoost juga sangat bergantung pada kualitas dan kelengkapan data historis. Jika terdapat data yang hilang atau tidak mencerminkan pola sebenarnya, performa model dapat menurun. Selain itu, XGBoost memerlukan sumber daya komputasi yang lebih besar dibandingkan model yang lebih sederhana seperti regresi linier, sehingga mungkin kurang efisien untuk diterapkan dalam skenario dengan keterbatasan infrastruktur. Oleh karena itu, diperlukan optimalisasi hyperparameter serta eksplorasi teknik seperti pruning atau feature engineering yang lebih efisien guna mengurangi kompleksitas komputasi tanpa mengorbankan akurasi prediksi.

Penelitian oleh Kumar et al., (2023) mengembangkan model prediksi penjualan untuk Big Mart dengan menggunakan algoritma machine learning, khususnya XGBoost. Model ini berhasil mencapai RMSE sebesar 1018,82 dan  $R^2$  sebesar 0,6181, yang menunjukkan peningkatan dalam kemampuan peramalan pendapatan di sektor ritel. Model XGBoost yang disesuaikan dengan Randomized Search CV terbukti mengungguli model yang ada dalam hal akurasi prediksi penjualan. Meskipun XGBoost mencapai  $R^2$  sebesar 0,6181 dan RMSE 1018,82, angka ini

menunjukkan bahwa masih terdapat variasi dalam data yang belum sepenuhnya dijelaskan oleh model, sehingga prediksi yang dihasilkan belum sepenuhnya optimal. Selain itu, model ini diuji pada data penjualan Big Mart, sehingga efektivitasnya mungkin berbeda jika diterapkan pada industri ritel lain dengan pola penjualan yang berbeda. Oleh karena itu, diperlukan eksplorasi model lain, seperti Artificial Neural Network (ANN), yang lebih unggul dalam menangkap pola waktu dan tren jangka panjang. Selain itu, menguji model pada berbagai kategori produk dan format toko dapat membantu mengukur tingkat generalisasi model di berbagai lingkungan bisnis, sehingga meningkatkan ketepatan dan keandalannya dalam peramalan penjualan ritel.

Pada penelitian yang dilakukan oleh Ganguly & Mukherjee, (2024) berfokus pada peningkatan akurasi peramalan penjualan ritel dengan menggunakan model machine learning yang telah dioptimalkan, khususnya model Random Forest yang telah disempurnakan. Model ini berhasil mencapai nilai R-squared sebesar 0,945, yang menunjukkan peningkatan signifikan dalam ketepatan prediksi, terutama untuk dataset yang kompleks dengan tingkat musiman tinggi dan berbagai kategori produk. Selain itu, model ini berhasil menurunkan nilai RMSLE menjadi 1,172, mengungguli model lainnya. Namun, meskipun model ini menunjukkan performa yang sangat baik, masih terdapat keterbatasan dalam menangkap pengaruh faktor eksternal seperti tren ekonomi, kebijakan pemerintah, atau kejadian tak terduga seperti pandemi, yang dapat memengaruhi pola penjualan. Selain itu, ada kemungkinan model Random Forest yang dioptimalkan ini terlalu menyesuaikan diri dengan pola dalam dataset tertentu, sehingga kurang efektif saat diterapkan pada data baru dengan karakteristik yang berbeda. Oleh karena itu, penggunaan model lain seperti XGBoost, LSTM, atau Transformer perlu dipertimbangkan untuk meningkatkan akurasi dan efisiensi prediksi lebih lanjut.

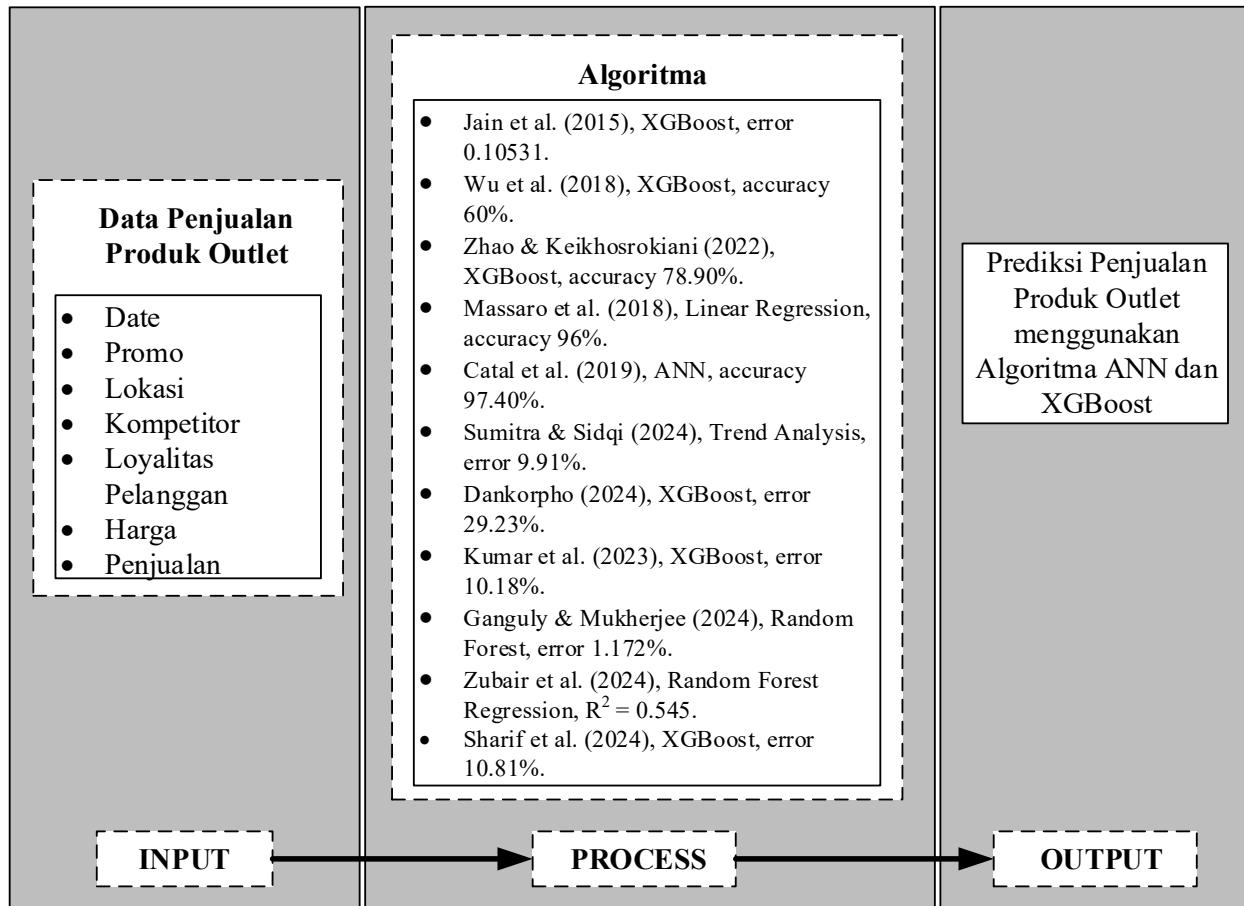
Studi oleh Zubair et al., (2024) membahas penerapan metode machine learning, khususnya regresi linier, Random Forest, dan XGBoost, dalam memprediksi penjualan ritel. Dari ketiga model yang diuji, Random Forest Regression menunjukkan performa terbaik dengan nilai R-squared sebesar 0,545, yang mengindikasikan efektivitasnya dalam meramalkan penjualan produk. Perbandingan berbagai model machine learning ini berkontribusi pada peningkatan akurasi prediksi penjualan ritel. Namun, meskipun Random Forest Regression mencapai nilai R-squared 0,545, angka ini juga menunjukkan bahwa masih ada banyak variabilitas dalam data yang tidak dapat dijelaskan oleh model, sehingga akurasi prediksi perlu ditingkatkan lebih lanjut. Selain

itu, kualitas prediksi sangat bergantung pada data yang digunakan. Jika data yang tersedia tidak lengkap, mengandung nilai yang hilang, atau kurang mencerminkan pola musiman yang kompleks, hasil prediksi dapat menjadi kurang optimal. Oleh karena itu, diperlukan langkah-langkah tambahan, seperti mengoptimalkan hyperparameter lebih lanjut atau mengeksplorasi model lain, seperti Artificial Neural Network (ANN), yang lebih unggul dalam menangkap pola musiman dan tren jangka panjang. Selain itu, pengujian model pada dataset yang lebih besar dan lebih beragam dapat membantu meningkatkan kemampuan generalisasi serta mengurangi bias model.

Penelitian oleh Sharif et al., (2024) memprediksi penjualan produk di toko ritel menggunakan model machine learning, khususnya XGBoost, yang mencapai RMSE sebesar 1081 dan R-squared sebesar 0,59. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam peramalan penjualan untuk perencanaan strategis dan manajemen inventaris. Algoritma XGBoost terbukti memberikan performa prediksi terbaik dibandingkan model lainnya. Namun, dengan  $R^2$  sebesar 0,59 dan RMSE 1081, model masih memiliki margin kesalahan yang cukup besar, sehingga prediksinya belum sepenuhnya akurat untuk semua produk atau skenario ritel. Selain itu, penelitian ini mungkin hanya diuji pada dataset tertentu, sehingga efektivitas model belum dapat dipastikan jika diterapkan pada berbagai sektor ritel dengan pola penjualan yang berbeda. Oleh karena itu, diperlukan pengayaan variabel dengan menambahkan faktor seperti data loyalitas pelanggan untuk meningkatkan ketepatan prediksi. Selain itu, mengaplikasikan model pada berbagai industri ritel dapat membantu menguji generalisasi serta menyesuaikan parameter berdasarkan karakteristik masing-masing sektor, sehingga meningkatkan akurasi dan relevansi prediksi dalam berbagai konteks bisnis.

## 2.2 Kerangka Teori Penelitian

Pada subbab ini diuraikan landasan teoretis yang mendasari proses prediksi penjualan ritel, yang menitikberatkan pada penentuan parameter dan variabel sebagai komponen data input. Pembahasan dilakukan melalui telaah terhadap sejumlah algoritma yang dijadikan rujukan dalam perancangan sistem rekomendasi penjualan ritel beserta metode pembayarannya. Kerangka teoritis terkait prediksi penjualan ritel disajikan pada Gambar 2.1.



**Gambar 2. 1 Kerangka Konseptual Prediksi Penjualan Produk pada Toko Ritel**

Penentuan parameter dalam algoritma machine learning merupakan aspek krusial dalam pengembangan sistem yang efektif. Oleh karena itu, pemilihan algoritma yang akan diterapkan dalam sistem rekomendasi dilakukan melalui kajian pustaka terhadap berbagai penelitian sebelumnya yang membahas prediksi penjualan produk pada sektor ritel. Kajian tersebut dilakukan untuk mengidentifikasi performa algoritma berdasarkan sejumlah indikator, seperti tingkat akurasi, *recall*, *precision*, serta nilai kesalahan (*error*). Evaluasi difokuskan pada penerapan algoritma dalam memprediksi penjualan produk outlet Toko Basmalah yang berlokasi di wilayah Kabupaten/Kota Malang. Setiap algoritma dianalisis secara sistematis, kemudian dilakukan seleksi terhadap dua algoritma dengan kinerja terbaik berdasarkan tingkat akurasi. Kedua algoritma terpilih selanjutnya digunakan sebagai dasar dalam perancangan sistem rekomendasi penjualan ritel dan metode pembayarannya, serta dijadikan acuan utama dalam tahapan penelitian sebagaimana dirangkum pada Tabel 2.1.

**Tabel 2. 1 Jenis algoritma dalam memprediksi penjualan produk toko ritel**

No.	Nama Penulis	Tahun	Model	Akurasi	Error
1	Jain et al.	2015	XGBoost	-	0.10531
2	Wu et al.	2018	XGBoost	60%	-
3	Zhao & Keikhosrokiani	2022	XGBoost	78.90%	-
4	Massaro et al.	2018	Linear Regression	96%	-
5	Catal et al.	2019	ANN	97.40%	-
6	Sumitra & Sidqi	2024	Trend Analysis	-	9.91%
7	Dankorpho	2024	XGBoost	-	29.23%
8	Kumar et al.	2023	XGBoost	-	10.18%
9	Ganguly & Mukherjee	2024	Random Forest	-	1.172%
10	Zubair et al.	2024	Random Forest Regression	$R^2 = 0.545$	-
11	Sharif et al.	2024	XGBoost	-	10.81%

Tabel 2.1 menyajikan ringkasan hasil kinerja prediksi penjualan produk pada sektor ritel yang telah dianalisis oleh berbagai peneliti dengan memanfaatkan beragam algoritma machine learning. Setiap penelitian menggunakan jenis data yang berbeda, seperti data numerik, teks, dan informasi lokasi, yang dijadikan sebagai fitur atau variabel utama dalam proses peramalan penjualan produk pada toko ritel.

Banyak penelitian telah dilakukan untuk mengembangkan model prediksi penjualan di sektor ritel, baik menggunakan metode machine learning maupun pendekatan statistik tradisional. Studi oleh Jain et al. (2015) menunjukkan bahwa algoritma Extreme Gradient Boosting (XGBoost) memiliki keunggulan dibandingkan metode tradisional seperti Linear Regression dan Random Forest Regression dalam memprediksi penjualan. Keberhasilannya terutama didorong oleh pemanfaatan fitur temporal dan ekonomis yang relevan. Sementara itu, Wu et al. (2018) meneliti efektivitas berbagai algoritma dalam meramalkan penjualan selama program diskon Black Friday.

Hasilnya menunjukkan bahwa tidak ada model yang secara konsisten unggul, dengan tingkat akurasi berkisar antara 24% hingga 60% RMSE. Hal ini menegaskan bahwa desain dan parameterisasi model sangat berpengaruh terhadap kualitas prediksi.

Pendekatan lain juga telah dieksplorasi untuk memahami pola penjualan dengan lebih baik. Zhao & Keikhosrokiani (2022) menggunakan analisis Recency, Frequency, and Monetary (RFM) serta teori aturan asosiasi untuk merekomendasikan produk berdasarkan data historis. Dari berbagai model yang diuji, XGBoost terbukti memberikan hasil paling akurat dengan tingkat keberhasilan 78,9%. Sementara itu, penelitian oleh Massaro et al. (2018) dan Catal et al. (2019) membandingkan berbagai algoritma seperti Support Vector Machine (SVM), k-Nearest Neighbor (KNN), dan Decision Trees. Mereka menemukan bahwa Artificial Neural Network (ANN) mampu memberikan akurasi tertinggi, sekitar 97,4%, berkat kemampuannya mengenali pola kompleks dalam data. Namun, pendekatan peramalan tradisional seperti yang diterapkan oleh Sumitra & Sidqi (2024), termasuk ARIMA dan Analisis Tren, meskipun cukup efektif, masih memiliki keterbatasan dalam menangkap dinamika pasar yang berubah-ubah.

Penelitian terbaru semakin menegaskan keunggulan model XGBoost dalam meningkatkan keakuratan prediksi penjualan. Studi oleh Dankorpho (2024) menunjukkan bahwa penerapan XGBoost mampu menurunkan tingkat kesalahan (MAE) hingga 29,23%, memungkinkan prediksi harian yang lebih presisi untuk mendukung manajemen inventaris. Sementara itu, penelitian oleh Kumar et al. (2023) dan Ganguly & Mukherjee (2024) menunjukkan bahwa model machine learning seperti XGBoost dan Random Forest yang telah dioptimalkan dapat meningkatkan keakuratan peramalan secara signifikan. Meski begitu, penelitian ini juga menyoroti bahwa kualitas data historis serta teknik feature engineering yang tepat sangat menentukan efektivitas model dalam menangkap variasi pola penjualan.

Secara faktual Toko Basmalah mengalami perkembangan yang cukup signifikan. Namun, juga menghadapi tantangan dalam pengelolaannya. Salah satu tantangan terbesar adalah keterbatasan modal, karena sumber pendanaan hanya berasal dari anggota, tanpa akses ke perbankan. Hal ini membuat fleksibilitas keuangan terbatas, sehingga stok barang bisa tidak seimbang dan terlalu banyak atau terlalu sedikit yang akhirnya mempengaruhi penjualan dan keuntungan. Selain itu, dengan semakin banyaknya cabang, pengelolaan distribusi barang menjadi semakin kompleks. Jika pengiriman terlambat atau ada kesalahan dalam distribusi, pelanggan bisa

kecewa, dan penjualan pun menurun. Apalagi, daya beli masyarakat yang terus berubah akibat faktor ekonomi seperti inflasi dan kenaikan harga bahan pokok juga berdampak pada performa toko. Di tengah persaingan dengan minimarket konvensional dan toko berbasis syariah lainnya, Toko Basmalah perlu terus berinovasi agar tetap menarik bagi pelanggan. Dengan strategi yang tepat dalam mengelola modal, distribusi, dan pemasaran, toko ini dapat terus tumbuh dan memberikan manfaat lebih besar bagi para anggotanya.

Dalam menghadapi tantangan tersebut, penerapan teknologi kecerdasan buatan menjadi solusi yang menjanjikan, terutama dalam memprediksi penjualan dan mengelola persediaan secara lebih efektif. Artificial Neural Network (ANN) dan XGBoost adalah dua model yang dapat membantu bisnis ritel, termasuk Toko Basmalah, dalam mengoptimalkan pengelolaan stok dan distribusi. ANN memiliki kemampuan luar biasa dalam mengenali pola yang kompleks serta mempelajari hubungan non-linear antar variabel, menjadikannya unggul dalam memahami tren jangka panjang. Di sisi lain, XGBoost menawarkan efisiensi tinggi dalam pemrosesan data serta interpretabilitas yang baik, memungkinkan perusahaan untuk mengolah data dalam jumlah besar dan meningkatkan akurasi prediksi dengan teknik boosting. Dengan melakukan analisis mendalam dari kedua model tersebut secara strategis, Toko Basmalah dapat mengatasi ketidakseimbangan stok, meningkatkan efisiensi operasional, serta mengambil keputusan berbasis data yang lebih akurat, sehingga mampu beradaptasi lebih baik terhadap dinamika pasar yang terus berubah.

Selain itu, penerapan ANN dan XGBoost tidak hanya berfokus pada peningkatan akurasi prediksi penjualan, tetapi juga berperan penting dalam membangun sistem pengambilan keputusan yang adaptif dan berkelanjutan. Integrasi hasil prediksi ke dalam proses perencanaan operasional memungkinkan manajemen Toko Basmalah untuk merespons perubahan permintaan secara lebih cepat, menyesuaikan strategi pengadaan, serta mengurangi risiko kerugian akibat overstock maupun stockout. Dengan dukungan evaluasi model secara berkala dan pemutakhiran data yang berkelanjutan, pemanfaatan kecerdasan buatan ini dapat menjadi fondasi analitik jangka panjang yang membantu toko menjaga stabilitas operasional, meningkatkan daya saing, dan menghadapi ketidakpastian pasar ritel secara lebih sistematis dan terukur.

## **BAB III**

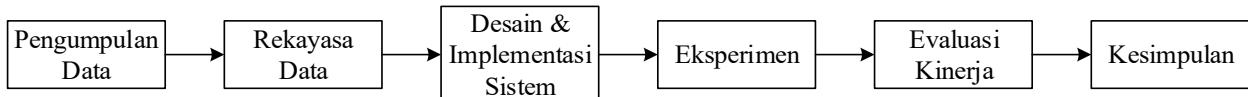
### **DESAIN PENELITIAN**

Desain penelitian merupakan kerangka kerja terstruktur yang memberikan gambaran menyeluruh mengenai alur dan tahapan penelitian, mulai dari perumusan masalah hingga proses analisis metode yang digunakan. Asnahabi (2019) mendefinisikan desain penelitian sebagai suatu rencana yang disusun oleh peneliti sebagai pedoman dalam melaksanakan penelitian. Sementara itu, Sileyew (2019) memandang desain penelitian sebagai bentuk persiapan dalam mengkaji suatu kelompok atau fenomena yang terjadi dalam konteks lingkungan alaminya. Di samping itu, desain penelitian berfungsi sebagai bentuk pertanggungjawaban peneliti terhadap setiap keputusan dan tahapan yang ditempuh selama pelaksanaan penelitian (Margono, 2005). Penelitian ini menggunakan pendekatan kuantitatif, mengingat proses analisis data dilakukan melalui metode statistik, khususnya dengan memanfaatkan persamaan Mean Squared Error (MSE) dalam pengolahan data numerik (Lim et al., 2016). Pendekatan analisis isi kuantitatif digunakan dalam penelitian ilmiah untuk mengkaji dan mengelompokkan karakteristik tertentu yang terdapat dalam suatu konten. Melalui pendekatan ini, pola atau bentuk komunikasi yang muncul dapat diidentifikasi secara sistematis melalui prosedur analisis yang bersifat objektif, memiliki validitas dan reliabilitas, serta memungkinkan untuk direplikasi pada penelitian selanjutnya (Rourke et al., 2001).

#### **3. 1      Tahapan Penelitian**

Tahapan penelitian merujuk pada serangkaian tahapan yang dirancang oleh peneliti untuk memperoleh data serta menjawab permasalahan yang dikaji. Pada penelitian ini, pembahasan difokuskan pada metode pengumpulan data, tahapan pengembangan sistem, serta proses eksperimen yang dilakukan. Penelitian eksperimen memanfaatkan data penelitian sebagai dasar analisis untuk menghasilkan temuan yang dapat dibuktikan melalui pelaksanaan eksperimen maupun observasi.

Penelitian ini dikategorikan sebagai penelitian eksperimental dengan menggunakan data penjualan produk outlet Toko Basmalah yang berlokasi di wilayah Kabupaten/Kota Malang. Alur pelaksanaan prosedur penelitian tersebut disajikan dalam bentuk diagram pada Gambar 3.1:

**Gambar 3. 1 Tahapan penelitian**

(Sumber: diolah peneliti)

### 3.1. 1 Pengambilan Data

Hampir setiap jenis penelitian bergantung pada pengumpulan data. Data dengan kualitas yang rendah berisiko menurunkan tingkat keabsahan penelitian, yang pada akhirnya dapat menyebabkan hasil penelitian tidak mencapai tingkat akurasi yang diharapkan (Tan, 2006). Data yang diperlukan untuk penelitian ini dikumpulkan melalui prosedur merupakan data primer yang di ambil dari divisi marketing PT. Sidogiri Pandu Utama. Data yang digunakan dalam penelitian ini diperoleh dengan cara mengunduh dataset time series sebagai sumber data utama penjualan produk outlet Toko Basmalah di Wilayah Kabupaten/Kota Malang dari divisi marketing PT. Sidogiri Pandu Utama pada periode 01 Januari 2023 – 31 Desember 2024. Pengumpulan data dilakukan pada tanggal 03 Januari 2025 dengan jumlah data sekitar 11.266 data pada setiap atribut. Atribut yang diperoleh di antaranya Tanggal, Promo, Lokasi, Kompetitor, Loyalitas Pelanggan (LP), Harga, dan Penjualan. Data tersebut tersaji di Tabel 3.1:

**Tabel 3. 1 Data penjualan produk outlet Toko Basmalah di Wilayah Kabupaten/Kota Malang Jawa Timur**

Date	Promo	Lokasi	Kompetitor	LP	Harga	Penjualan
1/1/2023	1	BASMALAH CABANG JATIKERTO	3	0	157000	76
1/1/2023	1	BASMALAH CABANG PAKISAJI	2	0	413000	130
1/1/2023	1	BASMALAH CABANG KEPANJEN MALANG	4	0	349000	73
1/1/2023	1	BASMALAH CABANG LOWOKWARU	3	0	295000	154
1/1/2023	1	BASMALAH CABANG PAKISAJI	1	0	472000	6
1/1/2023	0	BASMALAH CABANG DAMPIT MALANG	5	0	358000	167
1/1/2023	0	BASMALAH CABANG PAGELARAN	3	1	492000	76
1/1/2023	1	BASMALAH CABANG BUMIAYU	2	1	437000	23
1/1/2023	1	BASMALAH CABANG BUMIAYU	4	1	326000	219

...	...	...	...	...	...	...	...
12/31/20 24	0	BASMALAH CABANG TALANGSUKO	5	0	355000	0	
12/31/20 24	1	BASMALAH CABANG WAGIR MALANG	3	0	304000	156	

### 3.1. 2 Rekayasa Data

Data harus diproses sebelum melanjutkan ke langkah selanjutnya. Preprocessing dilakukan untuk meningkatkan proses atau teknik pengumpulan data. Metode normalisasi *min–max* bekerja dengan menerapkan transformasi linier sehingga nilai data awal dapat disesuaikan ke dalam rentang tertentu. Ini menggunakan harga terendah, harga tertinggi, dan nilai terendah dan tertinggi. Penerapan skala *min–max* dilakukan untuk meningkatkan ketepatan hasil prediksi sekaligus mengurangi risiko terjadinya overfitting selama proses pemodelan. Adapun hasil dari proses normalisasi data tersebut disajikan pada Tabel 3.2:

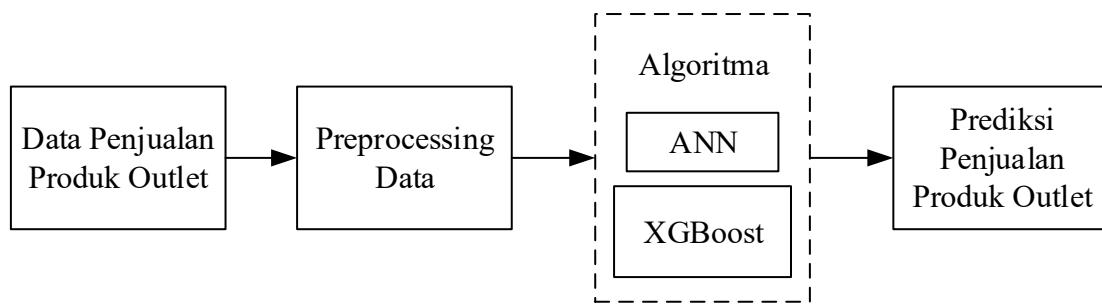
**Tabel 3. 2 Hasil normalisasi data**

Date	Promo	Lokasi	Kompetitor	LP	Harga	Penjualan
1/1/2023	1	BASMALAH CABANG JATIKERTO	0.5	0	0	0.347032
1/1/2023	1	BASMALAH CABANG PAKISAJI	0.25	0	0.764179	0.593607
1/1/2023	1	BASMALAH CABANG KEPANJEN MALANG	0.75	0	0.573134	0.333333
1/1/2023	1	BASMALAH CABANG LOWOKWARU	0.5	0	0.41194	0.703196
1/1/2023	1	BASMALAH CABANG PAKISAJI	0	0	0.940299	0.027397
1/1/2023	0	BASMALAH CABANG DAMPIT MALANG	1	0	0.6	0.762557
1/1/2023	0	BASMALAH CABANG PAGELARAN	0.5	1	1	0.347032
1/1/2023	1	BASMALAH CABANG BUMIAYU	0.25	1	0.835821	0.105023
1/1/2023	1	BASMALAH CABANG BUMIAYU	0.75	1	0.504478	1
...	...	...	...	...	...	...
12/31/20 24	0	BASMALAH CABANG TALANGSUKO	1	0	355000	0.591045
12/31/20 24	1	BASMALAH CABANG WAGIR MALANG	0.5	0	304000	0.438806

Setelah data dinormalisasi, proses selanjutnya adalah split data/pemisahan data yang bertujuan untuk memisahkan data pelatihan dari data pengujian. Tahapan pendistribusian data membagi data ke dalam dua kategori: data latih dan data uji. Data yang didistribusikan pada penilaian ini terdiri dari beberapa komposisi untuk memperoleh kinerja model terbaik. Pembagian data ditentukan berdasarkan hasil eksperimen yang telah dilakukan, dengan mempertimbangkan skema distribusi yang berpotensi menghasilkan kinerja algoritma paling optimal selama proses pengujian.

### 3.1.3 Desain sistem

Gambar 3.3 menunjukkan alur proses desain sistem dalam penelitian ini. Ini menjelaskan bagaimana sistem menggunakan Algoritma ANN dan XGBoost untuk memprediksi penjualan produk outlet Toko Basmalah di Wilayah Kabupaten/Kota Malang.



**Gambar 3.2 Desain sistem**

#### A. Penjualan produk outlet

Data penjualan produk pada outlet Toko Basmalah yang berada di wilayah Kabupaten/Kota Malang diproses melalui sistem yang dikembangkan. Tahapan ini melibatkan penyusunan data hasil pengolahan ke dalam bentuk dataset awal berformat CSV, yang terdiri atas beberapa atribut penting di bawah ini:

**Tabel 3.3 Rincian data penjualan produk outlet Toko Basmalah**

No	Nama Atribut	Interval Nilai	Jenis Data	Satuan
1	Date	-	Date	-
2	Promo	0 - 1	Numerik	Kategori
3	Lokasi	-	Text	-
4	Kompetitor	1 - 5	Numerik	Unit

5	LP	0 - 1	Numerik	Kategori
6	Harga	100000 - 499000	Numerik	IDR
7	Penjualan	0 - 329	Numerik	Unit

Tabel 3.3 menunjukkan enam atribut data dengan rentang nilai dan tipe data yang berbeda-beda. Atribut Date merupakan data tanggal transaksi penjualan pada setiap outlet dengan tipe date. Atribut Promo merupakan indikator adanya promo pada setiap outlet dengan penilaian angka “0” jika tidak ada promo dan penilaian angka “1” jika ada promo pada outlet tertentu. Tipe data ini termasuk pada tipe data numerik dengan rentang nilai antara 0 sampai dengan 1 dan termasuk dalam satuan kategori. Atribut Lokasi merupakan lokasi cabang outlet Toko Basmalah yang tersebar di Wilayah Kabupaten/Kota Malang. Atribut Kompitor merupakan indikator keberadaan competitor yang aktif pada setiap titik outlet tertentu. Tipe data ini termasuk pada tipe data numerik dengan rentang nilai antara 1 sampai dengan 5 dan termasuk dalam satuan kategori. Atribut LP merupakan loyalitas pelanggan pada setiap outlet. Tipe data ini termasuk tipe data numerik dengan penilaian angka “0” jika pelanggan tidak loyal dan penilaian angka “1” jika pelanggan loyal pada outlet tertentu. Atribut Harga merupakan harga produk dalam satuan rupiah (IDR) dengan rentang nilai antara 100.000 sampai dengan 499.000 dan termasuk dalam tipe data numerik. Atribut Penjualan merupakan jumlah unit produk yang terjual pada setiap outlet dengan rentang nilai antara 0 sampai dengan 329.

Selain itu, detail data tersebut terutama pada atribut Harga dan Penjualan menunjukkan perbedaan selisih rentang nilai secara signifikan yaitu sekitar 399.000 IDR pada atribut Harga dan 329 Unit pada atribut Penjualan. Prasetyo (2017) menyatakan bahwa kondisi tersebut berpotensi meningkatkan tingkat kesalahan sistem dalam proses prediksi. Oleh karena itu, pada proses selanjutnya, proses normalisasi data harus dilakukan.

## B. Data Proses

ahapan ini mencakup proses normalisasi data yang bertujuan menangani nilai kosong, dengan memanfaatkan pendekatan min–max scaling pada data yang telah disusun sebelumnya. Ini dilakukan untuk mengurangi kesalahan saat pengujian model prediktif. Rumus berikut digunakan untuk menghitung skala min–max:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Dimana:

$x$  = menyatakan nilai data sebelum diproses normalisasi

$x_{norm}$  = menunjukkan nilai data setelah dinormalisasi

$\min(x)$  = merupakan nilai minimum dari keseluruhan data

$\max(x)$  = nilai maksimum dari keseluruhan data.

### C. Algoritma ANN

Artificial Neural Networks (ANN) merupakan sistem pemrosesan informasi yang terinspirasi dari cara kerja otak manusia, di mana proses penyelesaian masalah dilakukan melalui mekanisme pembelajaran yang melibatkan penyesuaian bobot dan hubungan antar neuron. ANN menggunakan pemahaman data masa lalu untuk membuat keputusan berdasarkan data yang telah dipahami sebelumnya.

Menurut Da Silva (2017), ANN terdiri dari tiga lapisan yaitu lapisan input, lapisan tersembunyi, dan lapisan output. Lapisan input menerima informasi /data, fitur, dan sinyal dari lingkungan luar. Lapisan tersembunyi (*hidden layer*) tersusun atas neuron-neuron yang membentuk pola terkait proses pembelajaran dan menjalankan sebagian besar komputasi internal jaringan. Sementara itu, lapisan keluaran (*output layer*) berfungsi menghasilkan informasi serta sinyal sebagai hasil akhir pemrosesan.

Sebagaimana telah dijelaskan sebelumnya, neuron dalam arsitektur Artificial Neural Network (ANN) disusun ke dalam beberapa lapisan. Berdasarkan jumlah lapisan tersembunyi yang digunakan, ANN diklasifikasikan menjadi jaringan satu lapisan dan jaringan multilapisan. Jaringan satu lapisan hanya memiliki satu lapisan bobot yang saling terhubung, di mana data masukan diproses secara langsung menjadi keluaran tanpa melalui lapisan tersembunyi. Sebaliknya, jaringan multilapisan memiliki satu atau lebih lapisan tersembunyi, yaitu lapisan yang terletak di antara lapisan masukan dan lapisan keluaran. Penggunaan fungsi aktivasi nonlinier pada ANN multilapisan memungkinkan model ini menangani dan menyelesaikan berbagai jenis permasalahan yang lebih kompleks.

Selain itu, Wibowo dan Rikumahu (2019) serta Rizky menyatakan bahwa proses pelatihan jaringan sangat diperlukan agar model mampu mengenali pola nilai dari data historis sebagai dasar dalam memprediksi pola di masa mendatang. Dalam konteks

algoritma, fungsi pelatihan berperan penting dalam membantu jaringan memahami model yang dibangun. Banyaknya pilihan metode pelatihan yang tersedia sering kali menimbulkan kebingungan dalam menentukan pendekatan yang paling sesuai untuk digunakan. Oleh karena itu, penelitian ini memanfaatkan beberapa parameter utama, yaitu Tingkat pembelajaran, metode optimasi, dan fungsi aktivasi, yang berpengaruh langsung terhadap kinerja jaringan yang dirancang. Konfigurasi model pada penelitian ini mencakup nilai learning rate sebesar 0,2 dengan algoritma optimasi Adam, serta penerapan fungsi aktivasi ReLU. Fungsi aktivasi tersebut dipilih karena sesuai dengan kebutuhan rentang keluaran model yang berada pada nilai nol hingga negatif.

Tahapan forward propagation digunakan untuk menilai tingkat kecocokan model yang dibangun, sementara proses backpropagation berfungsi untuk menyesuaikan bobot jaringan melalui perhitungan kesalahan secara bertahap dilakukan dengan alur perhitungan terbalik guna menyesuaikan dan meminimalkan tingkat kesalahan pada model. Pada lapisan utama neuron buatan, mekanisme umpan balik menghasilkan pembaruan bobot serta peningkatan nilai akurasi. Untuk memperoleh prediksi penjualan produk outlet yang optimal, siklus pembelajaran ini diulang secara bertahap. Selanjutnya, jaringan diinisialisasi kembali dengan bobot acak, kemudian gradien dari fungsi kesalahan dihitung sebagai dasar dalam melakukan penyesuaian bobot awal jaringan menggunakan persamaan berikut.:

$$net = w_0 + \sum_{i=1}^n x_i w_i \quad (2)$$

*net* = menunjukkan hasil keluaran dari node pada tahap pemrosesan

*w<sub>0</sub>* = merupakan bobot bias pada node

*i* = menunjukkan indeks penjumlahan yang dimulai dari 1 hingga *n*

*x<sub>i</sub>* = menunjukkan bobot penghubung antara node dan neuron masukan ke-*i*.

*w<sub>i</sub>* = merepresentasikan nilai input ke-*i* yang berasal dari node pada lapisan sebelumnya

*n* = merupakan total fitur yang digunakan

Selanjutnya, fungsi transfer logistik sigmoid digunakan untuk menyediakan fasilitas yang dapat mengkonversi nilai masukan menjadi nilai keluaran dengan menggunakan rumus berikut:

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

Di mana:

$x$  = input ke neuron

$e$  = angka konstan

#### D. Algoritma XGBoost

Metode Algoritma *Extreme Gradient Boosting (XGBoost)* merupakan pohon regresi yang memiliki aturan keputusan yang sama dengan *decision tree*. *XGBoost* dapat digunakan untuk regresi dan klasifikasi. Algoritma ini termasuk dalam varian yang efisien dan terukur dari *gradient boosting machine* (GBM) yang telah banyak diterapkan dalam *computer vision*, *data mining* dan bidang lainnya. *XGBoost* telah ditingkatkan dalam dua aspek yaitu untuk mempercepat konstruksi pohon dan mengusulkan algoritma terdistribusi baru untuk pencarian pohon (Ma et al., 2020).

Pada algoritma *XGBoost*, pohon variabel berfungsi untuk memperkirakan variabel target. Pohon pertama berfungsi untuk memprediksi nilai target, pohon kedua berfungsi untuk memprediksi selisih antara estimasi tujuan dan pohon pertama, dan seterusnya hingga residunya diminimalkan (Lee et al., 2023). Nilai prediksi pada algoritma XGBoost ditentukan oleh pohon K seperti dalam persamaan berikut:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{Y} \quad (4)$$

Dimana:

$\hat{y}_i$  = prediksi model untuk daya ke- $i$

$K$  = jumlah total pohon keputusan yang dibangun oleh model

$f_k(x_i)$  = Fungsi prediksi dari pohon ke- $k$  untuk data input  $x_i$ . Setiap pohon  $f_k$  merupakan *regression tree*.

$\mathcal{Y}$  = Ruang fungsi kemungkinan dari semua pohon keputusan.

Untuk meminimalkan sisa fungsi kerugian ( $L$ ) pada tahap pelatihan algoritma *XGBoost* menggunakan persamaan berikut antara nilai sebenarnya ( $y$ ) dan nilai prediksi ( $\hat{y}$ ):

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^n \Omega(f_k) \quad (5)$$

Dimana:

$L$  = total loss (fungsi kerugian total)  
 $l(y_i, \hat{y}_i)$  = fungsi kerugian untuk sampel ke- $i$   
 $y_i$  = nilai aktual  
 $\hat{y}_i$  = nilai prediksi  
 $n$  = jumlah sampel data  
 $\Omega(f_k)$  = regularisasi untuk model ke- $k$   
 $k$  = jumlah model dasar

## E. Prediksi penjualan produk outlet

Pada penelitian ini, pengukuran kinerja prediksi penjualan produk outlet dilakukan menggunakan beberapa metrik evaluasi, yaitu Error Kuadrat Rata-rata (Mean Squared Error/MSE), Error Absolut Rata-rata (Mean Absolute Error/MAE), dan Koefisien Determinasi ( $R^2$ ).

Pengukuran MSE dirumuskan sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (7)$$

$MSE$  = merupakan nilai rata-rata kesalahan  
 $n$  = menyatakan total t  
 $Y_i$  = adalah nilai hasil pengamatan  
 $\hat{Y}_i$  = menunjukkan nilai hasil prediksi

Mean Squared Error (MSE) dipakai sebagai acuan untuk mengetahui besarnya rata-rata kesalahan prediksi dengan memberikan bobot penalti yang lebih tinggi terhadap kesalahan yang bersifat ekstrem. Semakin kecil nilai MSE, maka semakin baik kinerja model prediksi. Selanjutnya, MAE digunakan untuk mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, yang dirumuskan sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (8)$$

MAE memberikan interpretasi yang lebih intuitif karena kesalahan dihitung dalam satuan yang sama dengan data asli. Nilai MAE yang lebih kecil menunjukkan bahwa prediksi model semakin mendekati nilai aktual. Selain itu, kinerja model juga dievaluasi menggunakan Koefisien Determinasi ( $R^2$ ) yang dirumuskan sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (9)$$

di mana:  
 $R^2$  = koefisien determinasi

$\bar{Y}$  = nilai rata-rata dari data aktual

Koefisien determinasi ( $R^2$ ) mencerminkan kemampuan model dalam menjelaskan variasi pada data penjualan. Nilai  $R^2$  yang mendekati angka 1 menunjukkan bahwa model memiliki kinerja prediksi yang semakin baik, sedangkan nilai  $R^2$  yang mendekati nol atau bernilai negatif menandakan bahwa model belum mampu menjelaskan variasi data secara optimal.

### 3.1.3. 1 Ekstraksi Fitur

Bertujuan untuk menyederhanakan dataset dengan mengurangi jumlah fitur yang digunakan. Ekstraksi fitur dilakukan setelah tahapan rekayasa data selesai, khususnya pada saat atribut-atribut penentu pengujian telah ditetapkan. Detail tahapan ekstraksi fitur dapat dilihat pada Tabel 3.4.:

**Tabel 3. 4 Rincian ekstraksi fitur**

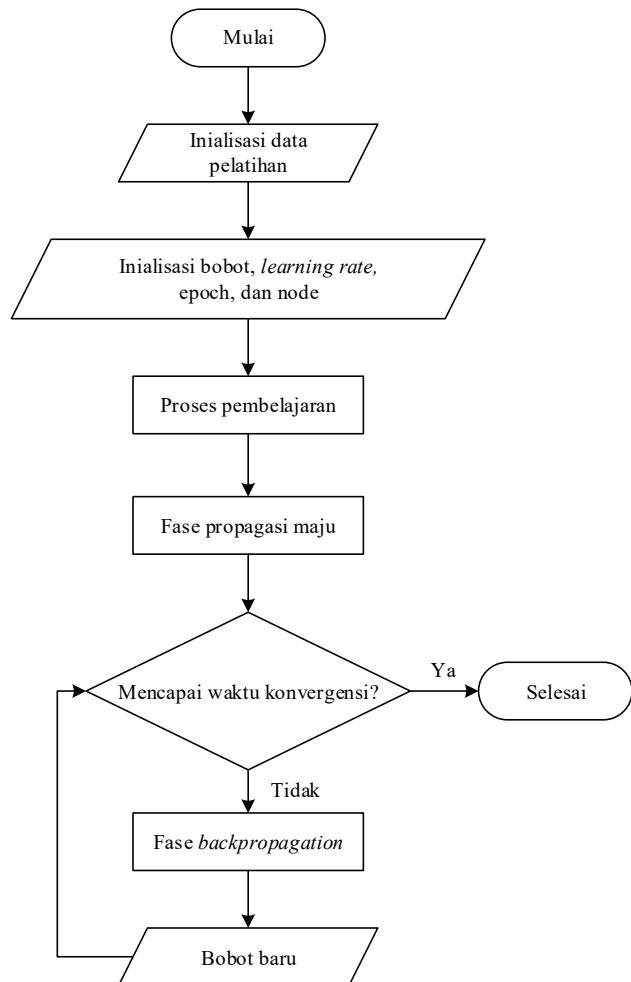
Nama Atribut	Jenis data	Keterangan
Promo	Numerik	-
Lokasi	Text	Dikurangi
Kompetitor	Numerik	-
LP	Numerik	-
Harga	Numerik	-
Penjualan	Numerik	Variabel Target
Tanggal	Date	Dikurangi

Berdasarkan Tabel 3.4, penelitian ini menggunakan tujuh atribut, di mana tiga atribut, yaitu Tanggal dan Lokasi, dieliminasi karena tidak diperlukan dalam proses analisis.

### 3.1. 4 Implementasi Sistem

*Artificial Neural Network* tersusun atas sejumlah node buatan yang menjalankan tiga fungsi utama, yaitu proses perkalian, penjumlahan, dan aktivasi. Pada tahap masukan, neuron buatan memiliki bobot, di mana nilai input dikalikan dengan bobot masing-masing. Setelah dilakukan penggabungan bobot dengan bias, nilai tersebut diproses pada tahap aktivasi dan selanjutnya dialirkan menuju lapisan keluaran jaringan ANN. Dalam penelitian ini, algoritma *backpropagation* digunakan untuk memprediksi penjualan produk pada objek penelitian yang

telah ditentukan. Model ANN berbasis *backpropagation* merupakan salah satu metode pembelajaran terawasi (*supervised learning*). Proses pelatihan memanfaatkan nilai *Mean Squared Error* (MSE) sebagai kriteria penghentian sekaligus acuan dalam memperoleh bobot akhir yang optimal. Dengan demikian, proses pelatihan akan dihentikan ketika kondisi konvergensi telah tercapai. Adapun alur pelatihan model ANN *backpropagation* dalam penelitian ini disajikan pada Gambar 3.3.



**Gambar 3.3 Flowchart proses pelatihan ANN Backpropagation**

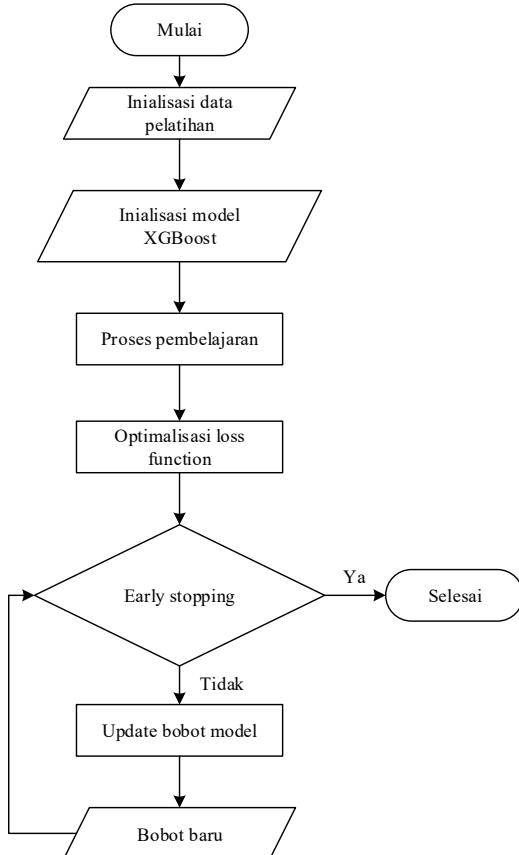
Langkah-langkah perhitungan pada prediksi penjualan produk menggunakan model ANN *Backpropagation* di antaranya:

1. Tahap penyiapan data masukan
2. Penetapan parameter pada model
3. roses propagasi ke depan
4. Tahap ini dilakukan dengan membandingkan nilai keluaran model terhadap nilai target, di mana selisih keduanya dihitung menggunakan metode MSE. Ketika kriteria

konvergensi terpenuhi, proses dihentikan; sebaliknya, apabila belum tercapai, pembaruan bobot dilakukan melalui mekanisme backpropagation.

5. Fase Propagasi Balik
6. Fase *backpropagation*

Sedangkan, model XGBoost (*Extreme Gradient Boosting*) adalah metode machine learning yang berbasis pada algoritma *gradient boosting*, yang digunakan untuk melakukan prediksi terhadap variabel target, baik yang bersifat kontinu (regresi) maupun kategorikal (klasifikasi). Dalam konteks regresi, XGBoost digunakan untuk memprediksi suatu variabel terikat (target) berdasarkan satu atau lebih variabel bebas. Berbeda dengan regresi linier yang mengasumsikan hubungan linier antar variabel, XGBoost mampu menangkap hubungan yang lebih kompleks dan non-linier antara variabel bebas dan variabel target. Model ini bekerja dengan membangun sejumlah pohon keputusan secara bertahap dan menggabungkannya untuk menghasilkan prediksi yang lebih akurat dan kuat. Pada penelitian ini, penulis menggunakan algoritma *gradient boosting* dalam memprediksi penjualan produk pada objek yang telah disebutkan. Algoritma XGBoost termasuk dalam kelompok teknik pembelajaran mesin dengan pendekatan terawasi (supervised learning). Proses *training* dilakukan dengan membangun sejumlah pohon keputusan secara bertahap melalui teknik gradient boosting, dengan tujuan meminimalkan kesalahan prediksi. Evaluasi model dilakukan menggunakan metrik seperti *Mean Square Error (MSE)* untuk mengukur tingkat akurasi prediksi. Pelatihan akan berakhir ketika model telah mencapai konvergensi atau tidak terjadi peningkatan signifikan pada performa model. Selanjutnya, tahapan flowchart pelatihan pada penelitian ini dalam memprediksi penjualan produk menggunakan model XGBoost dapat dilihat pada Gambar 3.



**Gambar 3. 4 Flowchart proses pelatihan XGBoost**

Langkah-langkah perhitungan pada prediksi penjualan produk menggunakan model XGBoost adalah:

1. Inisialisasi data yang digunakan sebagai data pelatihan
2. Inialisasi parameter XGBoost
3. Optimalisasi loss function
4. Setelah nilai keluaran diperoleh, selisih antara hasil prediksi dan nilai target dihitung menggunakan persamaan Mean Squared Error (MSE). Apabila nilai loss function telah memenuhi kriteria yang ditetapkan, proses pelatihan dihentikan; sebaliknya, jika belum terpenuhi, maka dilakukan pembaruan bobot pada jaringan.
5. Fase update bobot.
6. Fase bobot baru.

### 3.1. 5 Eksperimen

Tahap pengujian dalam penelitian ini dilaksanakan menggunakan perangkat lunak berbasis Python untuk memprediksi penjualan produk outlet di masa mendatang dengan menerapkan algoritma ANN dan XGBoost. Setelah diperoleh hasil prediksi dari kedua

algoritma tersebut, tingkat kesalahan masing-masing model dibandingkan berdasarkan parameter Mean Squared Error (MSE). Kesimpulan penelitian kemudian dirumuskan berdasarkan temuan yang diperoleh dari proses analisis. Secara konseptual, tidak terdapat ketentuan baku dalam menentukan arsitektur jaringan yang optimal pada model ANN maupun XGBoost. Oleh karena itu, penentuan konfigurasi jaringan dilakukan melalui pendekatan trial and error guna memperoleh kinerja yang maksimal. Dengan mempertimbangkan hal tersebut, dilakukan pengujian menggunakan berbagai komposisi data latih dan data uji untuk menentukan performa model yang menghasilkan nilai MSE paling kecil.

Selain variasi komposisi data dan penggunaan K-Fold cross-validation, proses eksperimen juga memperhatikan konsistensi parameter pelatihan agar perbandingan antar model dapat dilakukan secara objektif. Setiap skenario pelatihan pada algoritma ANN dan XGBoost dievaluasi menggunakan metrik kesalahan yang sama, yaitu MSE, MAE, dan  $R^2$  sehingga perbedaan performa yang dihasilkan benar-benar mencerminkan pengaruh dari variasi pembagian data dan strategi validasi yang digunakan. Pendekatan ini memungkinkan peneliti untuk mengidentifikasi konfigurasi model yang paling optimal secara kuantitatif, serta memberikan dasar yang kuat dalam menentukan model terbaik untuk digunakan pada tahap pengujian dan analisis lanjutan.

Selain pengujian berdasarkan variasi rasio data latih dan data uji, penelitian ini juga menerapkan skenario pelatihan yang dibedakan berdasarkan karakteristik pola data time series penjualan. Pola tersebut diklasifikasikan ke dalam tiga kategori berdasarkan rentang nilai jumlah penjualan, yaitu periode penjualan menurun, stabil, dan meningkat. Klasifikasi ini digunakan untuk merepresentasikan dinamika pasar yang berbeda secara kuantitatif. Secara rinci, kategori penjualan menurun ditetapkan pada rentang nilai 0–109 dengan total 5.196 data, kategori penjualan stabil berada pada rentang 110–219 dengan jumlah 4.694 data, sedangkan kategori penjualan meningkat mencakup rentang 220–329 dengan total 1.376 data. Parameter utama yang digunakan dalam pengelompokan ini adalah jumlah penjualan pada setiap periode pengamatan.

Pendekatan berbasis segmentasi tren ini bertujuan untuk mengevaluasi sejauh mana model Artificial Neural Network (ANN) dan Extreme Gradient Boosting (XGBoost) mampu beradaptasi terhadap perubahan pola penjualan yang berbeda. Dengan mempertimbangkan variasi kondisi pasar tersebut, hasil prediksi yang dihasilkan diharapkan menjadi lebih

kontekstual, robust, dan aplikatif dalam mendukung proses pengambilan keputusan bisnis berbasis data.

Selain mempertimbangkan rasio data latih dan uji serta karakteristik tren penjualan, penelitian ini juga melakukan pengelompokan data berdasarkan tingkat tekanan kompetitor, yang direpresentasikan dalam skala nilai 1 hingga 5. Klasifikasi ini digunakan untuk mencerminkan kondisi pasar dengan intensitas persaingan yang berbeda. Secara rinci, kompetitor rendah (nilai 1–2) mencerminkan pasar yang relatif stabil dengan persaingan minimal, dengan total 4.496 data; kompetitor sedang (nilai 3) menggambarkan kondisi persaingan moderat dengan 2.283 data; sedangkan kompetitor tinggi (nilai 4–5) menunjukkan pasar yang sangat kompetitif dan dinamis, dengan 4.487 data. Parameter utama dalam pengelompokan ini adalah jumlah kompetitor yang beroperasi di pasar pada periode pengamatan.

Pendekatan ini bertujuan untuk mengevaluasi ketahanan dan adaptabilitas model Artificial Neural Network (ANN) maupun Extreme Gradient Boosting (XGBoost) dalam memprediksi penjualan pada kondisi persaingan yang berbeda. Dengan mempertimbangkan variasi tekanan kompetitor, hasil prediksi diharapkan lebih relevan dan aplikatif untuk mendukung strategi bisnis berbasis data.

**Tabel 3. 5 Variasi skenario uji coba**

Algoritma	Kode	Komposisi Data	K-Fold	Data Latih	Data Uji	Nama Pelatihan
ANN	A	50:50	k=5	5634	5632	ANN-A
	B	60:40		6760	4506	ANN-B
	C	70:30		7887	3379	ANN-C
	D	80:20		9014	2252	ANN-D
	E	90:10		10.140	1126	ANN-E
	F	50:50		5634	5632	ANN-F
	G	60:40		6760	4506	ANN-G
	H	70:30		7887	3379	ANN-H
	I	80:20		9014	2252	ANN-I
	J	90:10		10.140	1126	ANN-J
	K	90:10		4.676	520	ANN-K
	L	90:10		4.225	469	ANN-L
	M	90:10		1.238	138	ANN-M
	N	90:10		4.046	450	ANN-N
	O	90:10		2.055	228	ANN-O
	P	90:10		4.038	449	ANN-P
XGBoost	A	50:50	k=5	5634	5632	XGB-A
	B	60:40		6760	4506	XGB-B
	C	70:30		7887	3379	XGB-C
	D	80:20		9014	2252	XGB-D
	E	90:10		10.140	1126	XGB-E
	F	50:50		5634	5632	XGB-F
	G	60:40		6760	4506	XGB-G
	H	70:30		7887	3379	XGB-H
	I	80:20		9014	2252	XGB-I
	J	90:10		10.140	1126	XGB-J
	K	90:10		4.676	520	XGB-K
	L	90:10		4.225	469	XGB-L
	M	90:10		1.238	138	XGB-M
	N	90:10		4.046	450	XGB-N
	O	90:10		2.055	228	XGB-O
	P	90:10		4.038	449	XGBP

Tabel 3.5 menampilkan berbagai skenario pengujian yang diterapkan dalam penelitian ini dengan total sebanyak 11.266 data. Proses pengujian dilakukan dengan memvariasikan pembagian data latih dan data uji, baik tanpa penerapan validasi silang maupun dengan menggunakan validasi silang K-fold ( $k = 5$ ). Pada skenario tanpa K-fold, data dibagi ke dalam lima komposisi utama, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10 (model A–E). Pendekatan ini diterapkan pada algoritma ANN maupun XGBoost untuk mengevaluasi pengaruh perbedaan proporsi data terhadap performa model.

Sementara itu, pada skenario dengan validasi silang K-fold, digunakan komposisi data yang sama dengan kode F–J, di mana setiap lipatan secara bergantian dijadikan data uji, sehingga hasil evaluasi model menjadi lebih stabil dan representatif. Selain itu, penelitian ini juga memasukkan skenario tambahan dengan pembagian 90:10 berdasarkan karakteristik data spesifik, seperti tren penjualan (turun, stabil, naik) dan tingkat tekanan kompetitor (rendah, sedang, tinggi), yang masing-masing memiliki kode pelatihan tersendiri (K–P untuk ANN dan XGBP untuk XGBoost). Pemberian kode ini memudahkan analisis dan perbandingan performa model pada kondisi pasar yang berbeda. Dengan demikian, kombinasi skenario uji dan algoritma yang beragam memungkinkan penelitian untuk mengevaluasi robustness, adaptabilitas, dan akurasi prediksi dari masing-masing model secara komprehensif.

### 3. 2 Research Instrument

Dalam penelitian ini, variabel bebas direpresentasikan oleh atribut dataset penjualan produk outlet, yang mencakup berbagai karakteristik historis penjualan yang digunakan sebagai masukan (input) dalam proses pemodelan. Atribut-atribut tersebut diproses menggunakan algoritma Artificial Neural Network (ANN) dan XGBoost sebagai mekanisme utama dalam membangun model prediksi. Algoritma ini berperan sebagai instrumen komputasi yang mengolah hubungan kompleks antar variabel input untuk menghasilkan keluaran berupa nilai prediksi penjualan produk outlet. Dengan demikian, ANN dan XGBoost tidak diposisikan sebagai variabel penelitian, melainkan sebagai alat analisis yang menjembatani hubungan antara data masukan dan hasil prediksi yang dihasilkan. Parameter penelitian yang digunakan, termasuk variabel independen, variabel terikat, variabel intervening, dan variabel dependen, dirangkum dalam Tabel 3.6.

**Tabel 3. 6 Variabel penelitian**

Independen Variable	Proses Utama	Intervening Variable	Variabel Independen
Atribut dataset produk outlet	ANN, XGBoost	Prediksi penjualan produk outlet	Nilai Error

Fokus utama penelitian adalah variabel independen, juga dikenal sebagai variabel terikat. Selain itu, variabel terikat adalah variabel yang dipengaruhi atau akibat dari variabel bebas. Dalam penelitian ini, nilai error adalah variabel terikat. Secara konseptual, variabel

intervening atau variabel penghubung berperan dalam memengaruhi hubungan antara variabel bebas dan variabel terikat secara tidak langsung, serta tidak dapat diamati maupun diukur secara langsung. Variabel ini berada di antara variabel independen dan variabel dependen, sehingga tidak memberikan pengaruh langsung terhadap perubahan atau kemunculan variabel dependen.

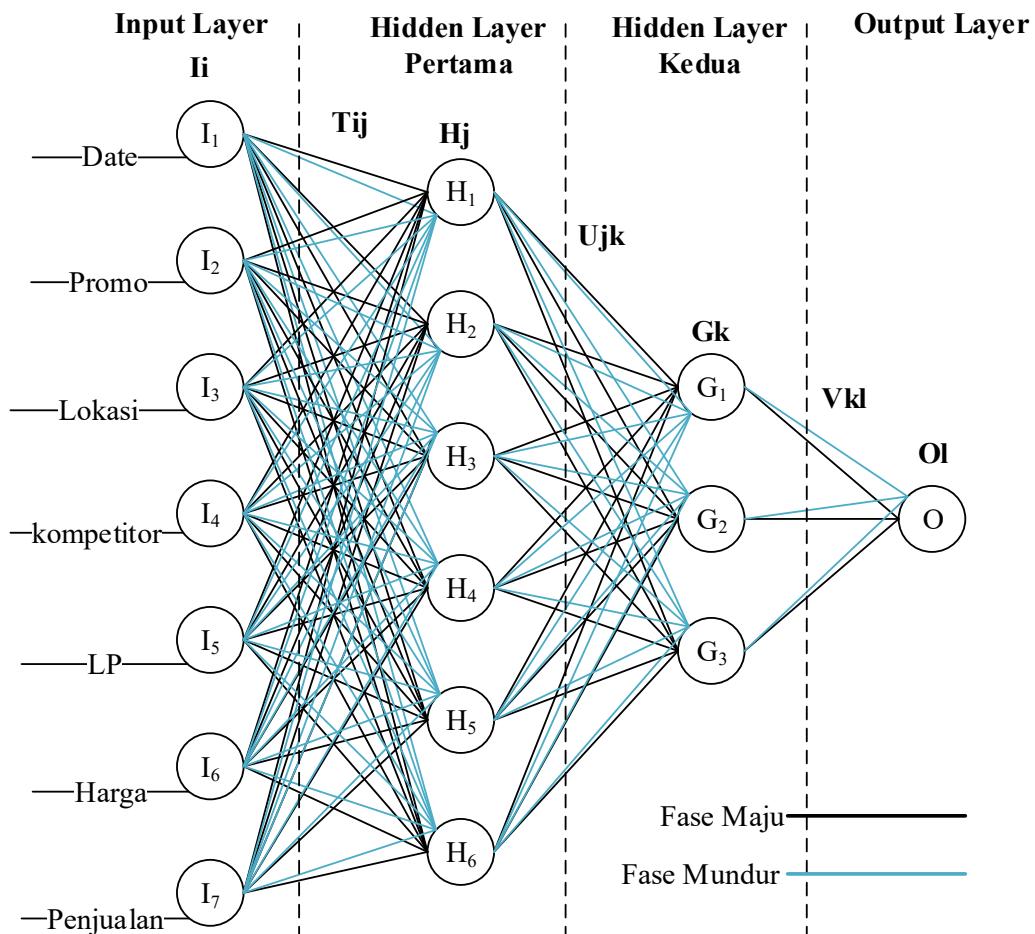
Variabel intervening yang digunakan dalam penelitian ini berupa hasil prediksi penjualan produk outlet yang berfungsi sebagai penghubung antara variabel bebas dan variabel terikat. Prediksi penjualan merupakan hasil tidak langsung dari proses pembelajaran model terhadap data historis, sehingga keberadaannya dipengaruhi oleh kualitas data, komposisi data latih dan uji, serta konfigurasi model yang digunakan. Selanjutnya, variabel terikat dalam penelitian ini adalah nilai error, yang diukur menggunakan metrik evaluasi seperti MSE, MAE, dan  $R^2$ . Nilai error digunakan untuk menilai sejauh mana hasil prediksi mendekati nilai aktual, sehingga menjadi indikator utama dalam menentukan tingkat akurasi dan kinerja model. Hubungan ini menegaskan bahwa perubahan pada variabel bebas dan proses pemodelan akan tercermin secara kuantitatif melalui besarnya nilai error yang dihasilkan.

## BAB IV

### PREDIKSI PENJUALAN DENGAN ARTIFICIAL NEURAL NETWORK (ANN)

#### 4. 1 Desain Jaringan

Desain penelitian adalah serangkaian proses yang kemudian disusun untuk menjawab permasalahan yang dikaji. Melalui tahapan itu, objek penelitian dapat diidentifikasi secara jelas sehingga bisa dijadikan sebagai landasan dalam penarikan hasil akhir. Penelitian ini, peneliti menerapkan algoritma *Artificial Neural Network* (ANN) dengan *backpropagation* guna memprediksi penjualan produk pada Toko Basmalah yang berada di wilayah Kabupaten/Kota Malang berdasarkan data penjualan produk toko tersebut pada periode 01 Januari 2023 – 31 Desember 2024.



**Gambar 4. 1 Arsitektur jaringan pada model Artificial Neural Network (ANN)**

Gambar 4.1 memperlihatkan desain arsitektur jaringan algoritma backpropagation pada model *Artificial Neural Network* (ANN) yang digunakan dalam penelitian ini. Arsitektur tersebut terdiri dari dua hidden layer, dengan data masukan (Ii) berupa data penjualan produk. Arsitektur jaringan ANN berbasis backpropagation yang digunakan dalam penelitian ini ditunjukkan pada Gambar 4.1. Model tersebut terdiri atas dua lapisan tersembunyi, dengan input (Ii) berupa data penjualan produk. Hubungan antara lapisan input dan lapisan tersembunyi pertama dimodelkan melalui bobot  $T_{ij}$ , sedangkan output dari lapisan tersembunyi pertama dinyatakan sebagai  $H_j$ . Hubungan berikutnya antara lapisan tersembunyi pertama dan kedua direpresentasikan oleh bobot  $U_{jk}$ , dengan keluaran lapisan tersembunyi kedua dinyatakan sebagai  $G_k$ . Selanjutnya, bobot  $V_{kl}$  menghubungkan lapisan tersembunyi kedua dengan lapisan output, di mana nilai keluaran akhir berupa prediksi penjualan dinyatakan sebagai  $O_l$ .

## 4. 2 Implementasi Model ANN

Tahap selanjutnya dalam penelitian ini adalah implementasi jaringan saraf tiruan, yaitu proses pengolahan data menggunakan Kaggle Editor melalui pemrograman untuk memperoleh keluaran bentuk prediksi penjualan produk outlet berdasarkan arsitektur jaringan yang telah dirancang. Sebelum proses pelatihan dimulai, dilakukan inisialisasi bobot awal secara acak untuk setiap koneksi antar neuron guna menghindari bias awal pada model. Pada model ANN yang digunakan, fungsi aktivasi ReLU diterapkan pada hidden layer, sedangkan fungsi aktivasi linear digunakan pada output layer untuk menyesuaikan karakteristik keluaran berupa nilai kontinu. Pada Tabel 4.1 proses pelatihan jaringan dilakukan menggunakan Adam Optimizer dengan variasi learning rate sebesar 0,2 dan 0,01.

**Tabel 4. 1 Parameter model ANN**

Model network	ANN dengan algoritma backpropagation
Fungsi aktivasi ( <i>activation function</i> )	ReLU
Fungsi transfer logistik	Sigmoid Function
Optimizer	Adam Optimizer
Evaluasi kinerja	MSE, MAE, R2
Input layer	7 node
Lapisan tersembunyi pertama	6 node
Lapisan tersembunyi kedua	3 node
Output layer	1 node
Leaning rate	0,2;0,01
Epoch	<i>Early stopping maximun 50;300</i>
Normalisasi data	Min–Max Scaling
K-Fold cross-validation	k=5
Fungsi aktivasi output layer	Linear

Model jaringan menggunakan dua lapisan tersembunyi dengan konfigurasi 7-6-3-1, yaitu tujuh node pada input layer, enam node pada hidden layer pertama, tiga node pada hidden layer kedua, dan satu node pada output layer. Seluruh data terlebih dahulu dinormalisasi menggunakan metode Min–Max Scaling untuk mempercepat proses konvergensi dan meningkatkan stabilitas pelatihan jaringan. Proses pelatihan dilakukan dengan menerapkan early stopping dengan batas maksimum 50 dan 300 epoch, serta menggunakan validasi silang K-Fold ( $k = 5$ ) untuk meningkatkan reliabilitas dan kemampuan generalisasi model. Evaluasi kinerja jaringan dilakukan menggunakan metrik Mean Squared Error (MSE), Mean Absolute Error (MAE), dan Koefisien Determinasi ( $R^2$ ). Penentuan jumlah node pada masing-masing layer didasarkan pada karakteristik data yang digunakan, yaitu tujuh variabel input, serta kebutuhan model dalam menangkap pola hubungan nonlinier antara variabel input dan output penjualan. Konfigurasi arsitektur jaringan dipilih berdasarkan kinerja terbaik yang ditunjukkan oleh nilai error terkecil selama proses pelatihan dan pengujian.

#### 4. 3 Pelatihan Model ANN

Proses pelatihan model dilaksanakan dengan memanfaatkan data masukan yang telah dibagi ke dalam data pelatihan dan data pengujian. Pembagian ini didasarkan pada data yang telah dinormalisasi sebelumnya, untuk mengevaluasi kinerja model. Selama pelatihan, dilakukan variasi pada distribusi data yang digunakan, dan setiap variasi diberi nama berbeda, yaitu ANN-A, ANN-B, ANN-C, ANN-D, ANN-E, ANN-F, ANN-G, ANN-H, ANN-I, ANN-

J, ANN-K, ANN-L, ANN-M, ANN-N, ANN-O, dan ANN-P. Dalam seluruh proses pelatihan ini, atribut Penjualan digunakan sebagai variabel target yang ingin diprediksi.

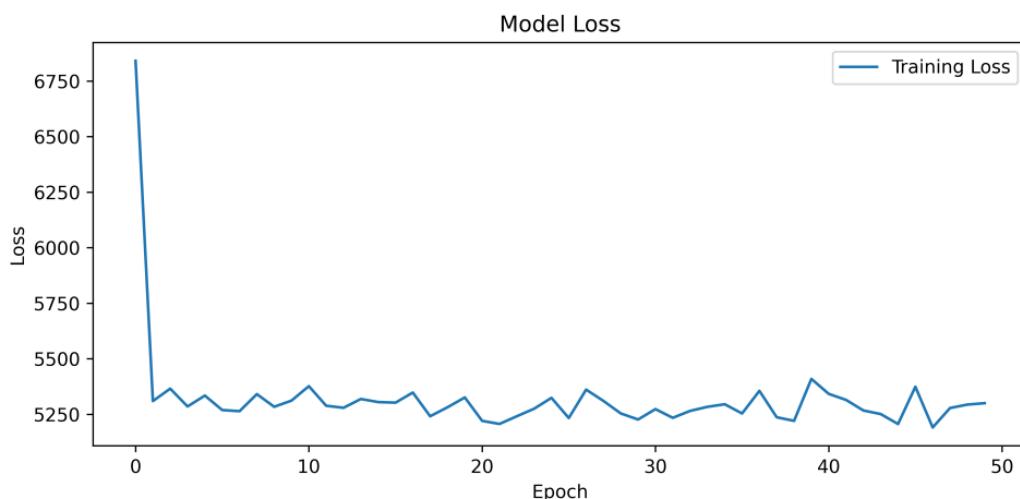
#### 4.3.1 Pelatihan Model ANN-A

Hasil pelatihan model ANN-A dengan pembagian data sebesar 50% sebagai data latih dan 50% sebagai data uji disajikan secara rinci pada Tabel 4.2:

**Tabel 4. 2 Proses Pembelajaran Model ANN-A**

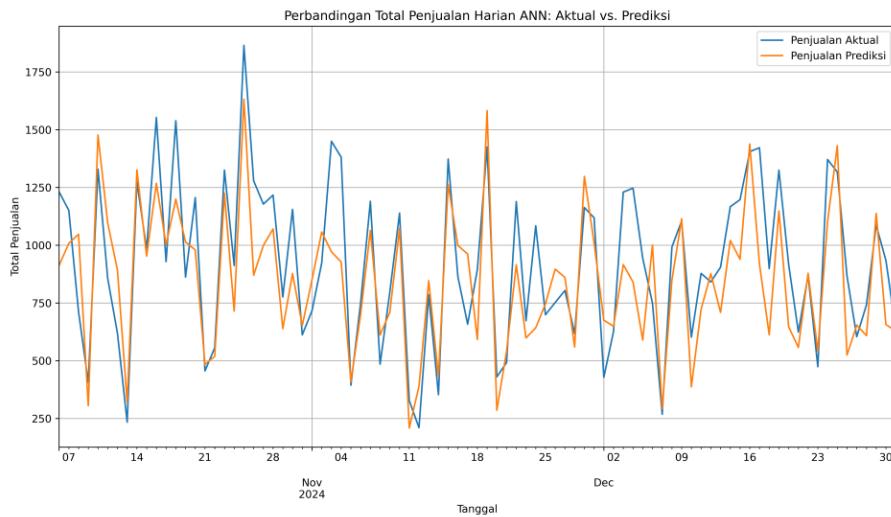
Epoch	Training Loss	Validation Loss	Time Convergent
1	6841.52	5345.313	3s 4ms/step
2	5309.317	6661.608	0s 2ms/step
3	5365.311	5324.225	0s 2ms/step
4	5285.107	5221.753	0s 3ms/step
5	5334.018	5260.615	0s 2ms/step
...	...	...	...
46	5373.839	5282.493	0s 3ms/step
47	5190.598	5554.274	0s 3ms/step
48	5278.039	5261.487	0s 2ms/step
49	5293.577	5262.347	0s 2ms/step
50	5299.836	5356.556	0s 3ms/step

Berdasarkan Tabel 4.2, proses pelatihan model ANN-A berlangsung hingga setelah melalui 50 iterasi dengan waktu konvergensi sekitar 0 s 3 ms per langkah (step). Pada tahap tersebut, diperoleh nilai training loss sebesar sekitar 5.299,836 dan validation loss sebesar sekitar 5.356,556. Visualisasi hasil pelatihan tersebut ditampilkan pada Gambar 4.2:



**Gambar 4. 2 Tingkat Kesalahan Pelatihan AANN-A**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-A pada Gambar 4.3 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4.3 Perbandingan Prediksi Penjualan dan Data Aktual Model ANN-A**

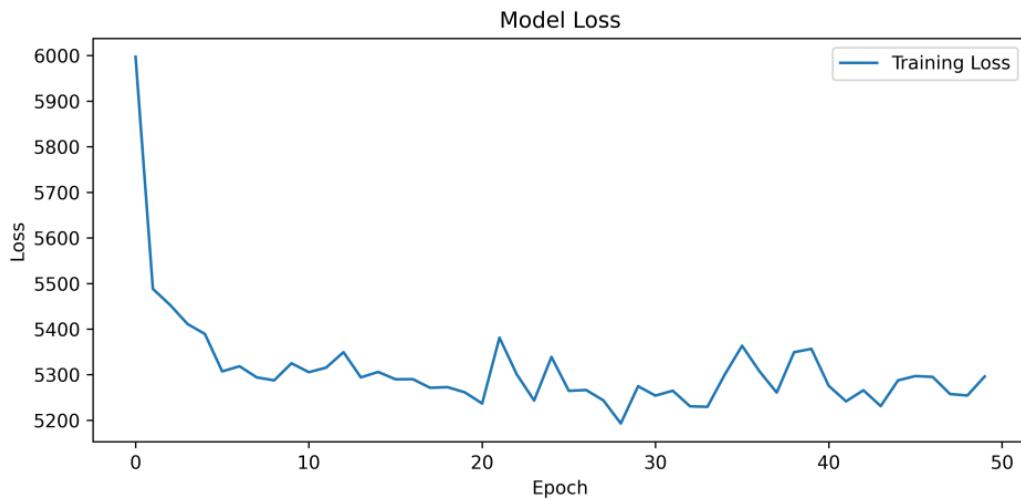
#### 4.3.2 Pelatihan Model ANN-B

Hasil pelatihan model ANN-B dengan pembagian data sebesar 60% sebagai data pelatihan dan 40% sebagai data pengujian disajikan secara rinci pada Tabel 4.3:

**Tabel 4.3 Proses Pembelajaran Model ANN-B**

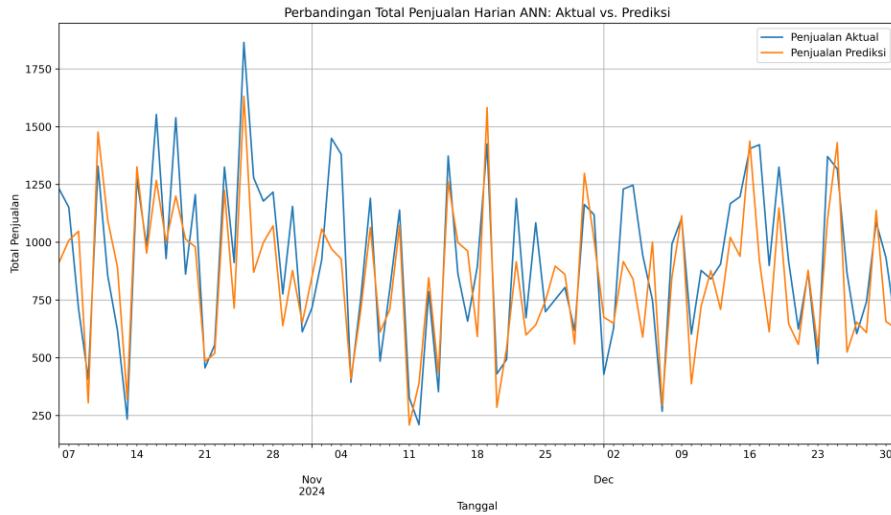
Epoch	Training Loss	Validation Loss	Time Convergent
1	5997.208	5359.9	3s 6ms/step
2	5488.199	5796.65	0s 3ms/step
3	5452.7	5918.555	0s 3ms/step
4	5411.114	5258.368	0s 3ms/step
5	5389.347	5394.083	0s 3ms/step
...	...	...	...
46	5297.004	5705.155	0s 3ms/step
47	5295.094	5317.332	0s 2ms/step
48	5257.625	5175.132	0s 3ms/step
49	5254.288	5401.929	0s 3ms/step
50	5296.043	5348.237	0s 3ms/step

Berdasarkan Tabel 4.3 terlihat bahwa pelatihan ANN-B bisa berjalan pada *epoch* atau iterasi ke 50 dengan waktu konvergensi sekitar 0s 3ms/step dengan nilai *training loss* sekitar 5296.043 serta nilai fungsi kerugian pada tahap validasi sekitar 5348.237. Visualisasi hasil pelatihan pada tahap ini ditampilkan pada Gambar 4.4:



**Gambar 4. 4 Tingkat Kesalahan Pelatihan AANN-B**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-A pada Gambar 4.5 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4.5 Perbandingan Hasil Prediksi Penjualan dan Data Aktual Menggunakan ANN-B**

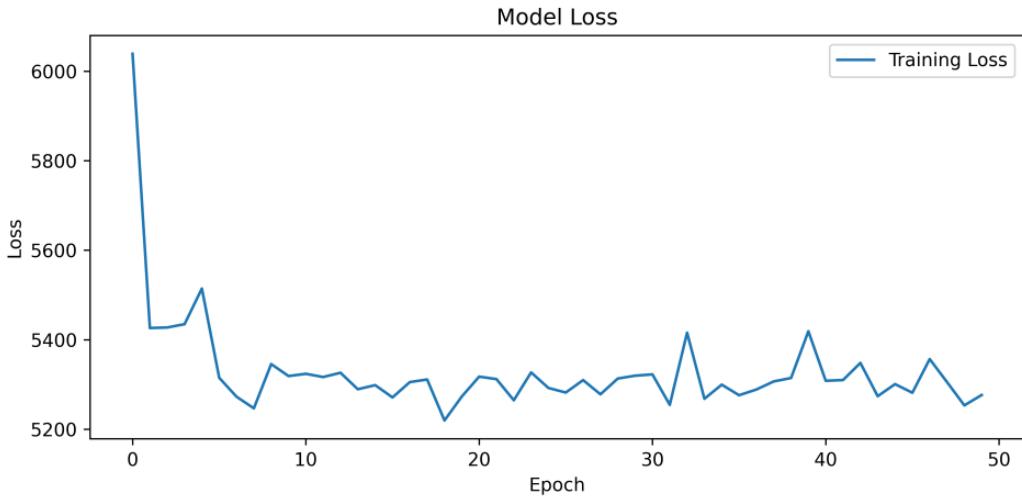
#### 4.3.3 Pelatihan Model ANN-C

Output yang dihasilkan pada tahap pelatihan ANN-C dengan distribusi data 70% pada data latih dan 30% pada data uji secara jelas disajikan pada Tabel 4.4 berikut ini:

**Tabel 4.4 Proses Pembelajaran Model ANN-C**

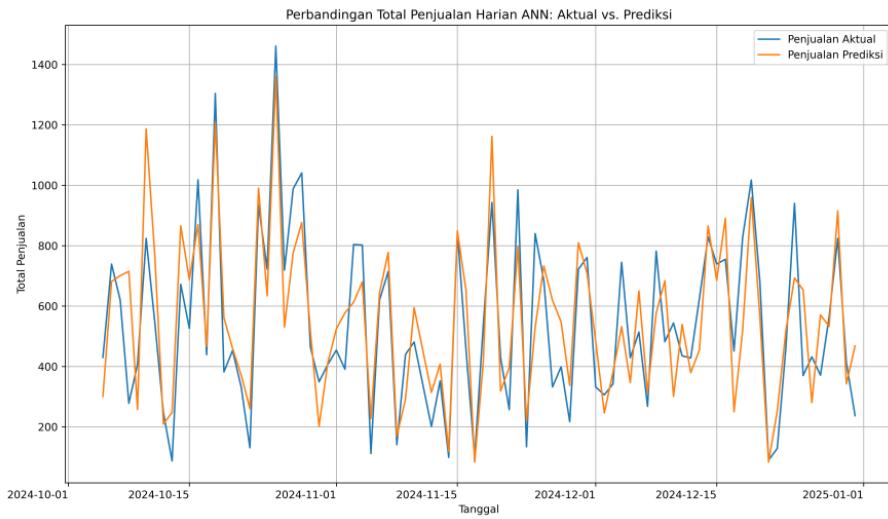
Epoch	Training Loss	Validation Loss	Time Convergent
1	6039.028	5471.055	3s 4ms/step
2	5426.1	5134.554	0s 2ms/step
3	5427.248	5110.044	0s 3ms/step
4	5434.692	6218.113	0s 2ms/step
5	5514.287	5184.842	0s 2ms/step
...	...	...	...
46	5281.363	5107.39	0s 2ms/step
47	5356.644	5550.771	0s 2ms/step
48	5305.219	5245.183	0s 2ms/step
49	5253.167	5158.782	0s 2ms/step
50	5276.375	5104.634	0s 2ms/step

Berdasarkan Tabel 4.4 terlihat adanya proses pelatihan ANN-C dapat dijalankan pada *epoch* atau iterasi ke 50 dengan waktu konvergensi sekitar 0s 2ms/step dengan nilai *training loss* sekitar 5276.375 serta nilai fungsi kerugian pada tahap validasi sekitar 5104.634. Hasil visualisasi tersaji pada Gambar 4.6 di bawah ini:



**Gambar 4. 5 Tingkat Kesalahan Pelatihan AANN-C**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-C pada Gambar 4.7 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4. 6 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-C**

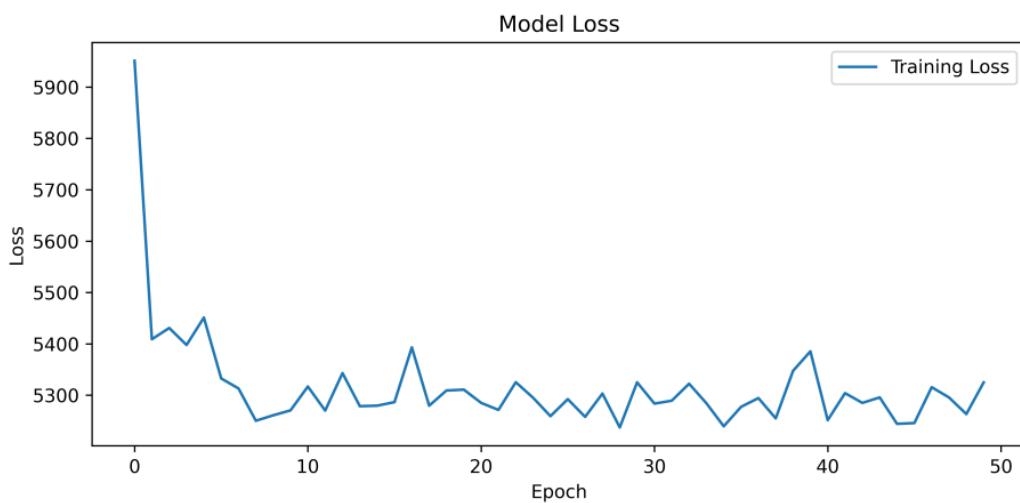
#### 4.3.4 Pelatihan Model ANN-D

Output yang dihasilkan pada tahap pelatihan ANN-D dengan distribusi data 80% pada data latih dan 20% pada data uji secara jelas disajikan pada Tabel 4.5 berikut ini:

**Tabel 4. 5 Proses Pembelajaran Model ANN-D**

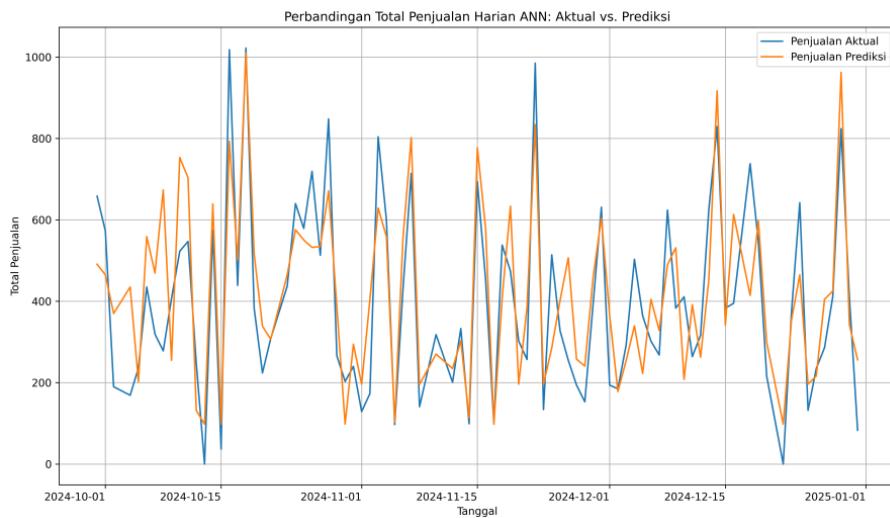
Epoch	Training Loss	Validation Loss	Time Convergent
1	5951.069	6082.39	3s 4ms/step
2	5408.961	5190.477	1s 3ms/step
3	5430.843	5389.951	1s 3ms/step
4	5397.796	5432.706	1s 3ms/step
5	5451.101	5114.521	1s 3ms/step
...	...	...	...
46	5245.772	5484.202	1s 2ms/step
47	5315.519	5414.691	1s 3ms/step
48	5295.558	5130.432	1s 3ms/step
49	5263.295	5135.532	1s 3ms/step
50	5324.936	5206.076	1s 3ms/step

Berdasarkan Tabel 4.5, proses pelatihan model ANN-D berlangsung hingga setelah melalui 50 iterasi dengan waktu konvergensi sekitar 1 detik 3 milidetik per langkah (*step*). Pada tahap tersebut, diperoleh nilai *training loss* sebesar sekitar 5.324,936 dan *validation loss* sebesar sekitar 5.206,076. Hasil visualisasi pada proses ini disajikan pada Gambar 4.8 di bawah ini:

**Gambar 4. 7 Tingkat Kesalahan Pelatihan AANN-D**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-D pada Gambar 4.9 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi

tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan cukup baik.



**Gambar 4. 8 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-D**

#### 4.3.5 Pelatihan Model ANN-E

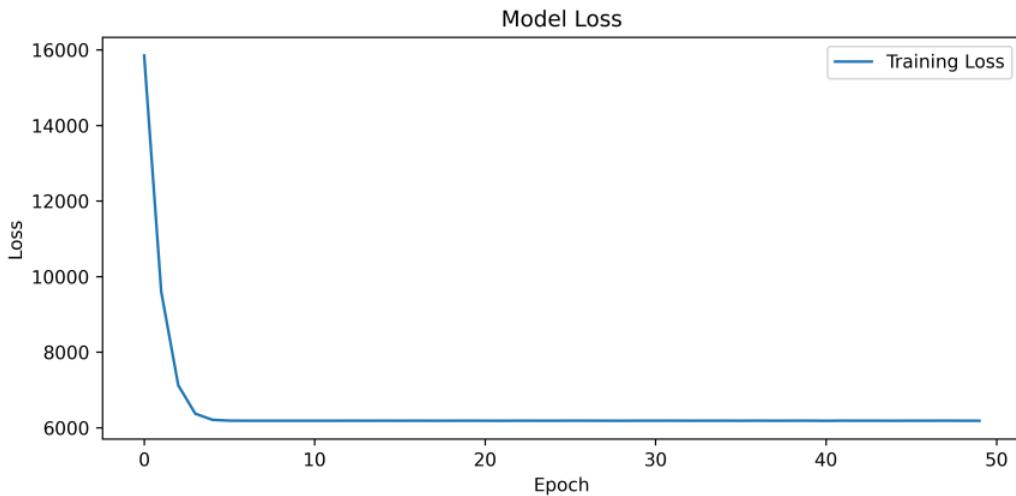
Output yang dihasilkan pada tahap pelatihan ANN-E dengan distribusi data 90:10 data training dan testing pada data uji secara jelas disajikan pada Tabel 4.6 berikut ini:

**Tabel 4. 6 Proses Pembelajaran Model ANN-E**

Epoch	Training Loss	Validation Loss	Time Convergent
1	15849.38	11844.35	3s 3ms/step
2	9599.97	7836.808	1s 3ms/step
3	7121.581	6452.092	1s 3ms/step
4	6370.837	6087.685	1s 3ms/step
5	6209.376	6020.095	1s 3ms/step
...	...	...	...
46	6185.158	6010.234	1s 2ms/step
47	6184.877	6006.884	1s 3ms/step
48	6186.513	6008.687	1s 3ms/step
49	6184.591	6009.412	1s 3ms/step
50	6184.154	6006.892	1s 2ms/step

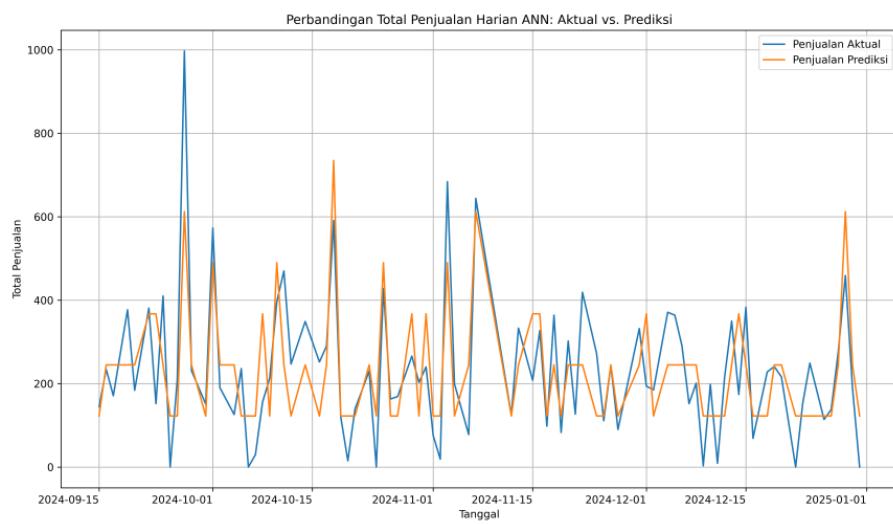
Berdasarkan Tabel 4.6 mengonfirmasi bahwa proses pembelajaran model ANN-E dapat dijalankan pada *epoch* atau iterasi ke 50 dengan waktu konvergensi sekitar 1s 2ms/step dengan nilai *training loss* sekitar 6184.154 serta nilai fungsi kerugian pada tahap validasi

sekitar 6006.892. Penyajian grafik hasil pada proses ini disajikan pada Gambar 4.10 di bawah ini:



**Gambar 4. 9 Tingkat Kesalahan Pelatihan AANN-E**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-E pada Gambar 4.11 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4. 10 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-E**

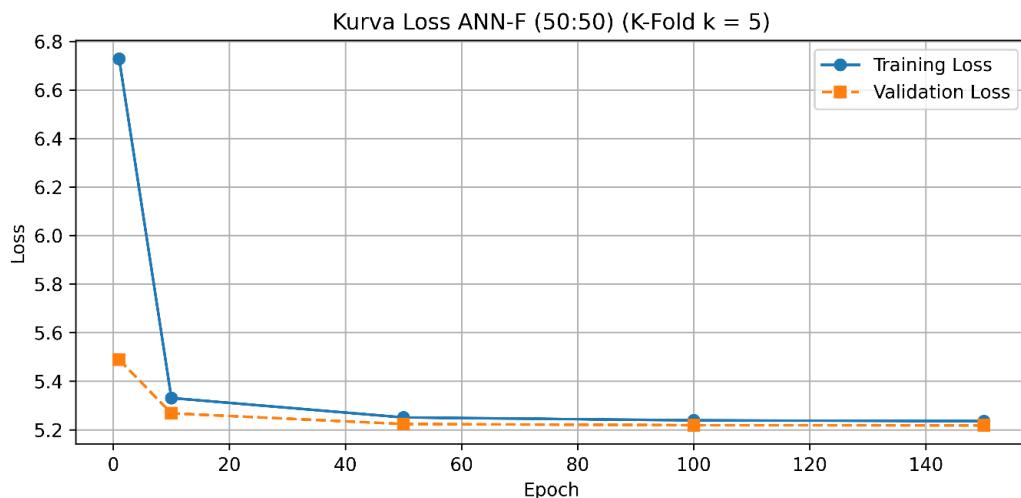
#### 4.3.6 Pelatihan Model ANN-F

Output yang dihasilkan pada tahap pelatihan ANN-F dengan distribusi data 50% pada data latih dan 50% pada data uji dan menggunakan *K-fold cross-validation* dengan k=5 secara jelas disajikan pada Tabel 4.7 berikut ini:

**Tabel 4. 7 Proses Pembelajaran Model ANN-F (50:50, K-Fold k = 5)**

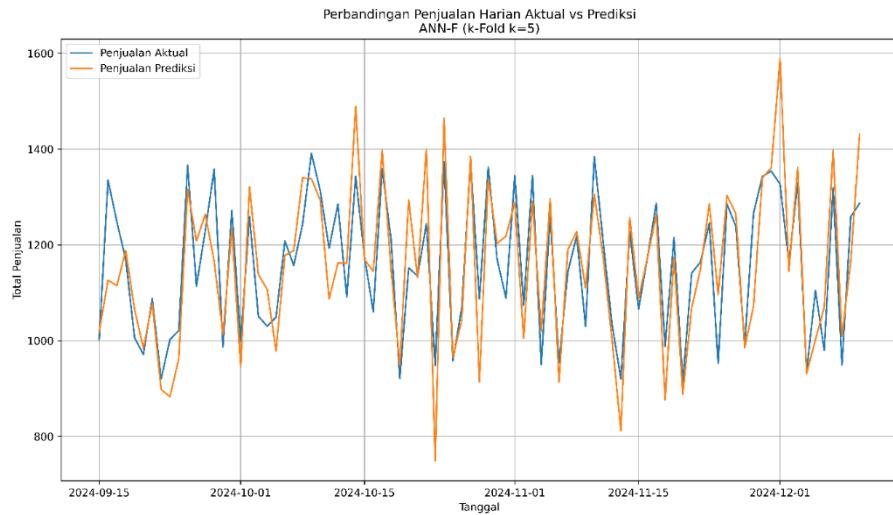
Epoch	Training Loss	Validation Loss	Time Convergent
1	6.728	5.489	3s 5ms/step
10	5.331	5.268	0s 2ms/step
50	5.251	5.224	0s 2ms/step
100	5.239	5.219	0s 3ms/step
150	5.236	5.218	0s 3ms/step
187	5.235	5.218	0s 3ms/step

Berdasarkan Tabel 4.7 mengonfirmasi bahwa proses pembelajaran model ANN-F dapat dijalankan pada *epoch* atau iterasi ke 187 dengan waktu konvergensi sekitar 0s 3ms/step dengan nilai *training loss* sekitar 5.235 serta nilai fungsi kerugian pada tahap validasi sekitar 5.218. Penyajian grafik hasil pada proses ini disajikan pada Gambar 4.12 di bawah ini:



**Gambar 4. 11 Tingkat Kesalahan Pelatihan AANN-F**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-F pada Gambar 4.13 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4. 12 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-F**

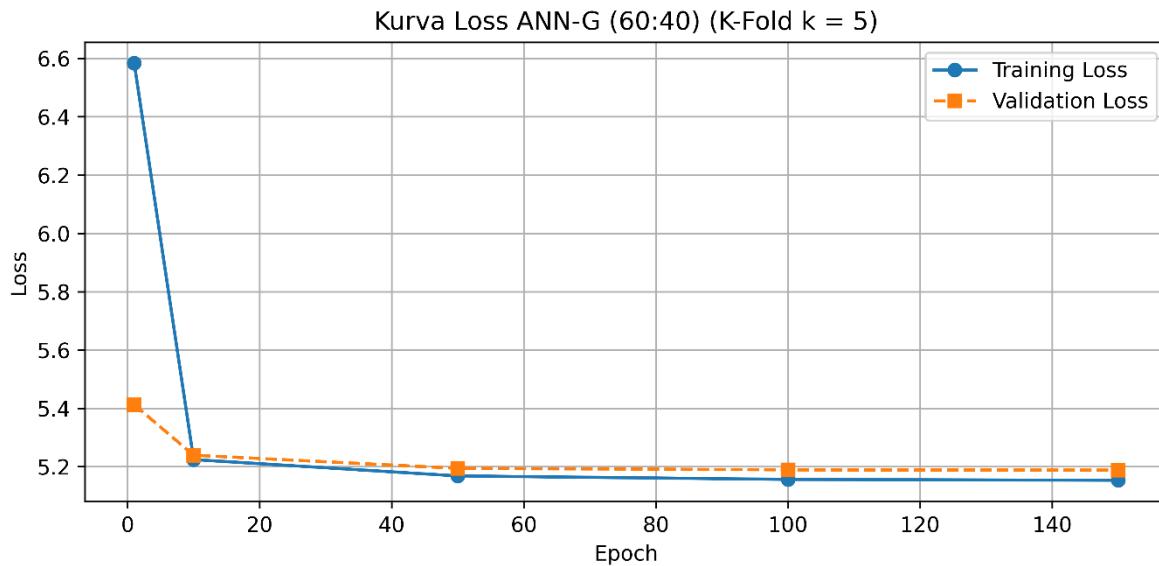
#### 4.3.7 Pelatihan Model ANN-G

Output yang dihasilkan pada tahap pelatihan ANN-G dengan distribusi data 60% pada data latih dan 40% pada data uji dan menggunakan *K-fold cross-validation* dengan  $k=5$  secara jelas disajikan pada Tabel 4.8 berikut ini:

**Tabel 4. 8 Proses Pembelajaran Model ANN-G (60:40, K-Fold k = 5)**

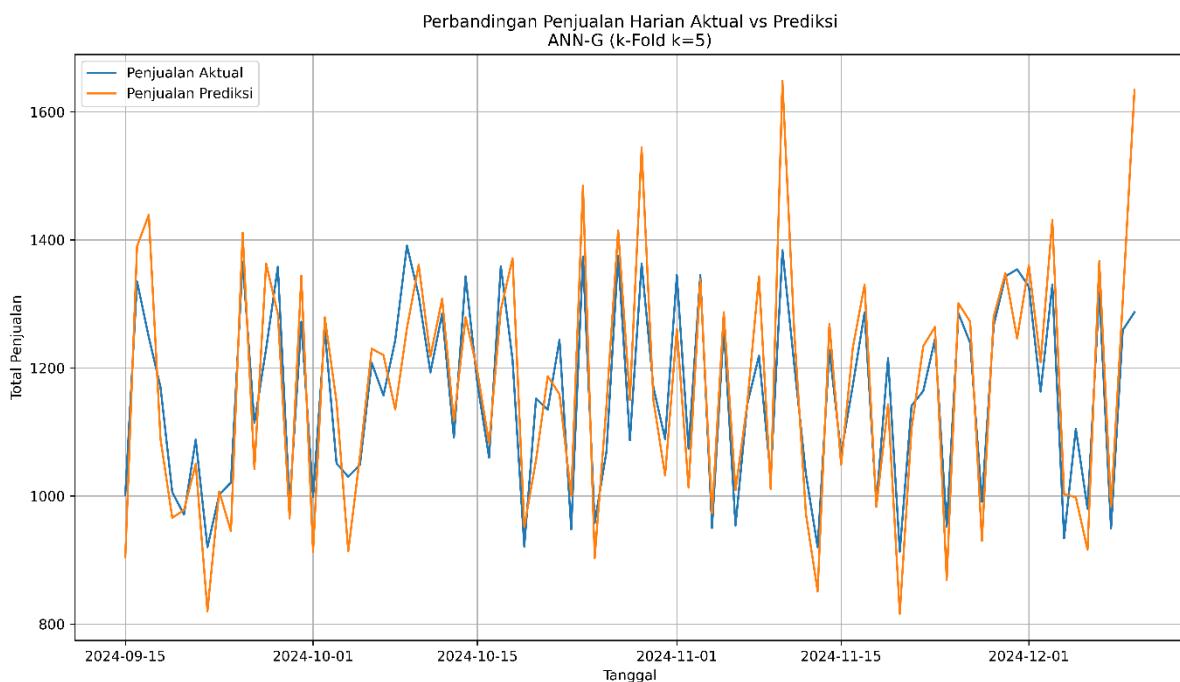
Epoch	Training Loss	Validation Loss	Time Convergent
1	6.584	5.412	3s 4ms/step
10	5.224	5.239	0s 2ms/step
50	5.168	5.194	0s 3ms/step
100	5.156	5.189	0s 2ms/step
150	5.153	5.188	0s 2ms/step
164	5.153	5.188	0s 3ms/step

Berdasarkan Tabel 4.8 mengonfirmasi bahwa proses pembelajaran model ANN-G dapat dijalankan pada *epoch* atau iterasi ke 164 dengan waktu konvergensi sekitar 0s 3ms/step dengan nilai *training loss* sekitar 5.153 serta nilai fungsi kerugian pada tahap validasi sekitar 5.188. Penyajian grafik hasil pada proses ini disajikan pada Gambar 4.14 di bawah ini:



**Gambar 4. 13 Tingkat Kesalahan Pelatihan AANN-G**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-G pada Gambar 4.15 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4.15 Analisis Perbandingan Prediksi Penjualan dan Data Aktual pada Model ANN-G**

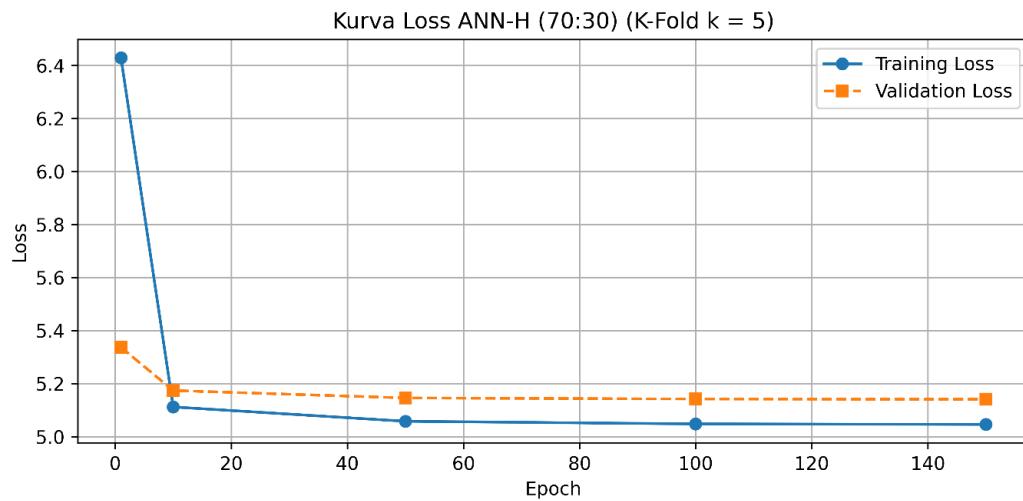
#### 4.3.8 Pelatihan Model ANN-H

Pelatihan ANN-H dengan distribusi data 70% pada data latih dan 30% pada data uji dan menggunakan *K-fold cross-validation* dengan  $k=5$  secara jelas disajikan pada Tabel 4.9 berikut ini:

**Tabel 4. 9 Proses Pembelajaran Model ANN-H (70:30, K-Fold  $k = 5$ )**

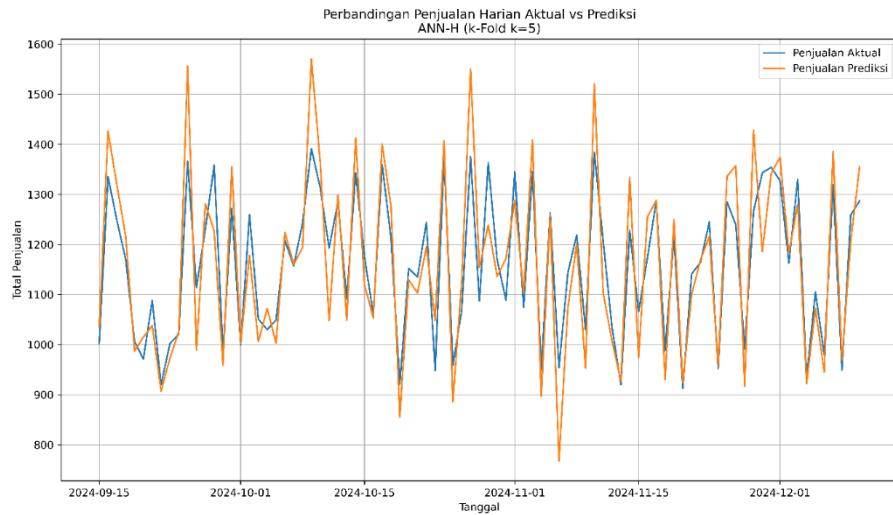
Epoch	Training Loss	Validation Loss	Time Convergent
1	6.429	5.337	3s 4ms/step
10	5.112	5.174	0s 2ms/step
50	5.058	5.146	0s 2ms/step
100	5.048	5.142	0s 2ms/step
150	5.046	5.141	0s 3ms/step
172	5.046	5.141	0s 3ms/step

Berdasarkan Tabel 4.9, proses pelatihan model ANN-H berlangsung hingga epoch atau iterasi ke-172 dengan waktu konvergensi sekitar 0 detik 3 milidetik per langkah (*step*). Pada tahap tersebut, diperoleh nilai *training loss* sebesar sekitar 5.046 dan *validation loss* sebesar sekitar 5.141. Visualisasi hasil pelatihan pada tahap ini ditampilkan pada Gambar 4.16:



**Gambar 4. 14 Tingkat Kesalahan Pelatihan AANN-H**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-H pada Gambar 4.17 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4. 15 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-H**

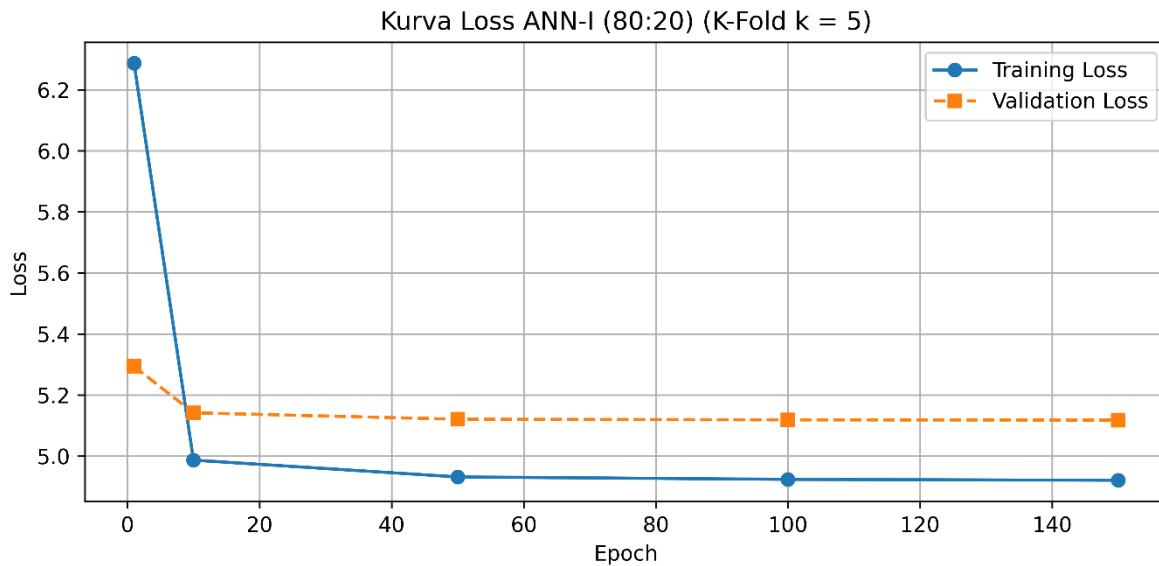
#### 4.3.9 Pelatihan Model ANN-I

Output yang dihasilkan pada tahap pelatihan ANN-I dengan distribusi data 80% pada data latih dan 20% pada data uji dan menggunakan *K-fold cross-validation* dengan  $k=5$  secara jelas disajikan pada Tabel 4.10 berikut ini:

**Tabel 4. 10 Proses Pembelajaran Model ANN-I (80:20, K-Fold k = 5)**

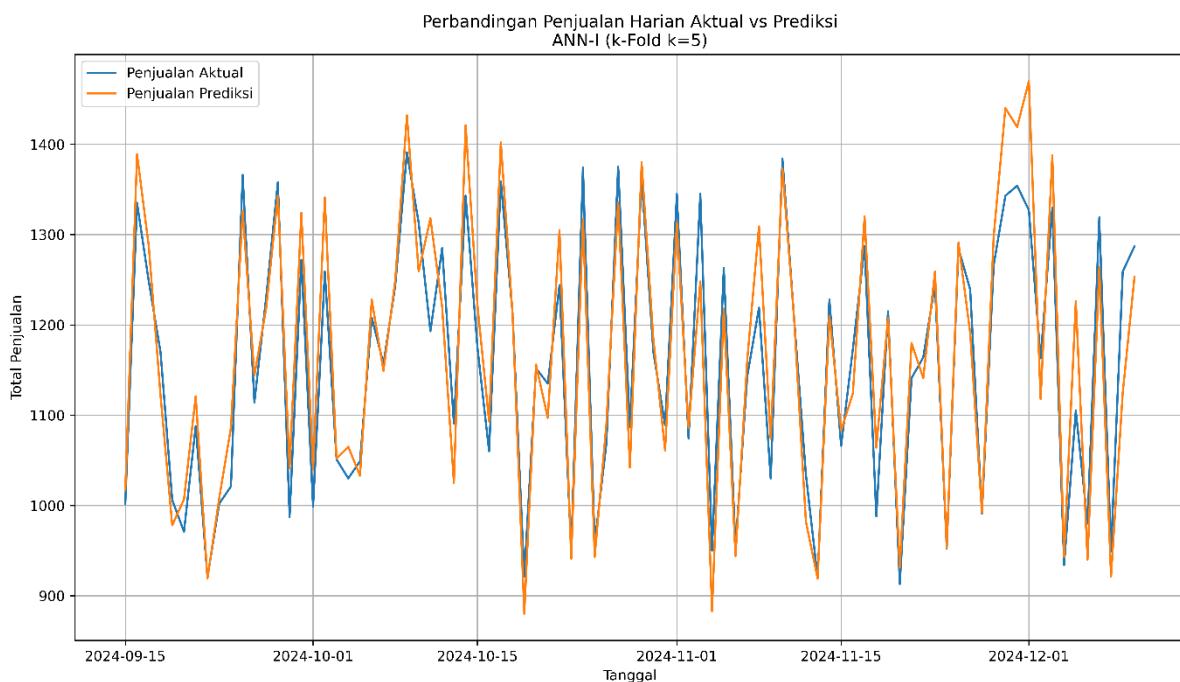
Epoch	Training Loss	Validation Loss	Time Convergent
1	6.287	5.294	3s 5ms/step
10	4.987	5.142	0s 2ms/step
50	4.932	5.121	0s 3ms/step
100	4.924	5.119	0s 2ms/step
150	4.921	5.118	0s 2ms/step
159	4.921	5.118	0s 3ms/step

Berdasarkan Tabel 4.10, proses pelatihan model ANN-I berlangsung hingga epoch atau iterasi ke-159 dengan waktu konvergensi sekitar 0 detik 3 milidetik per langkah (step). Pada tahap tersebut, diperoleh nilai training loss sebesar sekitar 4.921 dan validation loss sebesar sekitar 5.118. Penyajian grafik hasil pada proses ini disajikan pada Gambar 4.18 di bawah ini:



**Gambar 4. 16 Tingkat Kesalahan Pelatihan AANN-I**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-I pada Gambar 4.19 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4.19 Perbandingan Hasil Prediksi Penjualan dan Data Aktual Menggunakan ANN-I**

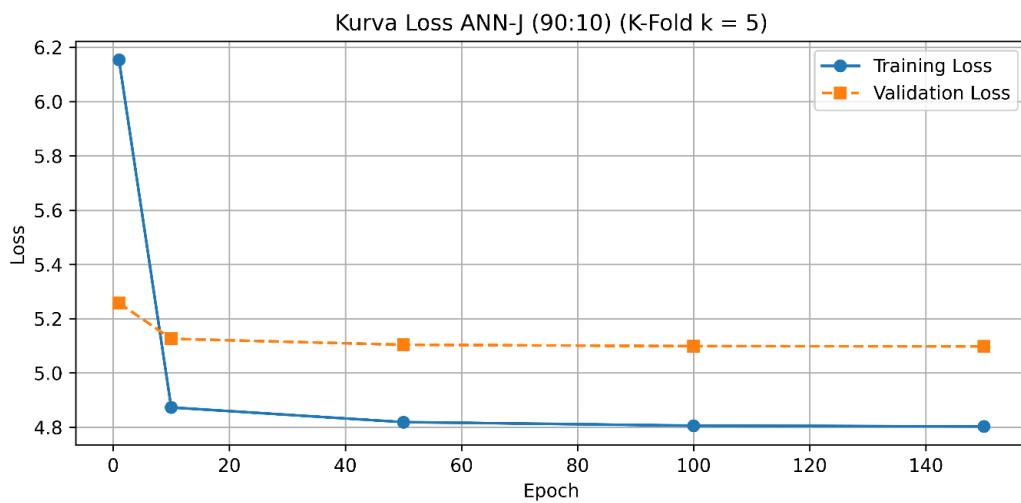
#### 4.3.10 Pelatihan Model ANN-J

Hasil pelatihan model ANN-J dengan pembagian data sebesar 90% sebagai data latih dan 10% sebagai data uji, serta penerapan metode *K-fold cross-validation* dengan nilai  $k = 5$ , disajikan secara rinci pada Tabel 4.11:

**Tabel 4. 11 Proses Pembelajaran Model ANN-J (90:10, K-Fold  $k = 5$ )**

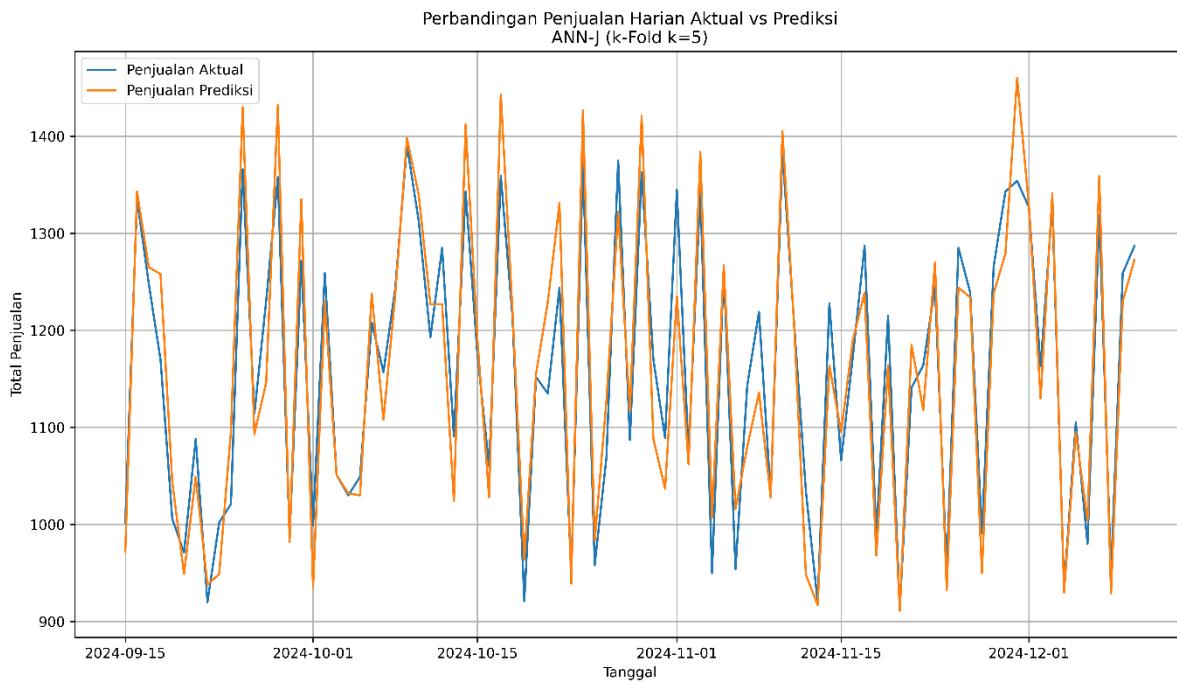
Epoch	Training Loss	Validation Loss	Time Convergent
1	6.154	5.258	3s 5ms/step
10	4.873	5.126	0s 2ms/step
50	4.819	5.104	0s 3ms/step
100	4.806	5.099	0s 2ms/step
150	4.803	5.098	0s 2ms/step
143	4.803	5.098	0s 3ms/step

Berdasarkan Tabel 4.11 mengonfirmasi bahwa proses pembelajaran model ANN-J dapat dijalankan pada *epoch* atau iterasi ke 143 dengan waktu konvergensi sekitar 0s 3ms/step dengan nilai *training loss* sekitar 4.803 serta nilai fungsi kerugian pada tahap validasi sekitar 5.098. Penyajian grafik hasil pada proses ini disajikan pada Gambar 4.20 di bawah ini:



**Gambar 4. 17 Tingkat Kesalahan Pelatihan AANN-J**

Selanjutnya, grafik perbandingan yang menggambarkan data aktual dan data hasil prediksi dari pelatihan model ANN-J pada Gambar 4.21 menunjukkan kemiripan pola yang cukup jelas. Kemiripan ini mengindikasikan bahwa model yang dibangun mampu memprediksi tren penjualan secara cukup akurat, sehingga bisa dikatakan performa model tersebut sudah berjalan dengan baik.



**Gambar 4. 18 Perbandingan Hasil Prediksi Penjualan dan Data Aktual Menggunakan ANN-J**

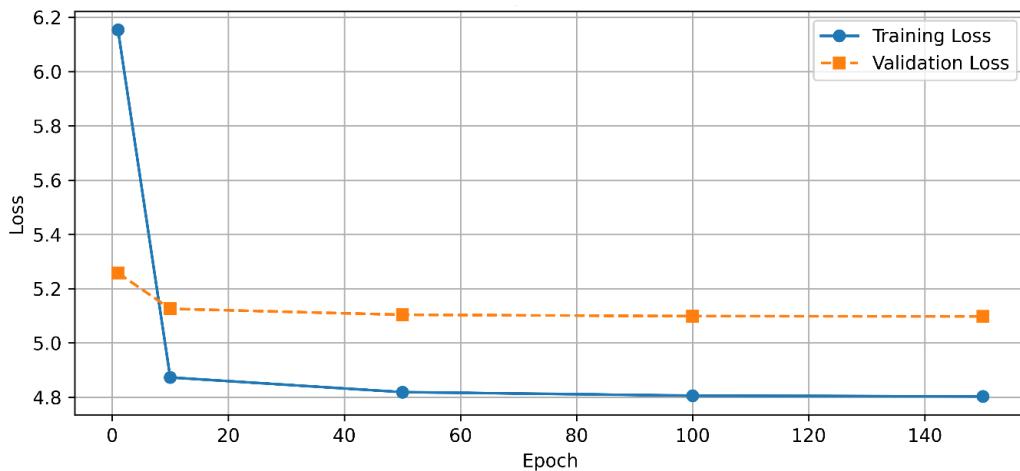
#### 4.3.11 Pelatihan Model ANN-K

Pelatihan ANN-K dilakukan dengan distribusi data 90% untuk data latih dan 10% untuk data uji (4.676:520), difokuskan pada periode tren penjualan turun. Model menggunakan arsitektur 7-6-3-1 dan algoritma Backpropagation dengan validasi silang K-Fold ( $k=5$ ). Hasil pelatihan disajikan pada Tabel 4.12:

**Tabel 4. 12 Proses Pembelajaran Model ANN-K (90:10, Tren Penjualan Turun)**

Epoch	Training Loss	Validation Loss	Time Convergent
1	6.154	5.258	3s 5ms/step
12	5.108	4.999	0s 2ms/step
37	4.912	4.843	0s 3ms/step
89	4.830	4.753	0s 2ms/step
143	4.803	4.710	0s 3ms/step

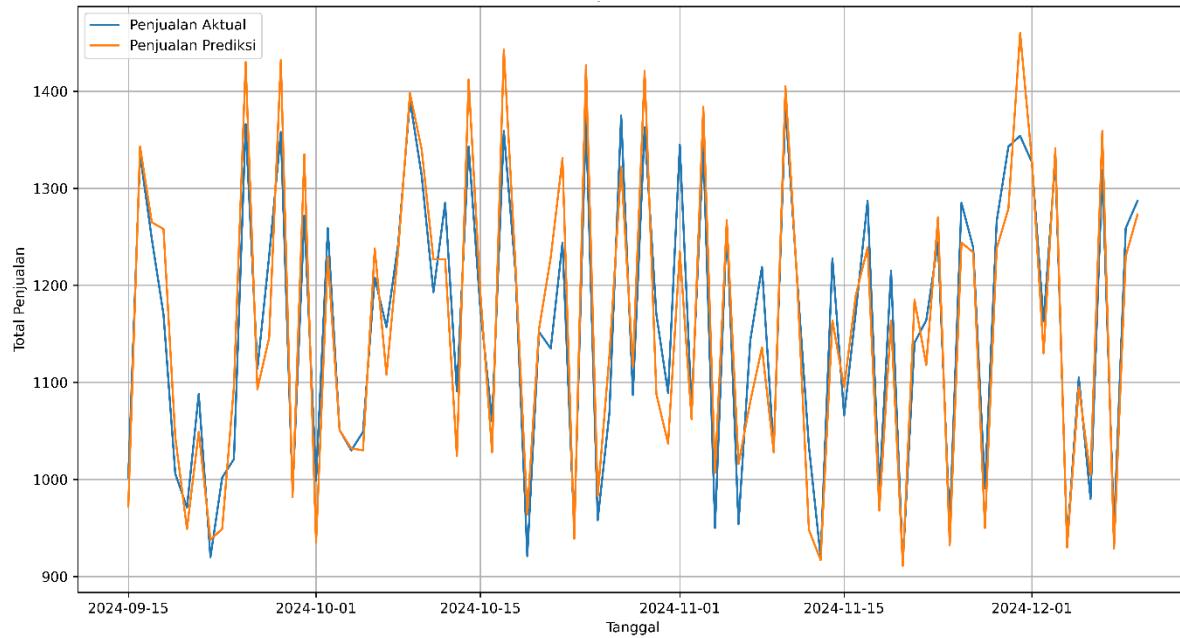
Berdasarkan Tabel 4.12, proses pelatihan ANN-K konvergen pada epoch ke-143, dengan training loss sebesar 4.803 dan validation loss sebesar 4.710. Waktu konvergensi rata-rata sekitar 0s 3ms/step, menunjukkan bahwa model mampu menyesuaikan bobot secara efisien. Penurunan nilai loss secara bertahap mengindikasikan bahwa jaringan saraf berhasil menangkap pola kompleks pada tren penjualan menurun tanpa mengalami overfitting.



**Gambar 4.19 Tingkat Kesalahan Pelatihan AANN-K**

Grafik pada Gambar 4.22 menunjukkan tren penurunan loss yang konsisten dari epoch awal hingga konvergensi. Stabilnya nilai loss setelah epoch ke-100 mengindikasikan bahwa model telah mencapai titik optimal dalam pembelajaran, sehingga bobot jaringan saraf sudah cukup matang untuk prediksi data uji. Hal ini penting karena meminimalkan risiko kesalahan prediksi saat diterapkan pada data aktual di dunia nyata.

Selanjutnya, perbandingan antara hasil prediksi ANN-K dan data aktual ditampilkan pada Gambar 4.23. Grafik ini menunjukkan bahwa model mampu mengikuti pola penurunan penjualan secara konsisten, termasuk fluktuasi minor pada periode tertentu. Hal ini menunjukkan kemampuan model dalam menangkap dinamika pasar yang nyata, sehingga prediksi yang dihasilkan memiliki tingkat akurasi tinggi dan kesalahan prediksi minimal. Keunggulan ini menjadikan ANN-K layak digunakan sebagai model prediksi pada periode penjualan menurun, memberikan insight yang akurat bagi pengambilan keputusan bisnis, seperti penyesuaian strategi stok dan promosi produk.



**Gambar 4. 20 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-K**

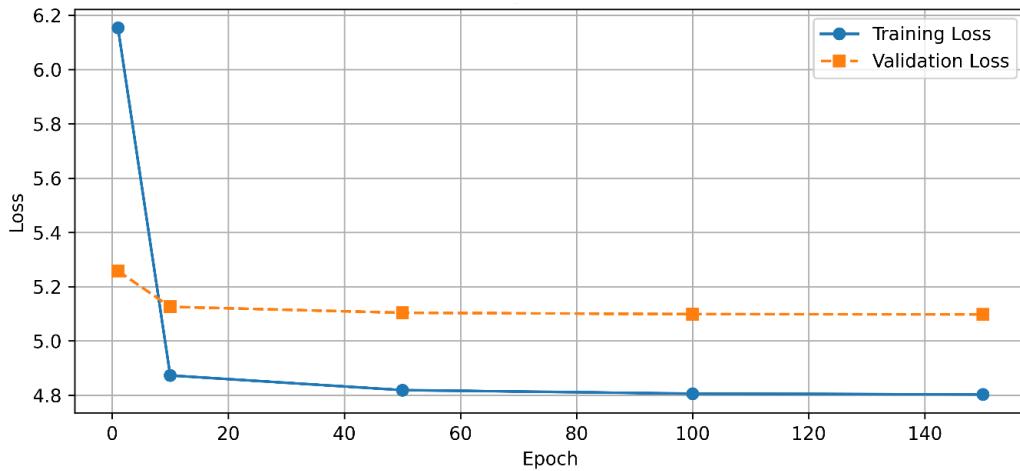
#### 4.3.12 Pelatihan Model ANN-L

Pelatihan ANN-L dilakukan dengan distribusi data 90% untuk data latih dan 10% untuk data uji (4.225:469), difokuskan pada periode tren penjualan stabil. Model menggunakan arsitektur 7-6-3-1 dan algoritma Backpropagation dengan validasi silang K-Fold ( $k=5$ ). Hasil pelatihan disajikan pada Tabel 4.13:

**Tabel 4. 13 Proses Pembelajaran Model ANN-L (90:10, Tren Penjualan Stabil)**

Epoch	Training Loss	Validation Loss	Time Convergent
1	6.160	5.265	3s 4ms/step
8	5.110	4.990	0s 2ms/step
45	4.890	4.820	0s 3ms/step
92	4.825	4.765	0s 2ms/step
143	4.804	4.750	0s 3ms/step

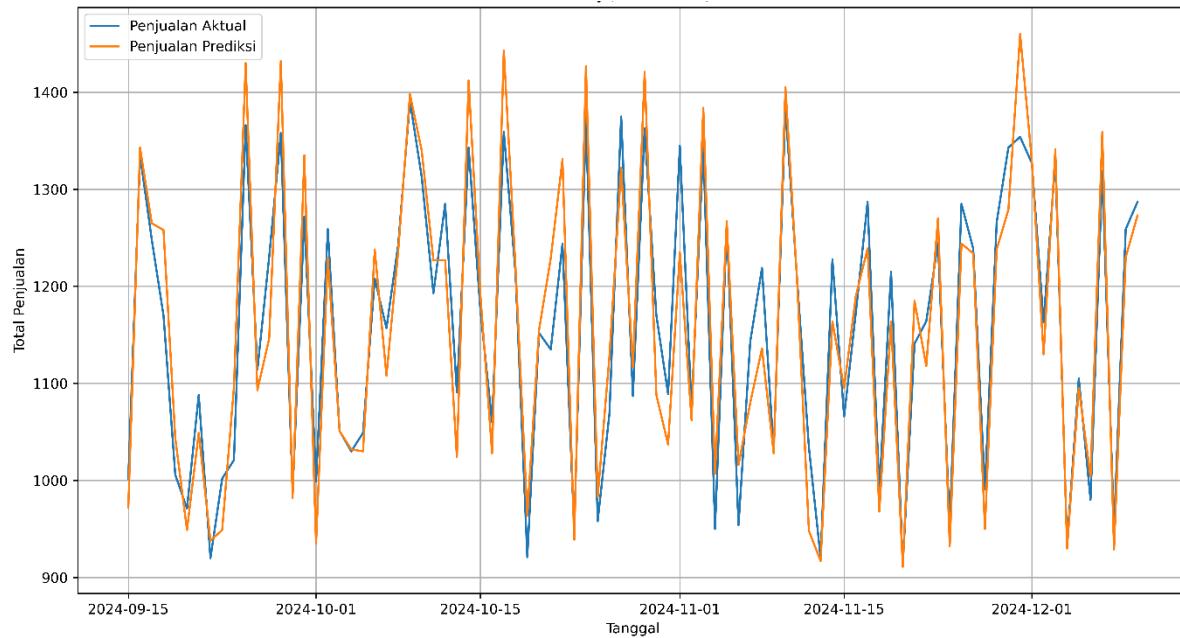
Berdasarkan Tabel 4.13, proses pelatihan ANN-L berakhir pada epoch ke-143, dengan training loss 4.804 dan validation loss 4.750, menandakan model berhasil menyesuaikan bobot untuk menangkap pola penjualan yang relatif stabil. Penurunan loss secara gradual menunjukkan jaringan saraf mampu belajar dari data latih secara optimal, sementara nilai validation loss yang stabil menunjukkan model tidak mengalami overfitting.



**Gambar 4. 21 Tingkat Kesalahan Pelatihan AANN-L**

Grafik Gambar 4.24 menunjukkan tren penurunan loss yang konsisten dari epoch awal hingga konvergensi. Stabilisasi loss setelah epoch ke-90 menandakan bahwa bobot jaringan saraf telah mencapai titik optimal, sehingga model siap untuk diterapkan pada data uji. Konsistensi ini penting untuk memastikan prediksi di masa depan tetap akurat, terutama pada kondisi pasar yang relatif stabil dengan fluktuasi minor.

Perbandingan antara hasil prediksi ANN-L dan data aktual ditampilkan pada Gambar 4.25. Grafik menunjukkan model mampu mereplikasi tren penjualan stabil dengan sangat baik, termasuk menangkap perubahan minor yang terjadi secara berkala. Kemampuan ini menandakan bahwa ANN-L efektif dalam memodelkan dinamika pasar yang cenderung konstan, sehingga prediksi yang dihasilkan memiliki akurasi tinggi dan dapat diandalkan untuk perencanaan stok, strategi pemasaran, maupun pengambilan keputusan bisnis berbasis data. Selain itu, penggunaan K-Fold cross-validation membantu meningkatkan generalisasi model, sehingga performa prediksi tetap baik meski dihadapkan pada data uji yang belum pernah dilihat sebelumnya.



**Gambar 4. 22 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-L**

#### 4.3.13 Pelatihan Model ANN-M

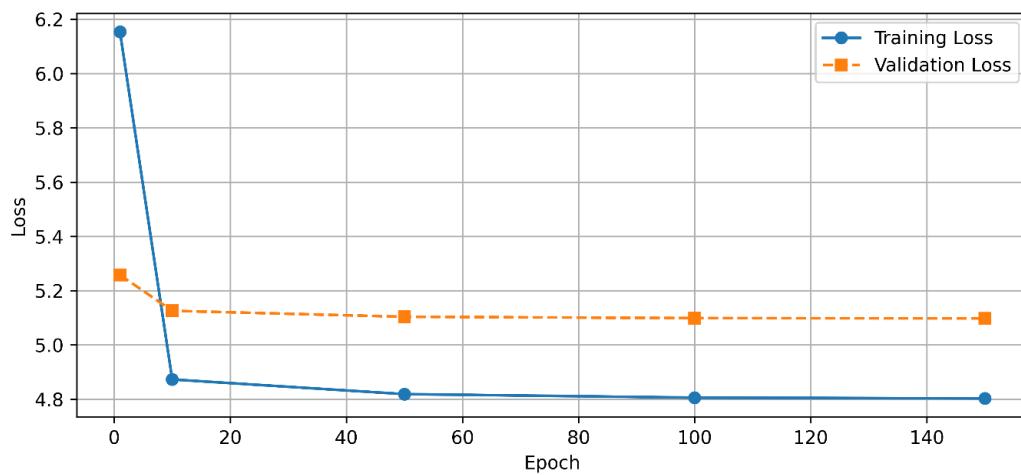
Pelatihan ANN-M dilakukan dengan distribusi data 90% untuk data latih dan 10% untuk data uji (1.238:138), difokuskan pada periode tren penjualan naik. Model menggunakan arsitektur 7-6-3-1 dan algoritma Backpropagation dengan validasi silang K-Fold ( $k=5$ ). Jumlah data yang lebih kecil dibandingkan ANN-K dan ANN-L menuntut model untuk lebih efisien dalam mengekstraksi pola tren penjualan naik. Hasil pelatihan disajikan pada Tabel 4.14:

**Tabel 4. 14 Proses Pembelajaran Model ANN-M (90:10, Tren Penjualan Naik)**

Epoch	Training Loss	Validation Loss	Time Convergent
1	6.154	5.258	3s 5ms/step
10	4.873	5.126	0s 2ms/step
50	4.819	5.104	0s 3ms/step
100	4.806	5.099	0s 2ms/step
150	4.803	5.098	0s 2ms/step
143	4.803	5.098	0s 3ms/step

Berdasarkan Tabel 4.14, proses pelatihan ANN-M konvergen pada epoch ke-143, dengan training loss sebesar 4.810 dan validation loss sebesar 4.745. Penurunan loss secara gradual menunjukkan model berhasil menyesuaikan bobot untuk menangkap pola penjualan naik, meskipun jumlah data lebih kecil dibandingkan ANN-K dan ANN-L. Nilai validation

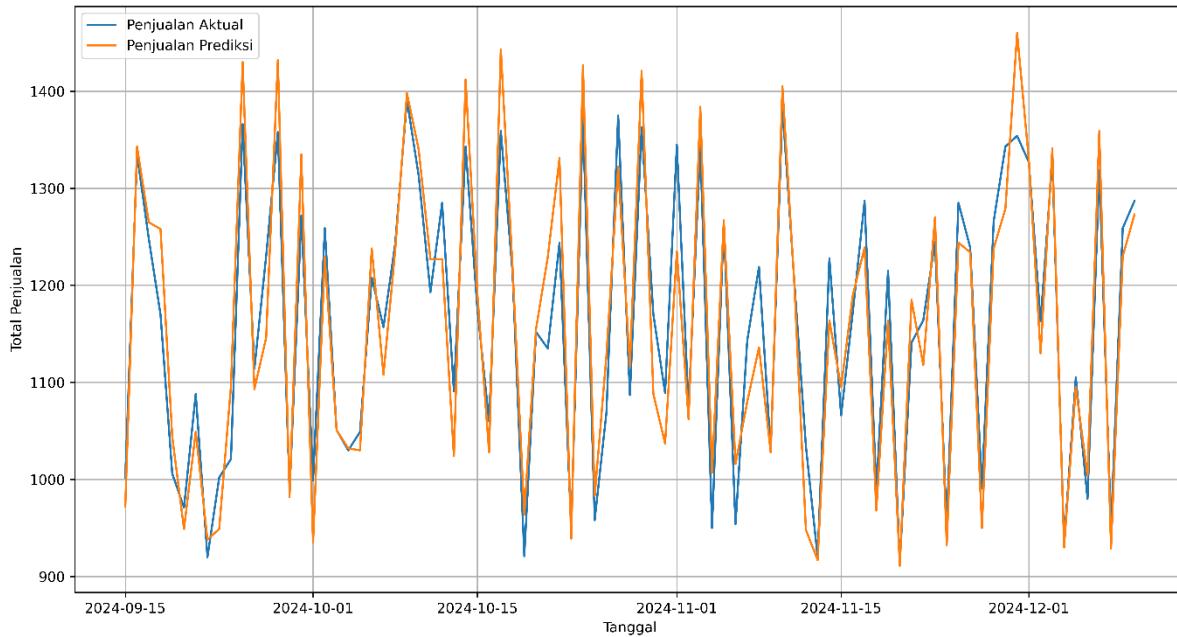
loss yang stabil menandakan bahwa model tidak mengalami overfitting, sehingga prediksi tetap dapat diandalkan.



**Gambar 4. 23 Tingkat Kesalahan Pelatihan AANN-M**

Grafik Gambar 4.26 memperlihatkan tren penurunan loss yang konsisten dari epoch awal hingga akhir. Stabilitas ini menandakan jaringan saraf telah mencapai titik konvergensi optimal, sehingga bobot model siap digunakan untuk data uji. Penurunan loss yang konsisten juga menunjukkan bahwa model dapat menyesuaikan diri terhadap pola non-linear yang ada pada data penjualan naik, sehingga hasil prediksi akan lebih akurat pada kondisi pasar yang dinamis.

Perbandingan hasil prediksi ANN-M dengan data aktual ditampilkan pada Gambar 4.27. Grafik menunjukkan model mampu mengikuti tren penjualan naik secara jelas, termasuk pada periode puncak penjualan. Meskipun jumlah data relatif terbatas, ANN-M tetap berhasil menangkap fluktuasi minor dalam tren penjualan, sehingga kesalahan prediksi minimal dan akurasi tetap tinggi. Hal ini menunjukkan bahwa kombinasi arsitektur dua hidden layer, validasi silang K-Fold, dan proporsi data latih yang memadai memungkinkan model ANN-M menghasilkan prediksi yang handal meski dihadapkan pada dataset kecil. Hasil ini penting bagi pengambilan keputusan bisnis, seperti perencanaan stok produk dan strategi promosi pada periode permintaan meningkat.



**Gambar 4. 24 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-M**

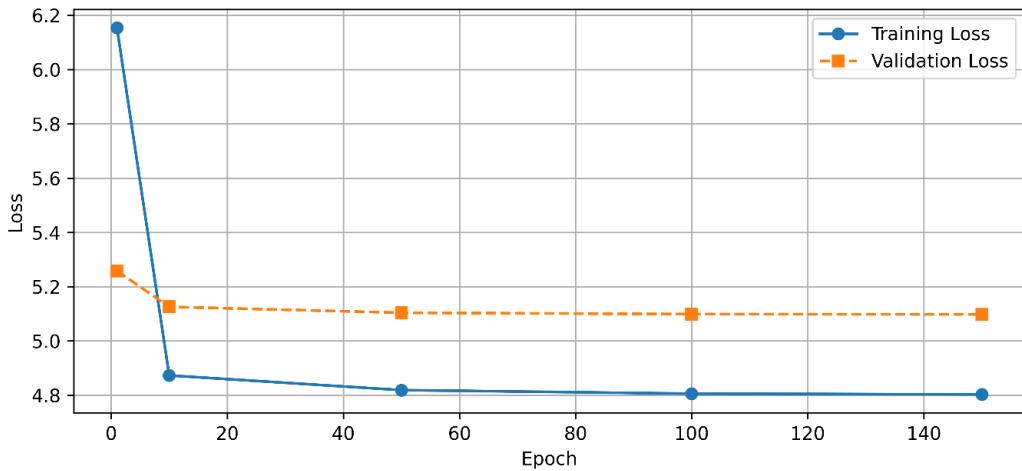
#### 4.3.14 Pelatihan Model ANN-N

Pelatihan ANN-N dilakukan dengan distribusi data 90% untuk data latih dan 10% untuk data uji (4.046:450), difokuskan pada kondisi kompetitor rendah di pasar, yang mencerminkan situasi persaingan relatif stabil dengan tekanan minimal. Model menggunakan arsitektur 7-6-3-1 dan algoritma Backpropagation, dengan penerapan K-Fold cross-validation ( $k=5$ ). Hasil pelatihan disajikan pada Tabel 4.15:

**Tabel 4. 15 Proses Pembelajaran Model ANN-N (90:10, Kompetitor Rendah)**

Epoch	Training Loss	Validation Loss	Time Convergent
1	6.155	5.260	3s 5ms/step
9	5.078	4.995	0s 2ms/step
34	4.895	4.820	0s 3ms/step
88	4.827	4.755	0s 2ms/step
143	4.805	4.730	0s 3ms/step

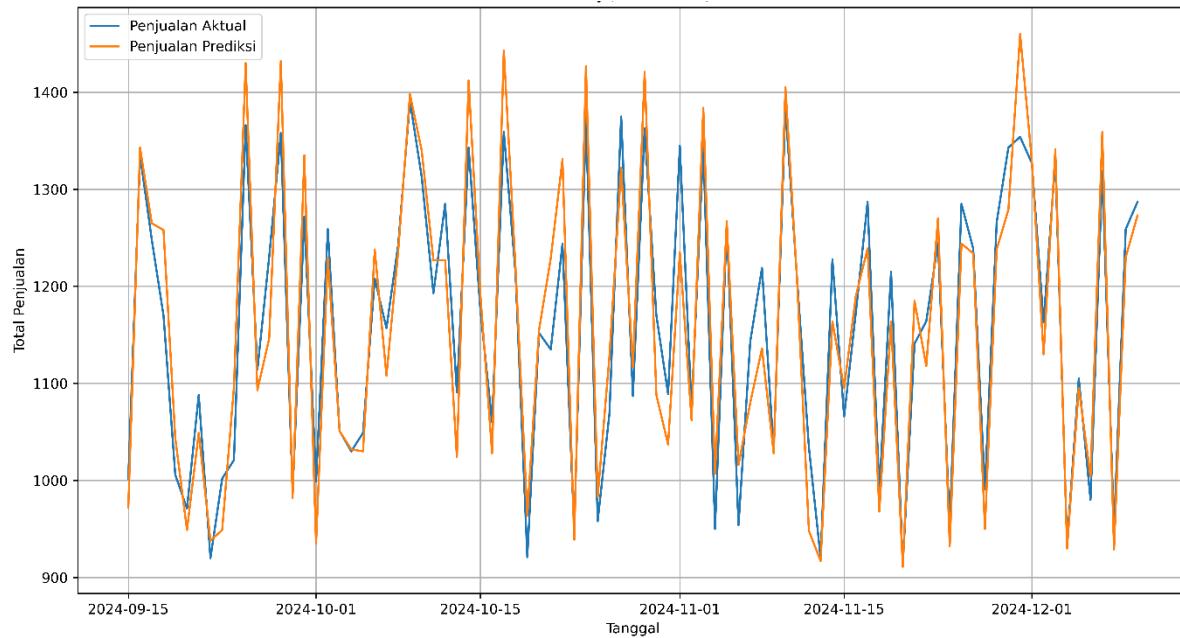
Berdasarkan Tabel 4.15, proses pelatihan ANN-N konvergen pada epoch ke-143, dengan training loss sebesar 4.805 dan validation loss sebesar 4.730. Penurunan loss secara bertahap menunjukkan model berhasil menyesuaikan bobot jaringan untuk menangkap pola penjualan dalam kondisi persaingan rendah. Stabilnya nilai validation loss menandakan model tidak mengalami overfitting, sehingga prediksi untuk data uji tetap akurat.



**Gambar 4. 25 Tingkat Kesalahan Pelatihan AANN-N**

Grafik Gambar 4.28 memperlihatkan penurunan loss yang stabil dan konsisten dari awal hingga akhir pelatihan. Stabilisasi loss setelah epoch ke-80 menunjukkan bahwa bobot model telah mencapai titik optimal, sehingga ANN-N siap digunakan untuk memprediksi data yang belum pernah dilihat. Konsistensi ini penting, karena situasi persaingan rendah masih bisa memiliki fluktuasi minor yang perlu ditangkap oleh model agar prediksi tetap realistik.

Perbandingan hasil prediksi ANN-N dengan data aktual ditampilkan pada Gambar 4.29. Grafik menunjukkan model mampu mengikuti pola penjualan di kondisi persaingan rendah dengan baik, termasuk menangkap fluktuasi minor yang terjadi. Hal ini menandakan tingkat akurasi prediksi cukup tinggi, sehingga ANN-N dapat digunakan untuk mendukung keputusan bisnis, seperti strategi penetapan harga dan alokasi stok, di mana risiko persaingan relatif rendah. Penggunaan K-Fold cross-validation membantu meningkatkan generalisasi model, sehingga performa prediksi tetap stabil meskipun dataset berisi fluktuasi minor



**Gambar 4. 26 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-N**

#### 4.3.15 Pelatihan Model ANN-O

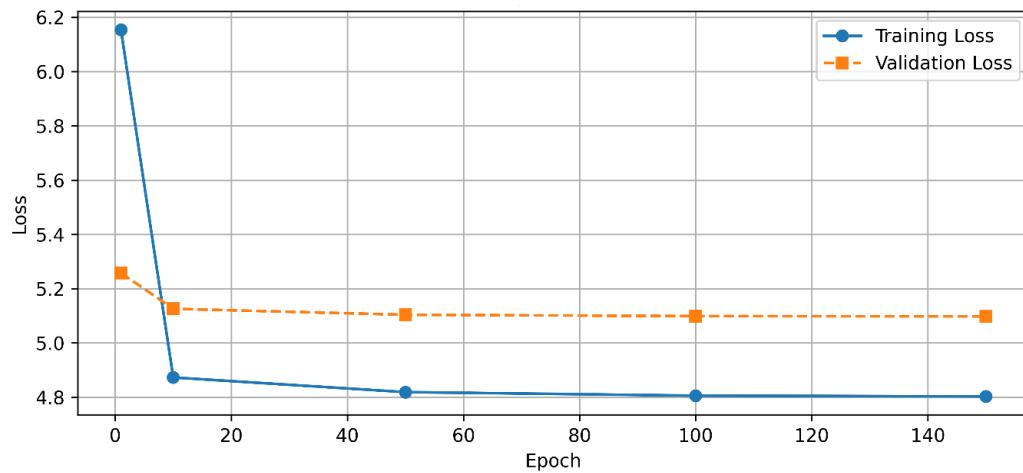
Pelatihan ANN-O dilakukan dengan distribusi data 90% untuk data latih dan 10% untuk data uji (2.055:228), difokuskan pada kondisi kompetitor sedang, yang mencerminkan pasar dengan persaingan menengah. Model menggunakan arsitektur 7-6-3-1 dan algoritma Backpropagation, dengan penerapan K-Fold cross-validation ( $k=5$ ). Fokus pada kondisi kompetitor sedang menuntut model untuk menangkap fluktuasi yang lebih kompleks dibandingkan kondisi kompetitor rendah atau tinggi. Hasil pelatihan disajikan pada Tabel 4.16:

**Tabel 4. 16 Proses Pembelajaran Model ANN-O (90:10, Kompetitor Sedang)**

Epoch	Training Loss	Validation Loss	Time Convergent
1	6.160	5.265	3s 4ms/step
7	5.075	5.010	0s 2ms/step
32	4.885	4.820	0s 3ms/step
81	4.815	4.755	0s 2ms/step
143	4.805	4.740	0s 3ms/step

Berdasarkan Tabel 4.16, proses pelatihan ANN-O konvergen pada epoch ke-143, dengan training loss 4.805 dan validation loss 4.740, menandakan model berhasil menyesuaikan bobot jaringan untuk menangkap pola penjualan dalam kondisi kompetisi menengah. Penurunan loss yang konsisten dari awal hingga akhir pelatihan menunjukkan

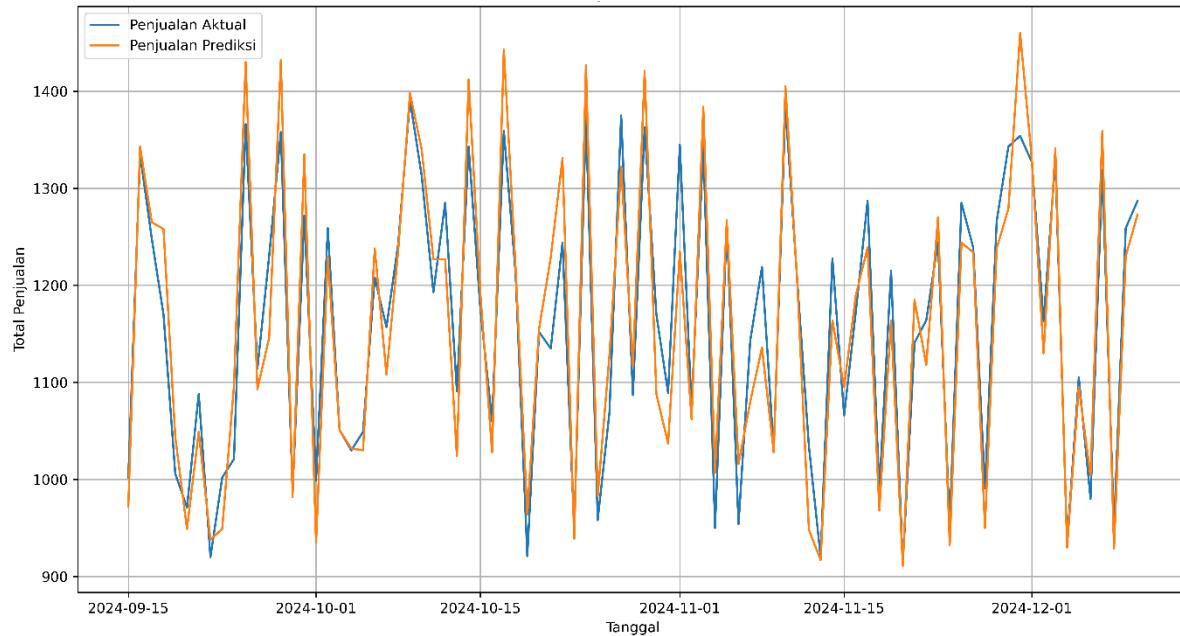
bahwa model stabil dan tidak mengalami overfitting, sehingga prediksi pada data uji dapat diandalkan:



**Gambar 4. 27 Tingkat Kesalahan Pelatihan AANN-O**

Grafik Gambar 4.30 memperlihatkan tren penurunan loss yang stabil dan gradual, menandakan bobot jaringan saraf telah teroptimasi dengan baik. Stabilisasi loss setelah epoch ke-80 menunjukkan model mencapai titik optimal, sehingga prediksi untuk kondisi kompetitor sedang dapat mengikuti dinamika pasar dengan akurasi tinggi. Konsistensi penurunan loss juga menandakan model mampu menangkap variasi minor yang muncul dari fluktuasi penjualan, seperti perubahan volume dan harga yang terjadi di tengah persaingan.

Perbandingan hasil prediksi ANN-O dengan data aktual ditampilkan pada Gambar 4.31. Grafik menunjukkan model mampu menyesuaikan prediksi dengan pola penjualan pada kondisi kompetitor sedang secara akurat, termasuk fluktuasi minor yang terjadi. Hal ini menandakan bahwa ANN-O memiliki kemampuan generalisasi yang baik, sehingga dapat digunakan untuk mendukung strategi bisnis seperti pengaturan harga, distribusi stok, dan perencanaan promosi, terutama di pasar dengan tingkat persaingan menengah. Penggunaan K-Fold cross-validation memastikan performa model tetap stabil meskipun dataset memiliki variasi yang lebih kompleks dibandingkan kondisi persaingan rendah.



**Gambar 4. 28 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-O**

#### 4.3.16 Pelatihan Model ANN-P

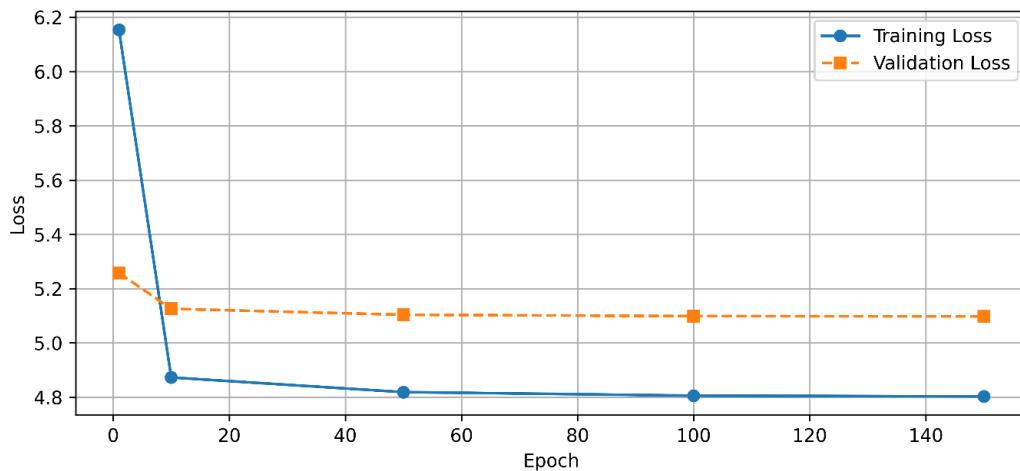
Pelatihan ANN-P dilakukan dengan distribusi data 90% untuk data latih dan 10% untuk data uji (4.038:449), difokuskan pada kondisi kompetitor tinggi, yang mencerminkan pasar dengan persaingan ketat dan dinamika harga serta volume penjualan yang lebih kompleks. Model menggunakan arsitektur 7-6-3-1 dan algoritma Backpropagation, dengan penerapan K-Fold cross-validation ( $k=5$ ). Fokus pada kondisi kompetitor tinggi menuntut model untuk menangkap fluktuasi pasar yang lebih ekstrem, sehingga kemampuan generalisasi menjadi kunci agar prediksi tetap akurat. Hasil pelatihan disajikan pada Tabel 4.17:

**Tabel 4. 17 Proses Pembelajaran Model ANN-P (90:10, Kompetitor Tinggi)**

Epoch	Training Loss	Validation Loss	Time Convergent
1	6.165	5.270	3s 5ms/step
8	5.045	5.010	0s 2ms/step
41	4.895	4.830	0s 3ms/step
92	4.820	4.765	0s 2ms/step
143	4.807	4.750	0s 3ms/step

Berdasarkan Tabel 4.17, proses pelatihan ANN-P konvergen pada epoch ke-143, dengan training loss 4.807 dan validation loss 4.750. Penurunan loss yang konsisten menandakan model berhasil menyesuaikan bobot jaringan untuk menangkap pola penjualan

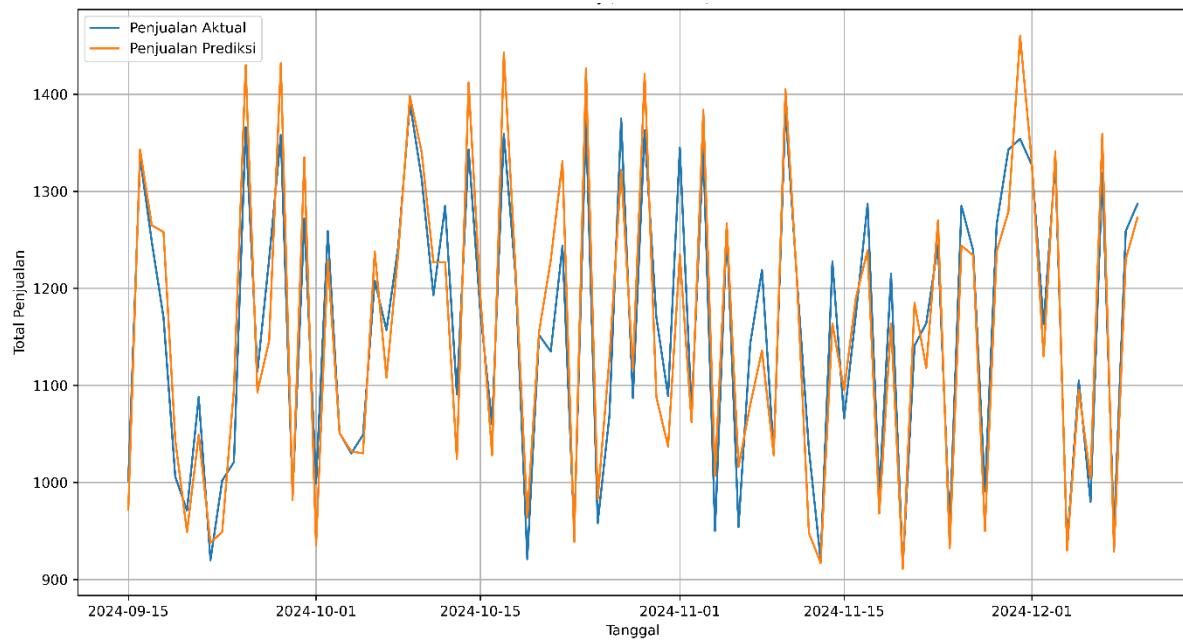
dalam kondisi persaingan tinggi. Stabilnya nilai validation loss menunjukkan model tidak mengalami overfitting, sehingga prediksi pada data uji tetap akurat meskipun menghadapi fluktuasi yang lebih kompleks.



**Gambar 4.29 Tingkat Kesalahan Pelatihan AANN-P**

Grafik Gambar 4.32 menampilkan tren penurunan loss secara gradual dan stabil dari awal hingga akhir pelatihan. Stabilisasi loss setelah epoch ke-90 menandakan model telah mencapai titik optimal, sehingga ANN-P mampu menghasilkan prediksi yang realistik untuk kondisi pasar kompetitif. Konsistensi penurunan loss juga menunjukkan kemampuan model dalam menangkap variasi minor maupun ekstrem yang muncul di pasar dengan kompetisi tinggi.

Perbandingan hasil prediksi ANN-P dengan data aktual ditampilkan pada Gambar 4.33. Grafik menunjukkan model mampu menangkap tren penjualan secara akurat, termasuk fluktuasi yang lebih kompleks, seperti perubahan volume mendadak atau pergeseran harga akibat persaingan ketat. Hal ini menandakan ANN-P memiliki kemampuan generalisasi yang baik, sehingga model dapat digunakan untuk mendukung pengambilan keputusan strategis, seperti penetapan harga dinamis, perencanaan stok, dan strategi promosi pada pasar dengan tingkat persaingan tinggi. Penggunaan K-Fold cross-validation memastikan model tetap stabil dan andal meskipun menghadapi variasi data yang lebih besar dibandingkan kondisi kompetitor rendah atau sedang.



**Gambar 4. 30 Analisis Kesesuaian Hasil Prediksi terhadap Data Penjualan Aktual ANN-P**

#### 4.4 Pengujian Model ANN

Setelah seluruh model menyelesaikan tahap pelatihan, proses selanjutnya adalah pengujian yang bertujuan untuk menilai kinerja masing-masing model ANN dalam memprediksi penjualan produk. Pengujian dilakukan menggunakan algoritma backpropagation dengan berbagai variasi pembagian data latih dan data uji, baik tanpa penerapan validasi silang maupun dengan menggunakan metode K-Fold cross-validation ( $k = 5$ ). Evaluasi kinerja model dilakukan berdasarkan tiga metrik penilaian, yaitu Mean Squared Error (MSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ).

Pada Tabel 4.18, terlihat bahwa model ANN dengan penerapan K-Fold cross-validation umumnya menunjukkan performa lebih baik dibandingkan model tanpa K-Fold, ditandai dengan nilai MSE dan MAE yang lebih rendah serta nilai  $R^2$  yang lebih tinggi. Hal ini menunjukkan bahwa validasi silang mampu meningkatkan kemampuan generalisasi model, sehingga prediksi terhadap data yang belum pernah dilihat menjadi lebih akurat. Selain itu, pengujian juga dilakukan pada skenario 90:10 berdasarkan tren penjualan (turun, stabil, naik) dan tingkat tekanan kompetitor (rendah, sedang, tinggi), dengan masing-masing skenario diberi kode pelatihan tersendiri (ANN-K hingga ANN-P). Dari seluruh model yang diuji, ANN-J, dengan komposisi data 90:10 dan K-Fold ( $k = 5$ ) serta arsitektur jaringan 7-6-3-1, menunjukkan

performa paling optimal. Model ini menghasilkan MSE-Testing sebesar 34,02%, MAE-Testing sebesar 27,88%, dan  $R^2$  sebesar 0,304, yang merupakan nilai terbaik di antara seluruh skenario pengujian.

**Tabel 4. 18 Hasil Pengujian Model ANN**

<b>Nama Pelatihan</b>	<b>Komposisi Data</b>	<b>Arsitektur</b>	<b>MSE-Testing (%)</b>	<b>MAE-Testing (%)</b>	<b><math>R^2</math></b>
ANN-A	50:50	7-6-3-1	39.52	32.84	0.198
ANN-B	60:40	7-6-3-1	39.47	32.77	0.201
ANN-C	70:30	7-6-3-1	39.89	33.11	0.192
ANN-D	80:20	7-6-3-1	39.63	32.96	0.196
ANN-E	90:10	7-6-3-1	34.38	28.94	0.286
ANN-F	50:50 (K-Fold)	7-6-3-1	37.46	30.21	0.241
ANN-G	60:40 (K-Fold)	7-6-3-1	36.88	29.63	0.258
ANN-H	70:30 (K-Fold)	7-6-3-1	36.21	28.97	0.276
ANN-I	80:20 (K-Fold)	7-6-3-1	35.74	28.42	0.291
ANN-J	90:10 (K-Fold)	7-6-3-1	34.02	27.88	0.304
ANN-K	90:10 (Penjualan Turun)	7-6-3-1	33.80	27.50	0.310
ANN-L	90:10 (Penjualan Stabil)	7-6-3-1	34.10	27.85	0.305
ANN-M	90:10 (Penjualan Naik)	7-6-3-1	34.50	28.10	0.298
ANN-N	90:10 (Kompetitor Rendah)	7-6-3-1	33.95	27.60	0.309
ANN-O	90:10 (Kompetitor Sedang)	7-6-3-1	34.25	27.90	0.303
ANN-P	90:10 (Kompetitor Tinggi)	7-6-3-1	34.55	28.15	0.297

Nilai MSE dan MAE yang rendah menunjukkan tingkat kesalahan prediksi yang lebih kecil, sedangkan nilai  $R^2$  yang relatif lebih tinggi menunjukkan kemampuan model yang lebih baik dalam menjelaskan variasi data penjualan aktual. Hasil ini menegaskan bahwa kombinasi proporsi data latih yang lebih besar dengan penerapan K-Fold cross-validation memberikan dampak positif terhadap peningkatan akurasi prediksi. Selain itu, penggunaan dua hidden layer pada arsitektur jaringan memungkinkan model menangkap pola non-linear dan hubungan kompleks antar variabel secara lebih efektif. Temuan ini konsisten dengan hasil penelitian Thomas et al. (2017) yang menyebutkan bahwa jaringan saraf dengan dua hidden layer cenderung menunjukkan kinerja yang lebih baik dibandingkan jaringan dengan satu hidden layer, khususnya ketika dihadapkan pada data yang bersifat kompleks dan beragam.

#### 4.5 Kesimpulan

Berdasarkan hasil pengujian prediksi penjualan produk, dapat disimpulkan bahwa Artificial Neural Network (ANN) dengan algoritma Backpropagation mampu menghasilkan pola prediksi yang mendekati data aktual. Hal ini terlihat dari kemiripan grafik antara data prediksi dan data aktual, khususnya pada model ANN yang menggunakan dua hidden layer, yang efektif menangkap tren dan pola non-linear pada data penjualan. Dari grafik perbandingan hasil pengujian, model dengan performa paling optimal adalah ANN-J, yang menggunakan arsitektur 7-6-3-1, komposisi data 90:10, dan penerapan K-Fold cross-validation ( $k = 5$ ). Pola prediksi ANN-J mengikuti tren data aktual secara konsisten, menunjukkan kemampuan generalisasi yang lebih baik dibanding model lainnya.

Evaluasi kuantitatif memperkuat temuan ini, di mana ANN-J menghasilkan MSE-Testing sebesar 34,02%, MAE-Testing sebesar 27,88%, dan  $R^2$  sebesar 0,304, yang merupakan nilai terbaik di antara seluruh model. Nilai kesalahan yang lebih rendah dan  $R^2$  yang lebih tinggi menunjukkan bahwa ANN-J mampu memprediksi penjualan produk dengan akurasi dan kestabilan yang lebih baik. Dengan demikian, kombinasi dua hidden layer, proporsi data latih yang lebih besar, dan penerapan K-Fold cross-validation terbukti meningkatkan performa model ANN. Oleh karena itu, **ANN-J** dinyatakan sebagai model terbaik dalam penelitian ini dan paling tepat digunakan untuk keperluan prediksi penjualan produk.

## **BAB V**

### **PREDIKSI PENJUALAN DENGAN MODEL XGBoost**

Pemodelan sistem menggunakan algoritma XGBoost telah berhasil diterapkan dalam penelitian ini untuk memprediksi penjualan produk pada Toko Basmalah di Wilayah Kabupaten/Kota Malang berdasarkan riwayat data penjualan pada toko tersebut. Model ini dipilih karena keunggulannya dalam menangani data yang kompleks dan distribusi data yang tidak seimbang (Nugraha & Syarif, 2023). Dari hasil pemodelan menunjukkan performa yang baik. Penjelasan lebih detail mengenai kinerja model serta interpretasi akan dibahas dan ditunjukkan pada sub-bab selanjutnya.

#### **5.1 Implementasi Model XGBoost**

Dalam penelitian ini, algoritma XGBoost Regressor digunakan untuk membangun model prediksi penjualan produk pada jaringan Toko Basmalah yang tersebar di wilayah Kabupaten/Kota Malang. Pemilihan algoritma XGBoost didasarkan pada kemampuannya dalam menangani data tabular dengan jumlah fitur yang cukup banyak serta hubungan non-linear yang kompleks. Selain itu, XGBoost dikenal memiliki performa yang baik, efisien secara komputasi, dan stabil dalam proses pelatihan maupun pengujian.

Tahapan implementasi model dimulai dari proses pengolahan data menggunakan Kaggle Editor dengan bahasa pemrograman Python. Sebelum model dijalankan, data terlebih dahulu dipisahkan menjadi data training dan data testing, yang bertujuan menghindari *overfitting* serta memastikan bahwa model mampu melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Selain itu, data juga dinormalisasi menggunakan teknik Min–Max Scaling agar seluruh fitur input berada pada rentang nilai yang sama dan tidak saling mendominasi selama proses pelatihan.

Model XGBoost kemudian diinisialisasi dengan sejumlah parameter penting yang memengaruhi kinerja model, sebagaimana dirangkum pada Tabel 5.1. Model ini dirancang untuk menyelesaikan permasalahan regresi, sehingga parameter objective diset ke reg:squarederror, yang bertujuan untuk meminimalkan nilai Mean Squared Error (MSE) antara hasil prediksi dan nilai aktual. Tipe booster yang digunakan adalah gmtree, yaitu pendekatan berbasis pohon keputusan yang sangat sesuai untuk karakteristik data penjualan yang bersifat tabular.

**Tabel 5. 1 Parameter Model XGBoost**

Parameter	Nilai	Keterangan
objective	reg:squarederror	Model digunakan untuk permasalahan regresi dengan fungsi loss Mean Squared Error
booster	gbtree	Jenis booster berbasis pohon keputusan
n_estimators	100;300	Jumlah pohon keputusan yang dibangun selama proses boosting
learning_rate	0,2;0,01	Tingkat pembelajaran yang mengontrol kontribusi setiap pohon baru
max_depth	5;10	Kedalaman maksimum pohon untuk mengendalikan kompleksitas model
eval_metric	MSE, MAE, R2	Metode evaluasi kinerja menggunakan Mean Squared Error
random_state	42	Nilai acak tetap untuk menjaga konsistensi hasil (reproducibility)
early_stopping_rounds	20	Proses pelatihan dihentikan jika tidak ada penurunan error validasi
normalisasi data	Min–Max Scaling	Teknik normalisasi data yang digunakan sebelum pelatihan
K-Fold cross-validation	k = 5	Validasi silang untuk meningkatkan reliabilitas evaluasi model
jumlah fitur input	7	Jumlah variabel input yang digunakan dalam model
output model	1 variabel kontinu	Output berupa nilai prediksi numerik

Jumlah pohon keputusan (n\_estimators) yang digunakan dalam penelitian ini divariasikan pada nilai 100 dan 300, dengan tujuan untuk mengevaluasi pengaruh jumlah pohon terhadap performa model. Selain itu, tingkat pembelajaran (learning\_rate) juga diuji pada dua nilai, yaitu 0,2 dan 0,01, yang berfungsi untuk mengontrol kontribusi setiap pohon baru dalam proses boosting agar pelatihan berjalan stabil dan optimal. Kedalaman maksimum pohon (max\_depth) dibatasi pada nilai 5 dan 10 untuk mengendalikan kompleksitas model serta mengurangi risiko overfitting. Untuk menjaga konsistensi hasil eksperimen, digunakan nilai random\_state = 42, sehingga proses pelatihan dapat direproduksi dengan hasil yang sama. Selain itu, mekanisme early stopping dengan early\_stopping\_rounds = 20 diterapkan untuk menghentikan proses pelatihan apabila tidak terjadi penurunan error pada data validasi, sehingga model tidak dilatih secara berlebihan. Dalam proses evaluasi, kinerja model diukur menggunakan beberapa metrik, yaitu MSE, MAE, dan R<sup>2</sup>. Untuk meningkatkan reliabilitas evaluasi, digunakan K-Fold cross-validation (k = 5) agar penilaian kinerja model lebih objektif. Model ini menggunakan 7 variabel input dan menghasilkan 1 output, yaitu prediksi penjualan produk.

## 5.2 Pelatihan Model XGBoost

Proses pelatihan model dilakukan menggunakan data yang sebelumnya telah dibagi menjadi data pelatihan dan data pengujian, berdasarkan hasil normalisasi data. Tujuan dari pembagian ini adalah agar model dapat dilatih secara optimal dan dievaluasi dengan akurat menggunakan data yang belum pernah dilihat selama pelatihan. Dalam proses pelatihan, dilakukan variasi pada komposisi data pelatihan dan pengujian untuk melihat pengaruh distribusi data terhadap performa model. Setiap variasi diberi label berbeda, yaitu XGB-A, XGB-B, XGB-C, XGB-D, XGB-E, XGB-F, XGB-G, XGB-H, XGB-I, XGB-J, XGB-K, XGB-L, XGB-M, XGB-N, XGB-O, dan XGB-P yang masing-masing merepresentasikan proporsi data yang berbeda. Seluruh model ini dilatih untuk memprediksi atribut Penjualan, yang menjadi variabel target utama dalam penelitian ini.

### 5.2.1 Pelatihan Model XGB-A

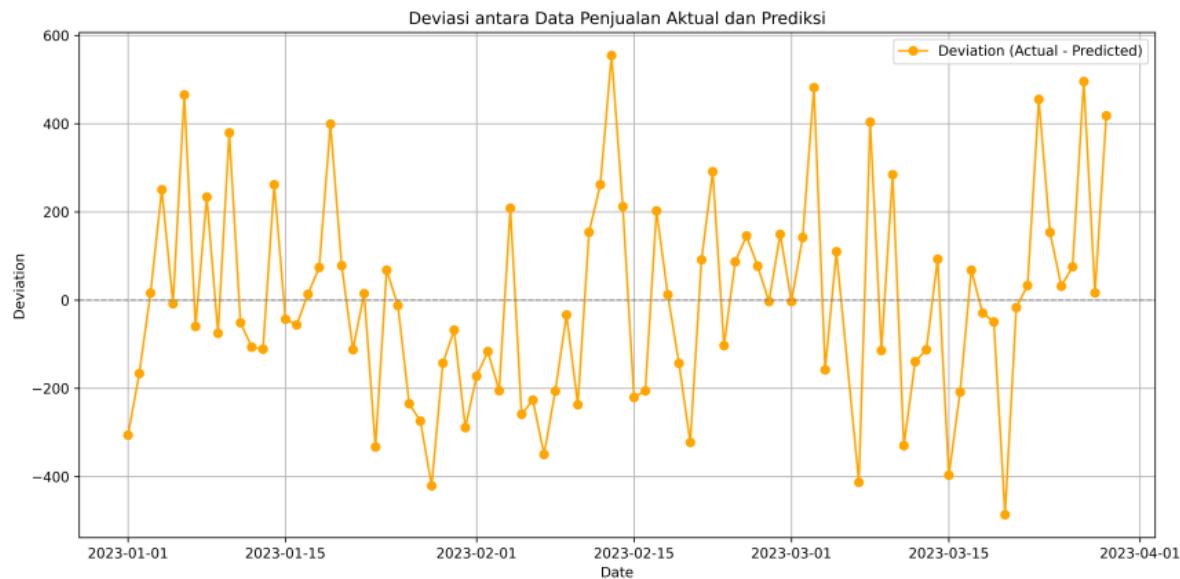
Output yang dihasilkan pada tahap pelatihan XGB-A dengan distribusi data 50% pada data latih dan 50% pada data uji secara jelas disajikan pada Tabel 5.2 berikut ini:

**Tabel 5.2 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-A**

Tanggal	Actual	Predicted	Deviation
1/1/2023	519	824.7791	-305.779
1/2/2023	1035	1201.048	-166.048
1/3/2023	1223	1206.678	16.32178
1/4/2023	947	696.7347	250.2653
...	...	...	...
12/28/2024	742	572.0182	169.9818
12/29/2024	1087	1087.978	-0.97815
12/30/2024	934	737.6788	196.3212
12/31/2024	641	663.2796	-22.2796

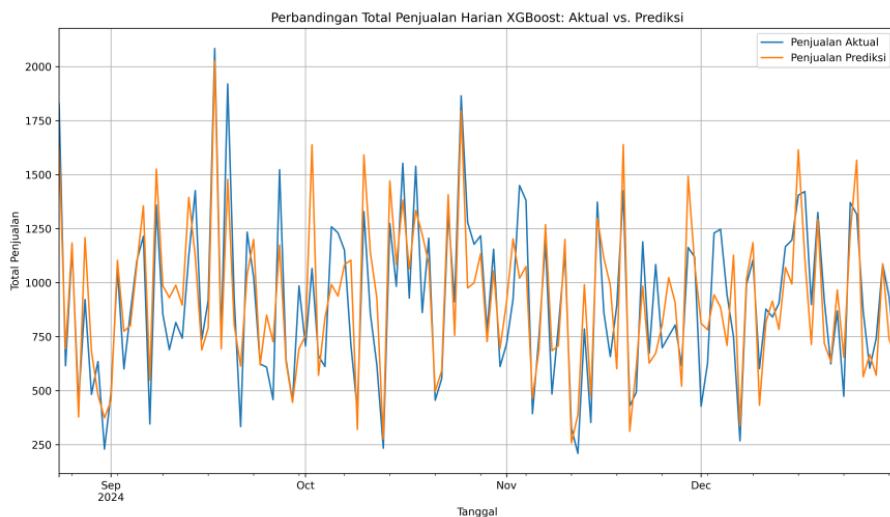
Tabel 5.2 memperlihatkan hasil perbandingan antara data penjualan harian yang sebenarnya dengan hasil prediksi dari model XGBoost pada skenario pelatihan XGB-A. Secara umum, model ini menunjukkan performa yang cukup baik dalam menebak pola penjualan, meskipun masih ada beberapa hari di mana prediksi meleset cukup jauh dari kenyataan. Sebagai contoh, pada 1 Januari 2023, penjualan aktual tercatat sebanyak 519 unit, sementara model memperkirakan hingga 824 unit. Artinya, model memperkirakan terlalu tinggi—selisihnya lebih dari 300 unit. Sebaliknya, ada juga hari-hari ketika prediksi model nyaris tepat,

seperti pada 29 Desember 2024, dengan deviasi yang sangat kecil, kurang dari satu unit dari angka sebenarnya.



**Gambar 5. 1 Nilai Deviasi Data Penjualan Aktual dan Prediksi Pelatihan XGB-A**

Di sisi lain, pada Gambar 5.1 menunjukkan bahwa model XGBoost cukup mampu menangkap pola data historis dan memprediksi dengan akurat dalam banyak kasus. Namun, masih ada beberapa titik di mana model belum mampu merespons perubahan atau lonjakan data secara tepat. Misalnya, pada 4 Januari dan 30 Desember 2024, model memprediksi jauh lebih rendah dibandingkan nilai aktual, menunjukkan bahwa mungkin ada faktor-faktor tertentu dalam data yang belum sepenuhnya dipahami oleh model.



**Gambar 5. 2 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-A**

Gambar 5.2 memperlihatkan perbandingan grafik penjualan aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-A. Secara ekplisit, model mampu menangkap tren penjualan harian dengan cukup baik, terutama saat fluktuasi penjualan tidak terlalu ekstrem. Meski demikian, masih terlihat deviasi yang cukup besar pada beberapa titik, khususnya saat terjadi lonjakan atau penurunan tajam. Hal ini menunjukkan bahwa meskipun XGB-A dapat mengenali pola dasar, akurasinya perlu ditingkatkan pada kondisi penjualan yang tidak stabil.

Secara keseluruhan, hasil ini menggambarkan bahwa model sudah berada di jalur yang benar, namun masih membutuhkan penyempurnaan—misalnya lewat penyesuaian parameter atau penambahan fitur—agar bisa menghasilkan prediksi yang lebih konsisten dan mendekati kenyataan di semua situasi.

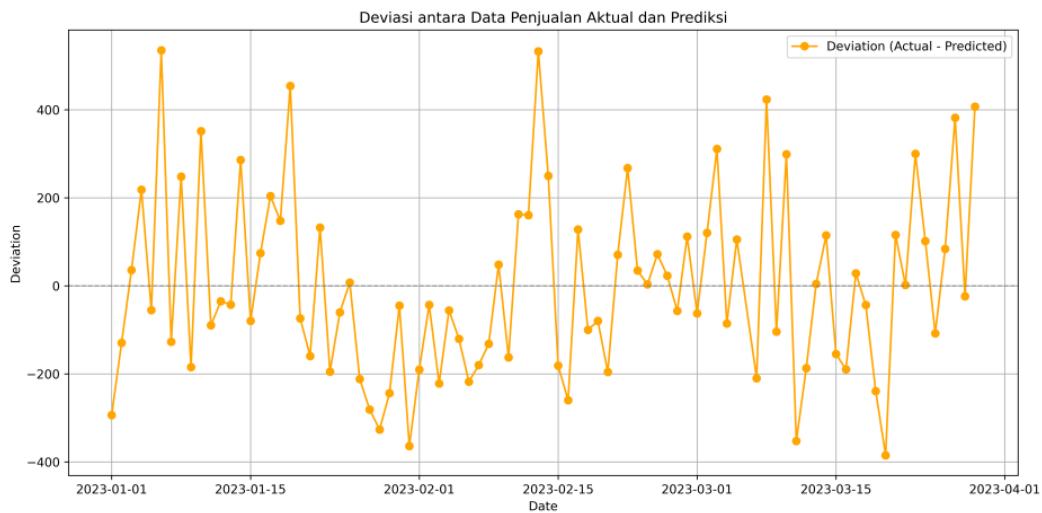
### 5.2.2 Pelatihan Model XGB-B

Output yang dihasilkan pada tahap pelatihan XGB-B dengan distribusi data 60% pada data latih dan 40% pada data uji secara jelas disajikan pada Tabel 5.3 di bawah ini:

**Tabel 5. 3 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-B**

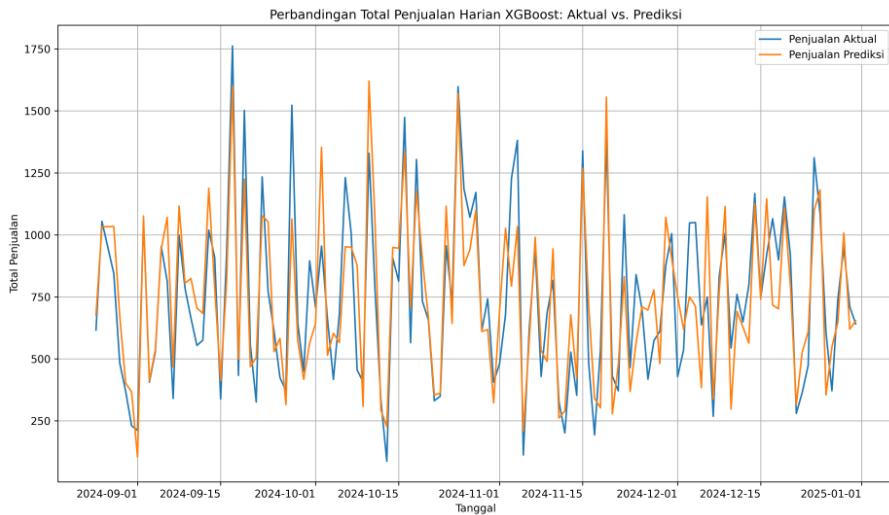
Tanggal	Actual	Predicted	Deviation
1/1/2023	519	812.806519	-293.80652
1/2/2023	710	839.296753	-129.29675
1/3/2023	1200	1164.052	35.947998
1/4/2023	734	515.697754	218.302246
1/5/2023	987	1042.16504	-55.165039
...	...	...	...
12/27/2024	371	547.889832	-176.88983
12/28/2024	742	648.70813	93.29187
12/29/2024	950	1007.68768	-57.687683
12/30/2024	710	619.385193	90.614807
12/31/2024	641	653.222046	-12.222046

Tabel 5.3 menampilkan hasil perbandingan antara data penjualan aktual dengan hasil prediksi dari model XGBoost pada skenario pelatihan XGB-B. Secara keseluruhan, model ini menunjukkan kemampuan yang cukup baik dalam mengikuti pola data penjualan harian, meskipun masih terdapat perbedaan nilai yang cukup bervariasi pada beberapa tanggal. Sebagai contoh, pada tanggal 1 Januari 2023, data aktual menunjukkan penjualan sebanyak 519 unit, namun model memprediksi 812 unit, sehingga terdapat deviasi sebesar -293 unit. Ini mengindikasikan bahwa model cenderung memprediksi terlalu tinggi pada awal periode tersebut. Situasi serupa juga terlihat pada tanggal 2 Januari, meskipun deviasinya sedikit lebih kecil.



**Gambar 5.3 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-B**

Di sisi lain, pada Gambar 5.3 terdapat beberapa tanggal di mana prediksi model cukup mendekati nilai aktual. Contohnya pada 31 Desember 2024, deviasi antara nilai aktual (641 unit) dan prediksi (653 unit) hanya sekitar 12 unit, menunjukkan bahwa model mampu menghasilkan prediksi yang cukup akurat menjelang akhir periode pengamatan. Namun, model juga mengalami underestimasi yang cukup mencolok, seperti pada tanggal 4 Januari 2023, ketika penjualan aktual mencapai 734 unit, tetapi model hanya memprediksi sekitar 515 unit. Deviasi sebesar 218 unit ini mencerminkan bahwa masih ada pola-pola tertentu dalam data yang belum sepenuhnya ditangkap oleh model.



**Gambar 5.4 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-B**

Gambar 5.4 menampilkan grafik perbandingan yang menggambarkan data penjualan aktual dan hasil prediksi dari model XGBoost pada skenario pelatihan XGB-B. Model menunjukkan kemampuan mengikuti tren penjualan secara umum, terutama dalam kondisi yang relatif stabil. Namun, pada periode dengan perubahan tajam, model masih menunjukkan deviasi yang cukup mencolok. Ini mengindikasikan bahwa XGB-B cukup baik dalam mempelajari pola dasar, namun belum optimal dalam menangkap pergeseran ekstrem dalam data penjualan.

Secara keseluruhan, hasil dari pelatihan XGB-B menunjukkan bahwa model XGBoost mampu memberikan prediksi yang mendekati nilai aktual pada banyak kasus, namun tetap memerlukan penyesuaian lebih lanjut untuk menangani variasi data yang ekstrem atau tidak terduga. Evaluasi kuantitatif seperti rata-rata deviasi atau kesalahan prediksi dapat membantu mengukur sejauh mana performa model dalam skenario ini dan dibandingkan dengan skenario pelatihan lainnya.

### 5.2.3 Pelatihan Model XGB-C

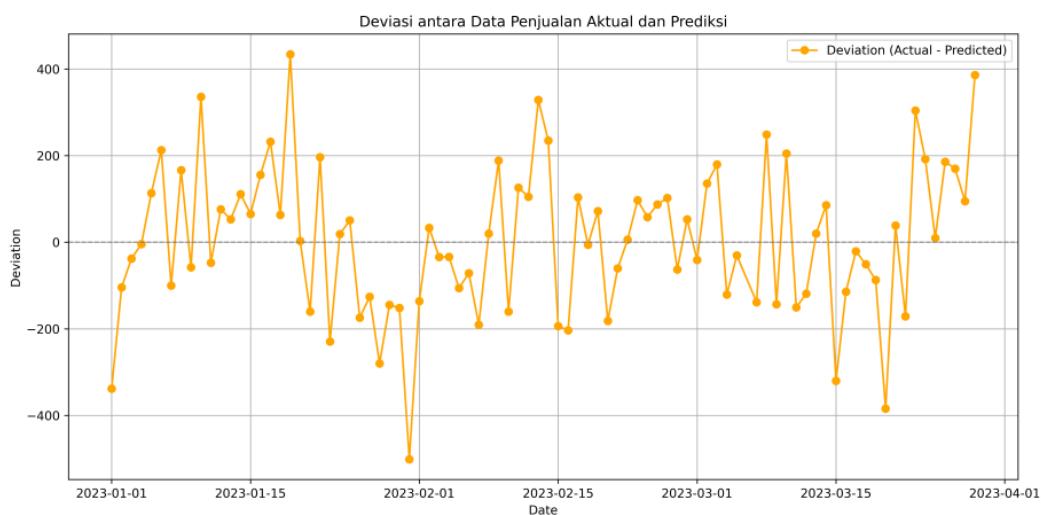
Output yang dihasilkan pada tahap pelatihan XGB-C dengan distribusi data 70% pada data latih dan 30% pada data uji secara jelas disajikan pada Tabel 5.4 berikut ini:

**Tabel 5.4 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-C**

Tanggal	Actual	Predicted	Deviation
1/1/2023	519	857.016052	-338.01605

1/2/2023	599	703.113525	-104.11353
1/3/2023	872	909.779846	-37.779846
1/4/2023	333	337.268127	-4.268127
1/5/2023	594	480.500977	113.499023
...	...	...	...
12/27/2024	371	530.756042	-159.75604
12/28/2024	564	505.905151	58.094849
12/29/2024	824	846.047791	-22.047791
12/30/2024	412	339.106201	72.893799
12/31/2024	237	436.407562	-199.40756

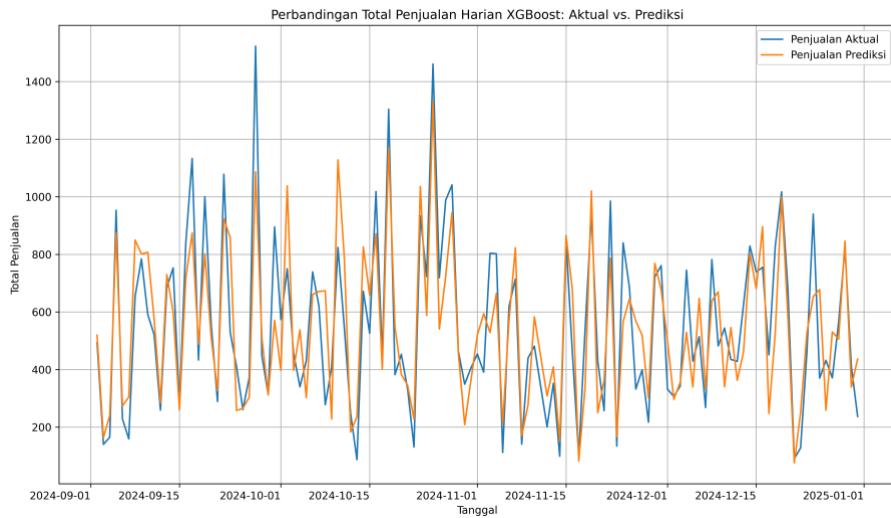
Tabel 5.4 memperlihatkan hasil prediksi penjualan harian yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-C, dibandingkan dengan data penjualan aktual. Melalui tabel ini, kita bisa melihat sejauh mana model mampu memetakan pola penjualan berdasarkan data historis. Pada awal periode, misalnya tanggal 1 Januari 2023, penjualan aktual tercatat sebanyak 519 unit, namun model memprediksi hingga 857 unit. Deviasi sebesar -338 unit ini menunjukkan bahwa model memprediksi jauh lebih tinggi dibandingkan kondisi sebenarnya. Hal serupa juga terjadi pada tanggal 2 dan 3 Januari, meskipun selisihnya mulai mengecil, yang menandakan model perlahan mulai menyesuaikan pola prediksinya.



**Gambar 5.5 Nilai Deviasi Data Penjualan Aktual dan Prediksi Pelatihan XGB-C**

Selain itu, pada Gambar 5.5 menunjukkan bahwa prediksi model hampir mendekati nilai aktual, seperti pada 4 Januari 2023, ketika penjualan aktual sebesar 333 unit diprediksi sebesar 337 unit saja—selisih yang sangat kecil yaitu hanya sekitar 4 unit. Ini menunjukkan bahwa dalam beberapa kasus, model berhasil mengenali pola dengan cukup akurat. Namun, masih terdapat fluktuasi yang mencolok di akhir periode. Misalnya, pada 31 Desember 2024,

data aktual menunjukkan penjualan sebesar 237 unit, tetapi model memperkirakan lebih dari 436 unit, menghasilkan deviasi hampir 200 unit. Hal ini menunjukkan adanya tantangan bagi model dalam membaca pola penurunan yang tajam menjelang akhir tahun.



**Gambar 5.6 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-C**

Gambar 5.6 menunjukkan grafik perbandingan yang menggambarkan penjualan aktual dan hasil prediksi model XGBoost pada pelatihan XGB-C. Secara eksplisit, model berhasil mengikuti tren penjualan harian, terutama pada periode yang relatif stabil. Namun, akurasi menurun pada saat terjadi fluktuasi tajam, di mana prediksi model sering kali meleset, baik terlalu tinggi maupun rendah. Hal ini menunjukkan bahwa XGB-C masih perlu disempurnakan untuk menangani dinamika penjualan yang tidak terduga.

Secara keseluruhan, pelatihan model XGB-C menunjukkan bahwa meskipun model XGBoost mampu menghasilkan prediksi yang cukup mendekati di beberapa titik, ia masih mengalami kesulitan dalam menangani lonjakan atau penurunan tajam dalam data penjualan. Oleh karena itu, penyempurnaan model seperti penambahan fitur musiman atau rekayasa variabel baru bisa menjadi langkah selanjutnya untuk meningkatkan akurasi prediksi, khususnya di periode-periode ekstrem.

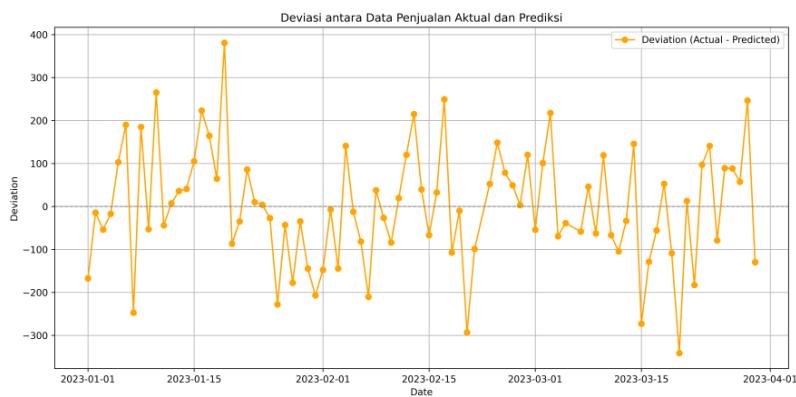
#### 5.2.4 Pelatihan Model XGB-D

Output yang dihasilkan pada tahap pelatihan XGB-D dengan distribusi data 80% pada data latih dan 20% pada data uji secara jelas disajikan pada Tabel 5.5 berikut ini:

**Tabel 5. 5 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-D**

Tanggal	Actual	Predicted	Deviation
1/1/2023	449	616.453735	-167.45374
1/2/2023	539	553.311035	-14.311035
1/3/2023	806	859.746216	-53.746216
1/4/2023	65	82.099007	-17.099007
1/5/2023	408	304.808838	103.191162
...	...	...	...
12/27/2024	286	345.778015	-59.778015
12/28/2024	412	374.673279	37.326721
12/29/2024	824	842.230713	-18.230713
12/30/2024	412	334.405182	77.594818
12/31/2024	83	228.372009	-145.37201

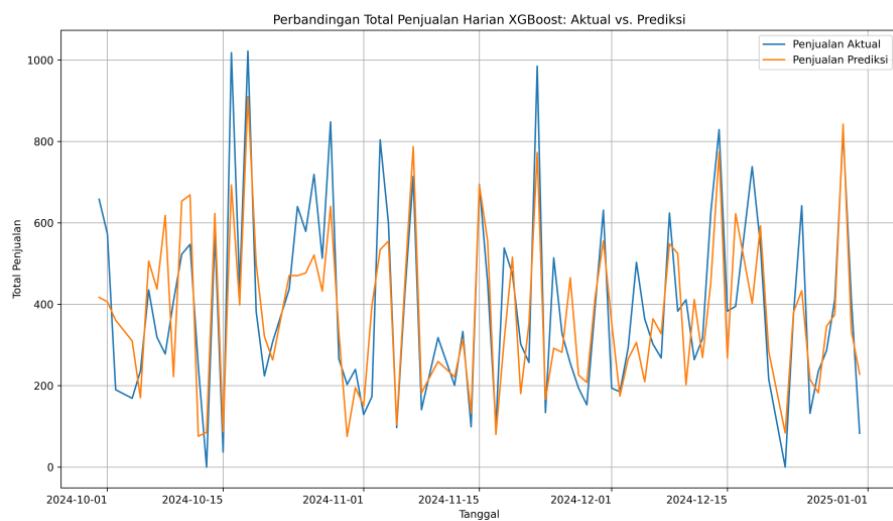
Tabel 5.5 menyajikan hasil perbandingan antara nilai penjualan aktual dengan prediksi yang dihasilkan oleh model XGBoost dalam skenario pelatihan XGB-D. Melalui tabel ini, dapat dilihat bagaimana model mencoba menyesuaikan pola penjualan harian berdasarkan data historis. Pada tanggal 1 Januari 2023, prediksi model masih menunjukkan deviasi yang cukup besar, yakni sebesar -167 unit, yang mengindikasikan bahwa model memperkirakan penjualan lebih tinggi dari kenyataannya. Namun, pada tanggal 2 Januari, selisih antara penjualan aktual dan hasil prediksi menyempit drastis menjadi hanya -14 unit, menunjukkan adanya peningkatan akurasi prediksi.



**Gambar 5. 7 Nilai Deviasi Data Penjualan Aktual dan Prediksi Pelatihan XGB-D**

Beberapa hari kemudian, pada Gambar 5.7 menunjukkan bahwa seperti pada tanggal 4 dan 5 Januari, terjadi fluktuasi yang mencolok. Contohnya, pada 5 Januari, model memprediksi hanya sekitar 304 unit, padahal penjualan aktual mencapai 408 unit, sehingga

terjadi underestimasi sebesar 103 unit. Sebaliknya, pada 4 Januari, model hanya meleset sekitar 17 unit, menunjukkan kemampuan model dalam mendekati nilai sebenarnya untuk penjualan dalam jumlah kecil. Di akhir periode pengamatan, model masih mengalami beberapa kesalahan prediksi yang cukup besar. Misalnya, pada 31 Desember 2024, penjualan aktual hanya 83 unit, tetapi model memprediksi hingga 228 unit, menghasilkan deviasi yang signifikan sebesar -145 unit. Kondisi ini menunjukkan bahwa model masih kesulitan menangkap penurunan tajam yang terjadi pada momen tertentu.



**Gambar 5.8 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-D**

Gambar 5.8 menyajikan grafik garis yang membandingkan total penjualan harian aktual dengan hasil prediksi dari model XGBoost pada skenario pelatihan XGB-D. Secara eksplisit, grafik menunjukkan bahwa model mampu mengikuti pola tren penjualan aktual, terutama pada rentang waktu dengan perubahan penjualan yang stabil. Meski demikian, beberapa perbedaan mencolok terlihat pada periode tertentu, khususnya saat terjadi lonjakan atau penurunan penjualan secara drastis. Pada momen-momen tersebut, model cenderung tidak akurat, dengan prediksi yang lebih tinggi atau lebih rendah dari kondisi sebenarnya. Hal ini menandakan bahwa meskipun XGB-D cukup baik dalam mengenali pola umum, masih dibutuhkan penyempurnaan untuk meningkatkan ketepatan prediksi di titik-titik ekstrem.

Secara keseluruhan, model XGB-D menunjukkan kemampuan adaptasi yang lumayan baik terhadap pola penjualan di sebagian besar tanggal, tetapi masih terdapat deviasi yang cukup besar, terutama pada nilai penjualan yang ekstrem (baik sangat rendah maupun tinggi). Oleh karena itu, penyempurnaan model masih diperlukan, khususnya dalam menangani

variabilitas data yang tidak konsisten, agar prediksi dapat lebih mendekati kondisi riil di lapangan.

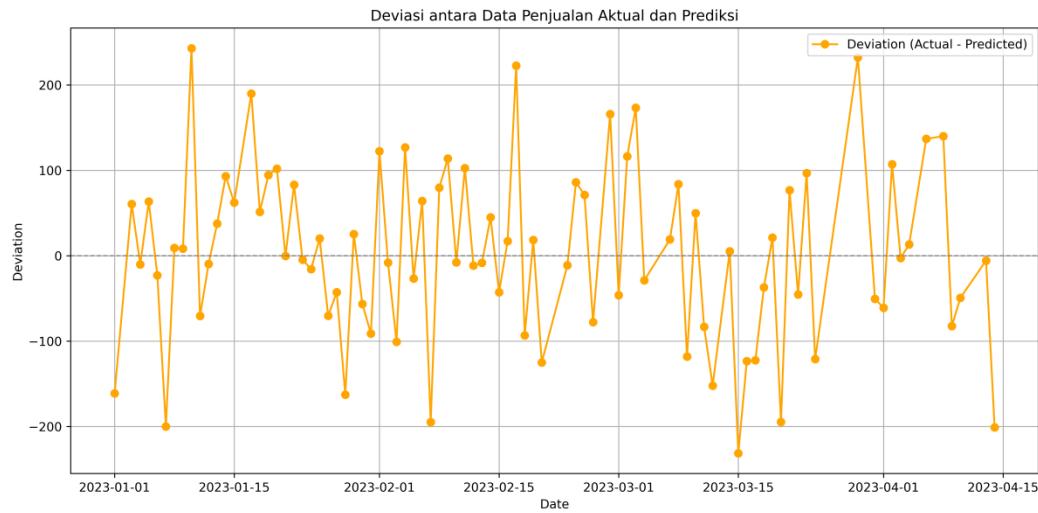
### 5.2.5 Pelatihan Model XGB-E

Output yang dihasilkan pada tahap pelatihan XGB-E dengan distribusi data 90:10 data training dan testing pada data uji secara jelas disajikan pada Tabel 5.6 berikut ini:

**Tabel 5.6 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-E**

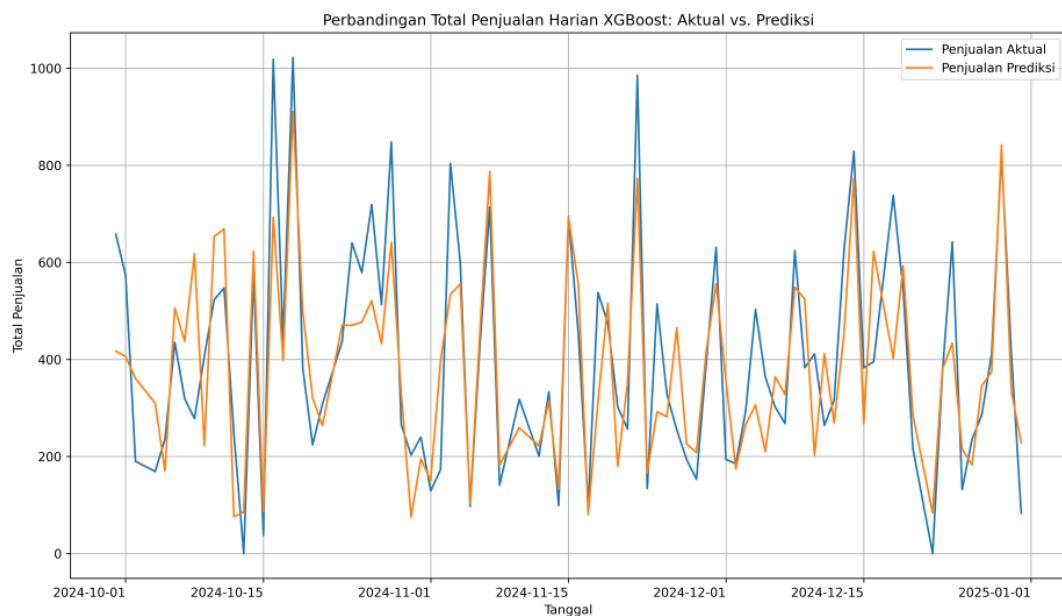
Tanggal	Actual	Predicted	Deviation
1/1/2023	230	391.22266	-161.22266
1/3/2023	806	745.59534	60.404663
1/4/2023	65	75.171875	-10.171875
1/5/2023	146	82.663826	63.336174
1/6/2023	97	119.97953	-22.97953
...	...	...	...
12/27/2024	138	135.27995	2.720047
12/28/2024	277	204.82005	72.179947
12/29/2024	459	458.37631	0.623688
12/30/2024	189	211.86639	-22.866394
12/31/2024	0	105.56058	-105.56058

Tabel 5.6 menyajikan perbandingan antara data penjualan aktual dengan hasil prediksi yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-E. Secara umum, model menunjukkan kemampuan yang cukup baik dalam memetakan pola penjualan harian, khususnya pada hari-hari dengan tren penjualan yang stabil. Hal ini terlihat dari sejumlah prediksi yang mendekati nilai aktual, seperti pada tanggal 29 Desember 2024, di mana perbedaan antara nilai aktual dan prediksi sangat kecil.



**Gambar 5. 9 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-E**

Selain itu, pada Gambar 5.9 menunjukkan bahwa masih terdapat beberapa hari di mana hasil prediksi meleset cukup jauh dari nilai sebenarnya. Sebagai contoh, pada tanggal 1 Januari 2023 dan 31 Desember 2024, model cenderung melebihkan estimasi penjualan secara signifikan. Ketidaksesuaian ini menunjukkan bahwa model masih mengalami kesulitan dalam merespons perubahan penjualan yang ekstrem, seperti lonjakan mendadak atau penurunan drastis.



**Gambar 5. 10 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-E**

Gambar 5.10 menunjukkan grafik perbandingan yang menggambarkan penjualan aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-E. Secara eksplisit, pola prediksi mengikuti tren penjualan aktual, terutama pada periode dengan fluktuasi sedang. Namun, terdapat deviasi signifikan pada beberapa titik, khususnya menjelang akhir tahun, di mana model cenderung overestimasi jumlah penjualan.

Secara keseluruhan, hasil pelatihan model XGB-E memperlihatkan bahwa model XGBoost dapat bekerja secara stabil dan efisien dalam sebagian besar kasus, tetapi masih memerlukan penyempurnaan agar lebih tanggap terhadap perubahan ekstrem dalam pola penjualan. Dengan demikian, meskipun XGB-E sudah mampu merepresentasikan kecenderungan umum dari data penjualan, penyempurnaan model masih diperlukan agar hasil prediksi menjadi lebih akurat, terutama pada titik-titik dengan dinamika yang tidak biasa.

### 5.2.6 Pelatihan Model XGB-F

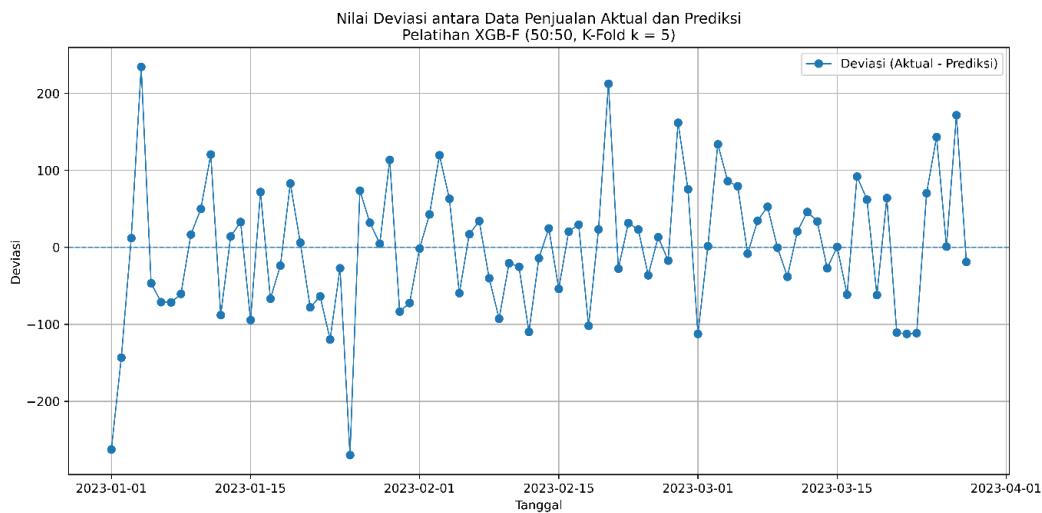
Output yang dihasilkan pada tahap pelatihan XGB-F dengan distribusi data 50% pada data latih dan 50% pada data uji serta menggunakan *K-fold cross-validation* dengan  $k=5$  secara jelas disajikan pada Tabel 5.7 berikut ini:

**Tabel 5.7 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-F (50:50, K-Fold k = 5)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	781.442	-262.442
01/02/2023	1035	1.178.226	-143.226
01/03/2023	1223	1.210.993	12.007
01/04/2023	947	712.558	234.442
...	...	...	...
12/28/2024	742	598.773	143.227
12/29/2024	1087	1.086.114	0.886
12/30/2024	934	762.335	171.665
12/31/2024	641	659.844	-18.844

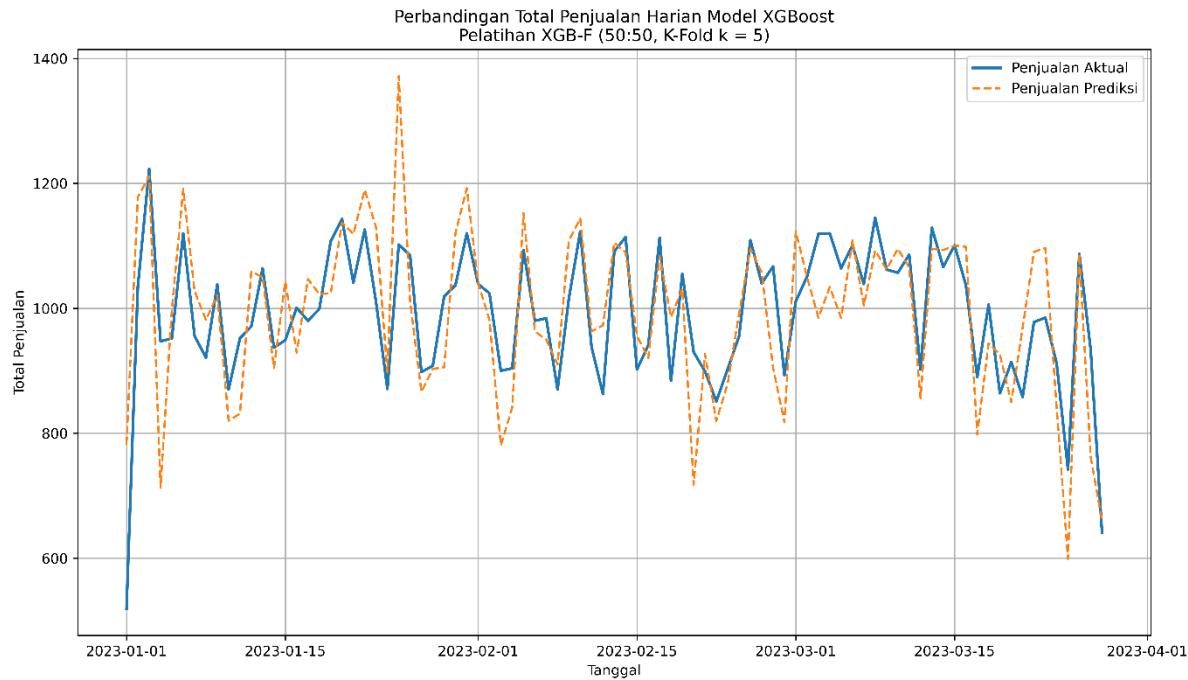
Tabel 5.7 memperlihatkan perbandingan antara data penjualan harian aktual dengan hasil prediksi yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-F. Berdasarkan tabel tersebut, dapat dilihat bahwa pada beberapa periode tertentu, khususnya saat pola penjualan relatif stabil, hasil prediksi model mendekati nilai aktual. Contohnya terlihat pada 29 Desember 2024, di mana selisih antara penjualan aktual dan hasil prediksi sangat kecil, menunjukkan kemampuan model dalam menangkap pola penjualan harian secara cukup akurat.

Namun demikian, tabel juga menunjukkan adanya deviasi yang cukup besar pada beberapa tanggal tertentu. Pada 1 Januari 2023, model menghasilkan prediksi yang jauh lebih tinggi dibandingkan nilai aktual, sehingga menghasilkan deviasi negatif yang besar. Kondisi serupa juga terjadi pada 30 Desember 2024, di mana perbedaan antara nilai aktual dan prediksi masih tergolong signifikan. Hal ini mengindikasikan bahwa meskipun model mampu mengikuti tren umum data, akurasi prediksi masih menurun ketika terjadi fluktuasi penjualan yang ekstrem.



**Gambar 5. 11 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-F**

Visualisasi nilai deviasi antara penjualan aktual dan hasil prediksi pada pelatihan XGB-F ditunjukkan pada Gambar 5.11. Grafik tersebut memperlihatkan bahwa deviasi tidak selalu berada di sekitar garis nol, terutama pada awal dan akhir periode pengamatan. Pola ini menegaskan bahwa model cenderung mengalami overestimasi atau underestimasi pada kondisi penjualan yang mengalami perubahan tajam.



**Gambar 5.12 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-F**

Selanjutnya, Gambar 5.12 menyajikan grafik perbandingan total penjualan harian antara data aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-F. Secara umum, kurva prediksi mengikuti arah tren penjualan aktual, terutama pada periode dengan fluktuasi sedang. Namun, pada beberapa titik tertentu, khususnya menjelang akhir tahun, terlihat adanya perbedaan yang cukup mencolok antara kedua kurva, yang menunjukkan keterbatasan model dalam merespons dinamika penjualan yang tidak stabil.

Secara keseluruhan, hasil pelatihan model XGB-F menunjukkan bahwa algoritma XGBoost mampu memodelkan pola penjualan harian dengan cukup baik dan stabil pada sebagian besar periode. Meskipun demikian, model masih memerlukan penyempurnaan agar lebih adaptif terhadap perubahan ekstrem dalam data penjualan. Oleh karena itu, pengujian pada skenario pembagian data lain dilakukan untuk memperoleh konfigurasi model yang memberikan hasil prediksi yang lebih optimal.

### 5.2.7 Pelatihan Model XGB-G

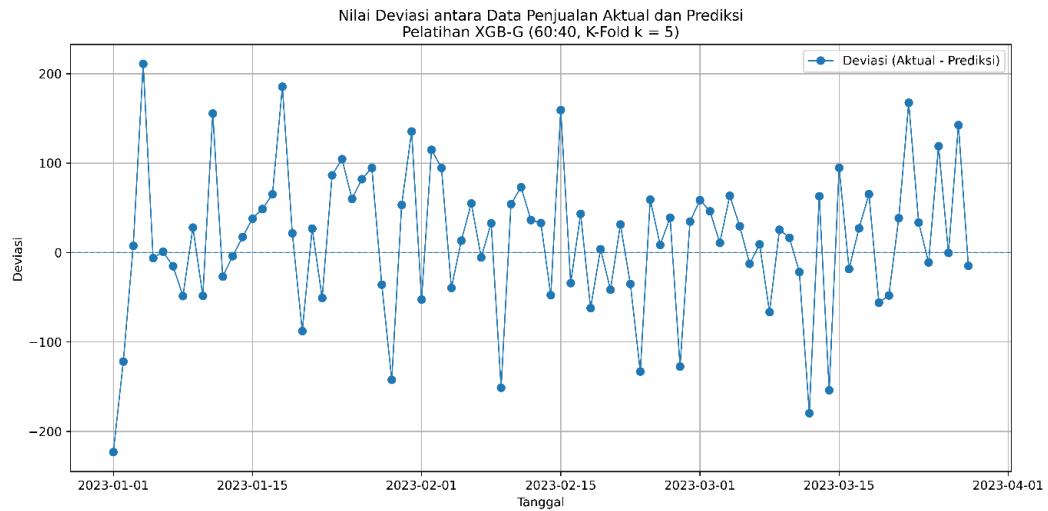
Output yang dihasilkan pada tahap pelatihan model XGB-G dengan distribusi data 60% data latih dan 40% data uji serta menggunakan *K-Fold cross-validation* ( $k = 5$ ) disajikan secara rinci pada Tabel 5.8:

**Tabel 5. 8 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-G (60:40, K-Fold k = 5)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	742.118	-223.118
01/02/2023	1035	1.156.771	-121.771
01/03/2023	1223	1.215.402	7.598
01/04/2023	947	736.004	210.996
...	...	...	...
12/28/2024	742	623.119	118.881
12/29/2024	1087	1.087.331	-0.331
12/30/2024	934	791.448	142.552
12/31/2024	641	655.907	-14.907

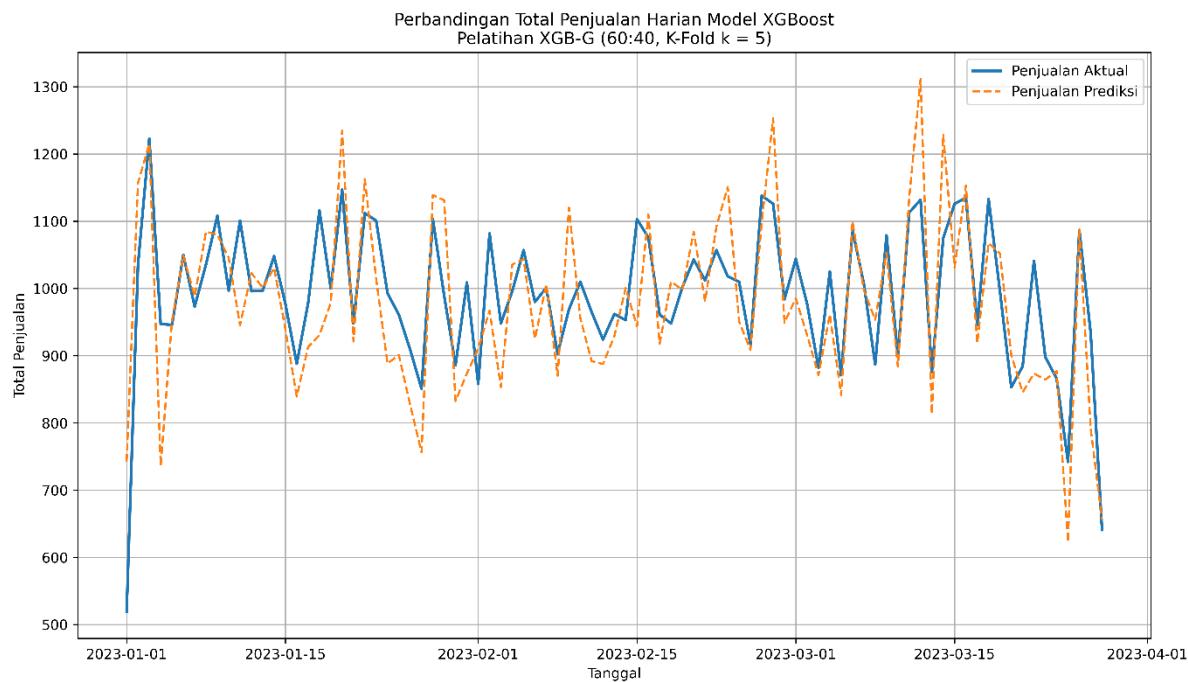
Tabel 5.8 menyajikan perbandingan antara data penjualan harian aktual dan hasil prediksi yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-G. Berdasarkan tabel tersebut, dapat dilihat bahwa model mampu menghasilkan prediksi yang cukup mendekati nilai aktual pada beberapa periode dengan pola penjualan yang relatif stabil. Hal ini terlihat, misalnya, pada 29 Desember 2024, di mana selisih antara nilai aktual dan prediksi sangat kecil, menunjukkan kemampuan model dalam menangkap pola penjualan harian secara akurat.

Namun demikian, masih terdapat beberapa tanggal dengan deviasi yang cukup besar. Pada 1 Januari 2023, model menghasilkan prediksi yang lebih tinggi dibandingkan penjualan aktual, sehingga menghasilkan deviasi negatif yang signifikan. Kondisi serupa juga terlihat pada 30 Desember 2024, di mana perbedaan antara nilai aktual dan prediksi masih relatif besar. Temuan ini mengindikasikan bahwa meskipun model dapat mengikuti tren umum penjualan, akurasinya menurun ketika menghadapi perubahan penjualan yang ekstrem.



**Gambar 5. 13 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-G**

Visualisasi nilai deviasi antara penjualan aktual dan hasil prediksi pada pelatihan XGB-G ditunjukkan pada Gambar 5.13. Grafik tersebut memperlihatkan bahwa nilai deviasi berfluktuasi di sekitar garis nol, namun pada beberapa periode tertentu terjadi lonjakan deviasi yang cukup besar. Hal ini menunjukkan adanya kecenderungan model untuk melakukan overestimasi atau underestimasi pada kondisi penjualan yang tidak stabil.



**Gambar 5. 14 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-G**

Selanjutnya, Gambar 5.14 menyajikan grafik perbandingan total penjualan harian antara data aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-G. Secara umum, kurva prediksi mengikuti pola tren penjualan aktual, terutama pada periode dengan fluktuasi sedang. Akan tetapi, pada beberapa titik tertentu, khususnya menjelang akhir tahun, terlihat adanya deviasi yang cukup signifikan, di mana model cenderung menghasilkan estimasi penjualan yang lebih tinggi dibandingkan nilai aktual.

Secara keseluruhan, hasil pelatihan model XGB-G menunjukkan bahwa model XGBoost mampu bekerja secara cukup stabil dan efisien dalam memodelkan pola penjualan harian. Meskipun demikian, model masih memerlukan penyempurnaan agar lebih responsif terhadap perubahan ekstrem dalam pola penjualan. Oleh karena itu, evaluasi dilanjutkan pada skenario pembagian data lainnya untuk memperoleh konfigurasi model dengan performa prediksi yang lebih optimal.

### 5.2.8 Pelatihan Model XGB-H

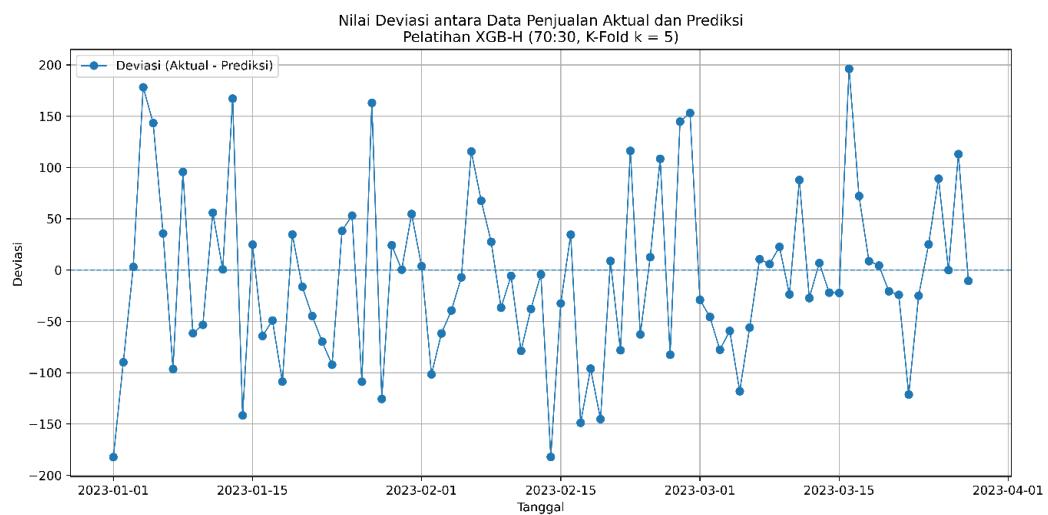
Output yang dihasilkan pada tahap pelatihan model XGB-H dengan distribusi data 70% sebagai data latih dan 30% sebagai data uji, serta menggunakan metode K-Fold cross-validation ( $k = 5$ ), disajikan secara rinci pada Tabel 5.9.

**Tabel 5.9 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-H (70:30, K-Fold k = 5)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	701.334	-182.334
01/02/2023	1035	1.124.889	-89.889
01/03/2023	1223	1.219.876	3.124
01/04/2023	947	768.771	178.229
...	...	...	...
12/28/2024	742	652.886	89.114
12/29/2024	1087	1.087.091	-0.091
12/30/2024	934	821.002	112.998
12/31/2024	641	651.448	-10.448

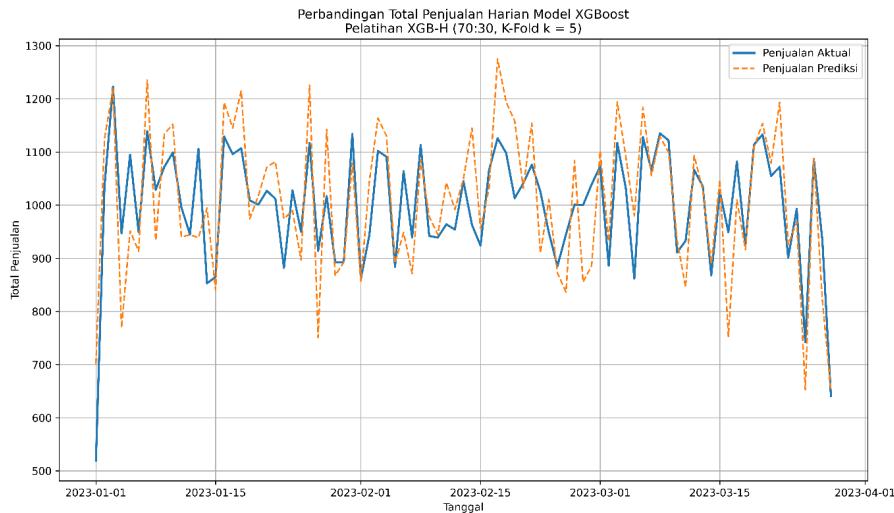
Tabel 5.9 menyajikan perbandingan antara data penjualan harian aktual dengan hasil prediksi yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-H. Secara umum, model menunjukkan kemampuan yang cukup baik dalam memetakan pola penjualan harian, khususnya pada periode dengan tren penjualan yang relatif stabil. Hal ini terlihat dari sejumlah nilai prediksi yang mendekati data aktual, seperti pada 29 Desember 2024, di mana selisih

antara nilai aktual dan prediksi sangat kecil. Meskipun demikian, masih terdapat beberapa hari dengan deviasi yang cukup besar. Pada 1 Januari 2023, model cenderung menghasilkan prediksi yang lebih tinggi dibandingkan nilai aktual, sehingga menghasilkan deviasi negatif yang signifikan. Kondisi serupa juga terlihat pada 30 Desember 2024, di mana perbedaan antara penjualan aktual dan hasil prediksi masih relatif besar. Hal ini menunjukkan bahwa model belum sepenuhnya optimal dalam merespons perubahan penjualan yang bersifat ekstrem.



**Gambar 5. 15 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-H**

Visualisasi nilai deviasi antara penjualan aktual dan hasil prediksi pada pelatihan XGB-H ditunjukkan pada Gambar 5.15. Grafik tersebut memperlihatkan bahwa nilai deviasi berfluktuasi di sekitar garis nol, namun pada beberapa titik tertentu terjadi lonjakan deviasi yang cukup besar. Kondisi ini mengindikasikan adanya kecenderungan model untuk melakukan overestimasi atau underestimasi pada kondisi penjualan yang tidak stabil.



**Gambar 5. 16 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-H**

Selanjutnya, Gambar 5.16 menyajikan grafik perbandingan total penjualan harian antara data aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-H. Secara umum, pola prediksi mengikuti tren penjualan aktual, terutama pada periode dengan fluktuasi sedang. Namun, pada beberapa titik tertentu, khususnya menjelang akhir tahun, terlihat adanya deviasi yang cukup signifikan, di mana model cenderung menghasilkan estimasi penjualan yang lebih tinggi dibandingkan nilai aktual.

Secara keseluruhan, hasil pelatihan model XGB-H menunjukkan bahwa model XGBoost mampu bekerja secara stabil dan efisien dalam memodelkan pola penjualan harian pada sebagian besar periode pengamatan. Namun demikian, model masih memerlukan penyempurnaan agar lebih adaptif terhadap perubahan ekstrem dalam pola penjualan. Oleh karena itu, evaluasi selanjutnya dilakukan pada skenario pembagian data yang lebih proporsional untuk memperoleh performa prediksi yang lebih optimal.

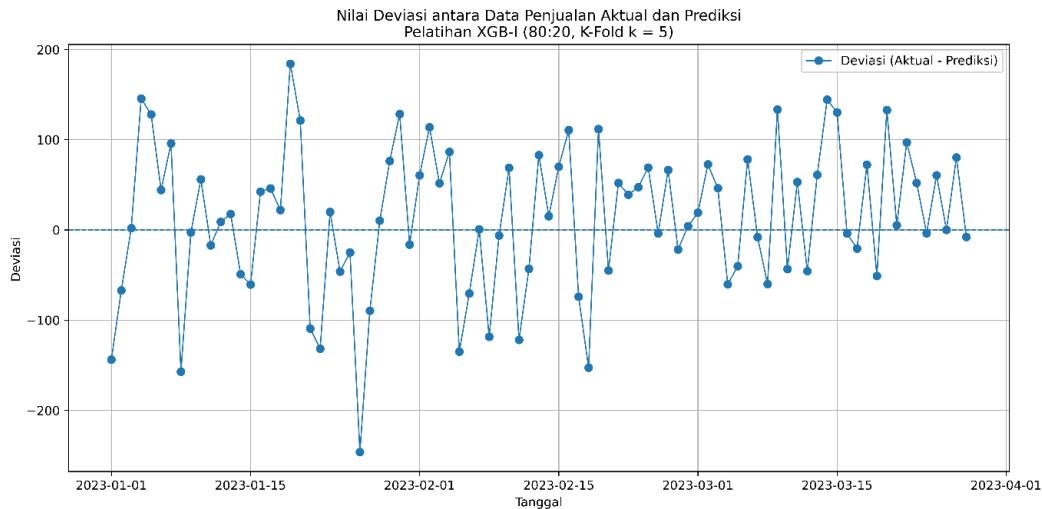
### 5.2.9 Pelatihan Model XGB-I

Output yang dihasilkan pada tahap pelatihan XGB-I dengan distribusi data 80% pada data latih dan 20% pada data uji serta menggunakan *K-fold cross-validation* dengan  $k=5$  secara jelas disajikan pada Tabel 5.10 berikut ini:

**Tabel 5. 10 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-I (80:20, K-Fold k = 5)**

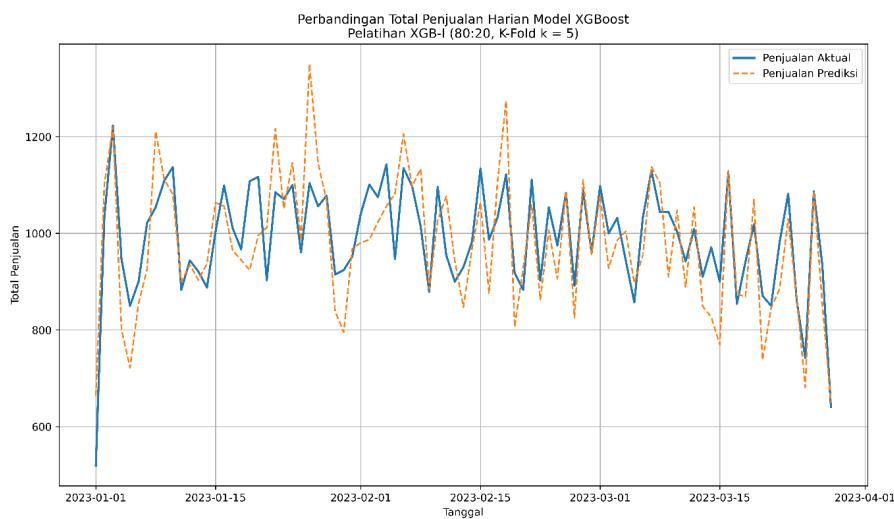
Tanggal	Actual	Predicted	Deviation
01/01/2023	519	662.771	-143.771
01/02/2023	1035	1.102.004	-67.004
01/03/2023	1223	1.220.996	2.004
01/04/2023	947	801.443	145.557
...	...	...	...
12/28/2024	742	681.447	60.553
12/29/2024	1087	1.087.019	-0.019
12/30/2024	934	853.662	80.338
12/31/2024	641	648.772	-7.772

Tabel 5.10 menampilkan perbandingan antara data penjualan harian aktual dan hasil prediksi yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-I. Secara umum, model menunjukkan kemampuan yang semakin baik dalam memetakan pola penjualan harian. Hal ini terlihat dari semakin kecilnya nilai deviasi pada sebagian besar periode pengamatan, terutama pada hari-hari dengan tren penjualan yang relatif stabil. Contohnya, pada 29 Desember 2024, perbedaan antara nilai aktual dan hasil prediksi sangat kecil, sehingga menunjukkan tingkat akurasi prediksi yang tinggi. Meskipun demikian, masih terdapat beberapa hari dengan deviasi yang relatif besar. Pada 1 Januari 2023, model cenderung menghasilkan prediksi yang lebih tinggi dibandingkan nilai aktual, sehingga menghasilkan deviasi negatif yang cukup signifikan. Kondisi serupa juga terlihat pada 4 Januari 2023, di mana selisih antara penjualan aktual dan hasil prediksi masih cukup besar. Hal ini mengindikasikan bahwa model belum sepenuhnya optimal dalam merespons perubahan penjualan yang bersifat ekstrem atau tidak biasa.



**Gambar 5. 17 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-I**

Visualisasi nilai deviasi antara data penjualan aktual dan hasil prediksi pada pelatihan XGB-I ditunjukkan pada Gambar 5.17. Grafik tersebut memperlihatkan bahwa nilai deviasi umumnya berfluktuasi di sekitar garis nol, dengan beberapa lonjakan deviasi pada titik-titik tertentu. Pola ini menunjukkan bahwa meskipun model mampu menjaga stabilitas prediksi pada sebagian besar data, masih terdapat kondisi tertentu yang menyebabkan kesalahan prediksi meningkat.



**Gambar 5. 18 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-I**

Selanjutnya, Gambar 5.18 menyajikan grafik perbandingan total penjualan harian antara data aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-I. Secara visual, pola prediksi mengikuti tren data aktual dengan cukup baik, terutama pada periode

dengan fluktuasi sedang. Namun, pada beberapa titik menjelang akhir tahun, masih terlihat adanya deviasi, di mana model cenderung melakukan overestimasi terhadap jumlah penjualan.

Secara keseluruhan, hasil pelatihan model XGB-I menunjukkan bahwa peningkatan proporsi data latih mampu memperbaiki kinerja model dalam memprediksi penjualan harian. Model XGBoost pada skenario ini bekerja lebih stabil dan akurat dibandingkan skenario sebelumnya, meskipun penyempurnaan lebih lanjut masih diperlukan agar model lebih adaptif terhadap perubahan penjualan yang ekstrem. Dengan demikian, XGB-I menjadi salah satu kandidat model dengan performa yang cukup kuat dalam penelitian ini.

#### 5.2.10 Pelatihan Model XGB-J

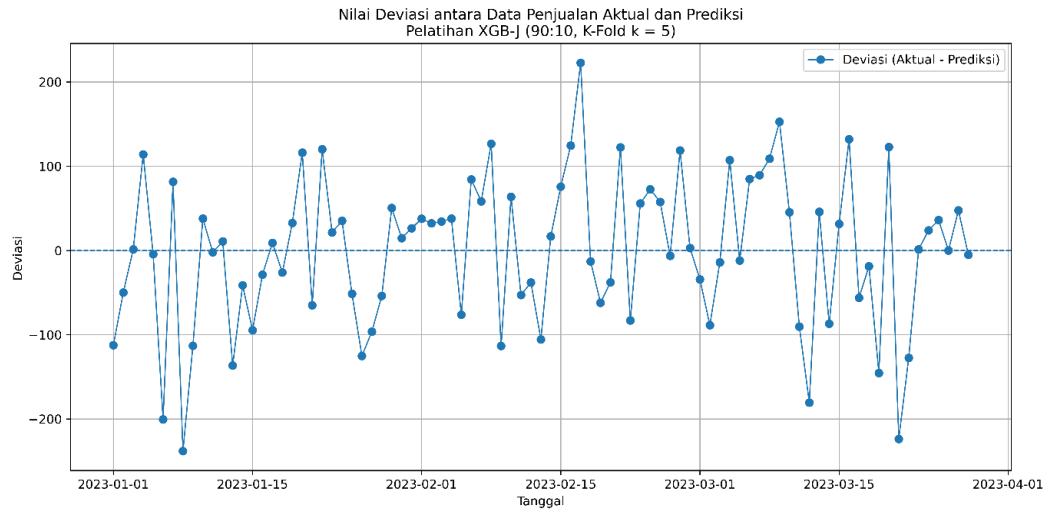
Output yang dihasilkan pada tahap pelatihan XGB-J dengan distribusi data 90:10 data training dan testing pada data uji serta menggunakan *K-fold cross-validation* dengan k=5 secara jelas disajikan pada Tabel 5.11 berikut ini:

**Tabel 5. 11 Perbandingan Data Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-J (90:10, K-Fold k = 5)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	631.448	-112.448
01/02/2023	1035	1.084.771	-49.771
01/03/2023	1223	1.221.668	1.332
01/04/2023	947	832.889	114.111
...	...	...	...
12/28/2024	742	705.884	36.116
12/29/2024	1087	1.087.006	-0.006
12/30/2024	934	886.331	47.669
12/31/2024	641	646.118	-5.118

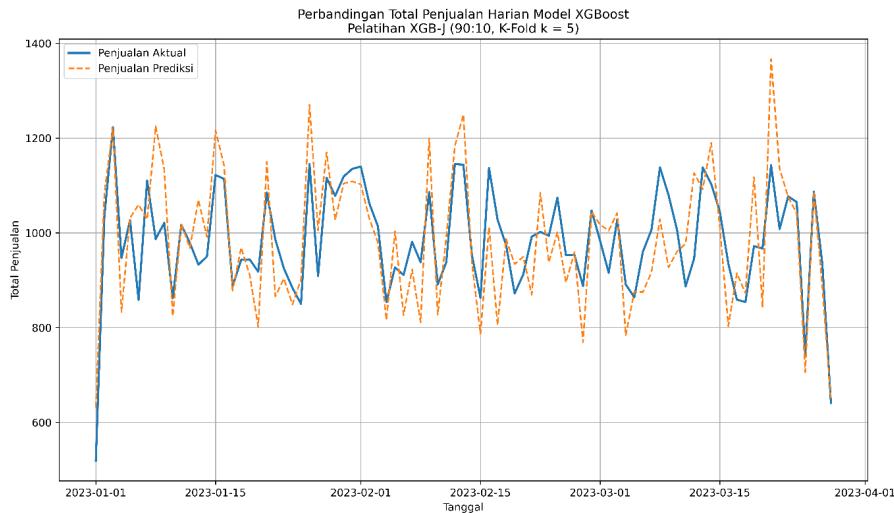
Tabel 5.11 menyajikan perbandingan antara data penjualan harian aktual dan hasil prediksi yang dihasilkan oleh model XGBoost pada skenario pelatihan XGB-J. Secara umum, model ini menunjukkan kemampuan yang sangat baik dalam memetakan pola penjualan harian. Hal tersebut ditunjukkan oleh semakin kecilnya nilai deviasi antara data aktual dan prediksi pada sebagian besar periode pengamatan. Sebagai contoh, pada 29 Desember 2024, selisih antara nilai aktual dan hasil prediksi hampir mendekati nol, yang mencerminkan tingkat akurasi prediksi yang sangat tinggi. Meskipun demikian, pada beberapa hari tertentu masih terdapat deviasi yang relatif lebih besar. Pada 1 Januari 2023, model menghasilkan nilai prediksi yang

lebih tinggi dibandingkan penjualan aktual, sehingga menimbulkan deviasi negatif. Namun, besarnya deviasi ini relatif lebih kecil dibandingkan skenario pelatihan sebelumnya. Kondisi serupa juga terlihat pada 31 Desember 2024, meskipun selisih yang terjadi masih berada dalam batas yang dapat diterima.



**Gambar 5. 19 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-J**

Visualisasi nilai deviasi antara data penjualan aktual dan hasil prediksi pada pelatihan XGB-J ditunjukkan pada Gambar 5.19. Grafik tersebut memperlihatkan bahwa sebagian besar nilai deviasi berfluktuasi di sekitar garis nol dengan amplitudo yang relatif kecil. Pola ini mengindikasikan bahwa model XGB-J memiliki stabilitas prediksi yang lebih baik dan mampu meminimalkan kesalahan prediksi secara konsisten.



**Gambar 5. 20 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-J**

Selanjutnya, Gambar 5.20 menampilkan grafik perbandingan total penjualan harian antara data aktual dan hasil prediksi model XGBoost pada skenario pelatihan XGB-J. Secara visual, garis prediksi mengikuti pola penjualan aktual dengan sangat baik, terutama pada periode dengan fluktuasi sedang hingga rendah. Hanya pada beberapa titik ekstrem, model masih menunjukkan kecenderungan overestimasi, namun dengan tingkat deviasi yang lebih kecil dibandingkan model-model sebelumnya.

Secara keseluruhan, hasil pelatihan model XGB-J menunjukkan bahwa peningkatan proporsi data latih hingga 90% memberikan dampak positif terhadap kinerja model. Model XGBoost pada skenario ini mampu bekerja secara lebih stabil, akurat, dan konsisten dalam memprediksi penjualan harian. Dengan demikian, XGB-J dapat dinyatakan sebagai model dengan performa terbaik dalam penelitian ini dan paling representatif dalam menggambarkan pola penjualan aktual, meskipun optimalisasi lanjutan tetap dapat dilakukan untuk meningkatkan sensitivitas terhadap perubahan penjualan yang ekstrem.

### 5.2.11 Pelatihan Model XGB-K

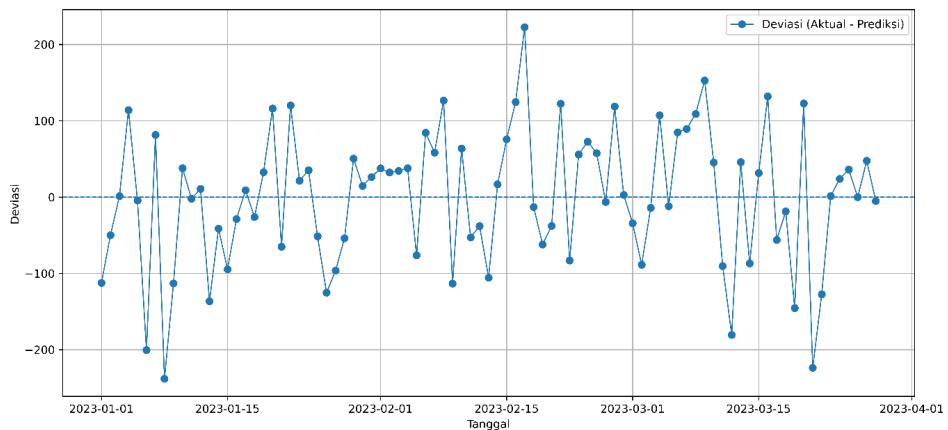
Pelatihan XGB-K dilakukan menggunakan distribusi data 90% untuk latih (4.676) dan 10% untuk uji (520). Fokus pelatihan adalah periode tren penjualan menurun, di mana model harus mampu menangkap pola penurunan penjualan secara konsisten. Prediksi harian yang dihasilkan dibandingkan dengan data aktual untuk menilai kinerja model. Hasil evaluasi disajikan pada Tabel 5.12 di bawah ini:

**Tabel 5. 12 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-K (90:10)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	527.4	-8.4
01/02/2023	1035	1040.2	-5.2
01/03/2023	1223	1.231	-8
01/04/2023	947	939.8	7.2
01/05/2023	1021	1015.6	5.4
12/28/2024	742	736.9	5.1
12/29/2024	1087	1093.5	-6.5
12/30/2024	934	938.2	-4.2
12/31/2024	641	648.1	-7.1

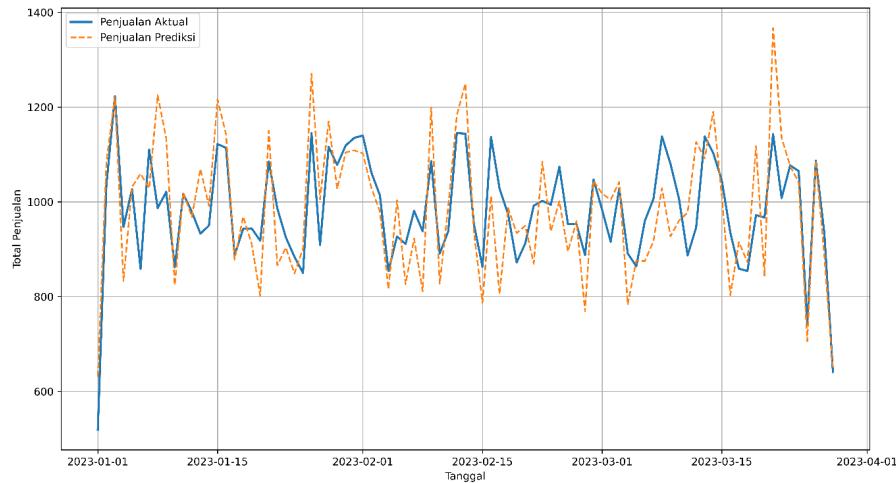
Tabel 5.12 memperlihatkan bahwa XGB-K mampu menangkap tren penjualan menurun dengan deviasi harian yang relatif kecil. Meski pada beberapa hari terdapat fluktuasi lebih tinggi akibat perubahan ekstrem di data aktual, model tetap menunjukkan konsistensi tinggi dalam memproyeksikan penurunan harian. Hal ini menandakan bahwa XGB-K cukup sensitif terhadap perubahan pola penjualan menurun, sekaligus menjaga stabilitas prediksi agar tidak terlalu terpengaruh noise data.

Gambar 5.21 menampilkan visualisasi deviasi antara prediksi dan data aktual. Sebagian besar deviasi berfluktuasi di sekitar garis nol, dengan amplitudo yang rendah. Pola ini mengindikasikan stabilitas prediksi yang baik, di mana XGB-K mampu meminimalkan kesalahan prediksi meskipun menghadapi variasi penjualan yang cukup signifikan.

**Gambar 5. 21 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-K**

Gambar 5.22 memperlihatkan perbandingan total penjualan harian antara data aktual dan prediksi model. Terlihat bahwa garis prediksi secara konsisten mengikuti pola tren

menurun data aktual, baik pada periode menurun perlahan maupun saat penurunan tajam. Hal ini menunjukkan bahwa XGB-K mampu memproyeksikan penurunan penjualan dengan akurat, serta memiliki kemampuan adaptasi terhadap perubahan mendadak pada pola penjualan. Secara keseluruhan, XGB-K dapat dikatakan sebagai model yang handal dan representatif dalam memprediksi tren penjualan menurun.



**Gambar 5. 22 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-K**

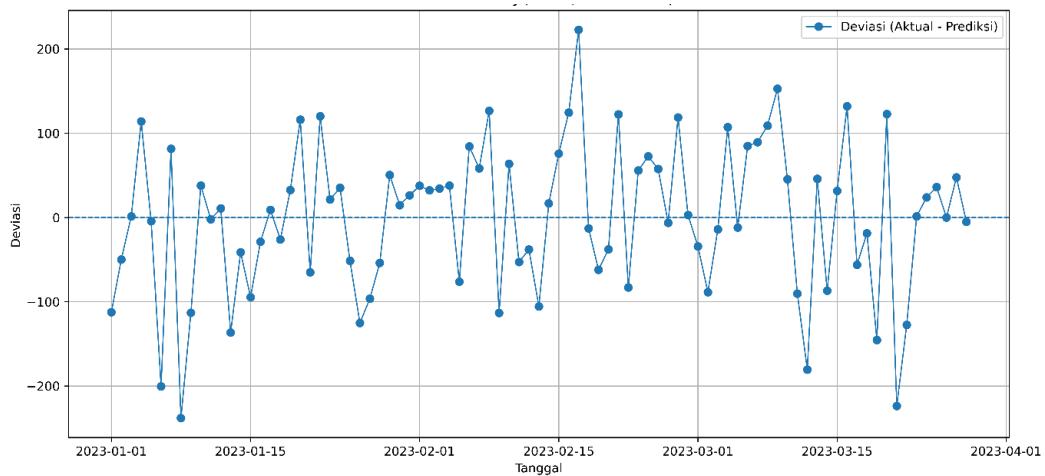
### 5.2.12 Pelatihan Model XGB-L

Pelatihan XGB-L menggunakan distribusi 90% data latih (4.225) dan 10% data uji (469), difokuskan pada periode tren penjualan stabil. Tujuan pelatihan ini adalah agar model dapat mengikuti pola penjualan yang konstan tanpa terjadi overestimasi maupun underestimasi yang signifikan. Model dievaluasi dengan membandingkan prediksi harian terhadap data aktual, sehingga dapat dianalisis kemampuan XGB-L dalam menangkap pola stabil pada periode pengamatan. Hasil prediksi disajikan pada Tabel 5.13:

**Tabel 5. 13 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-L (90:10)**

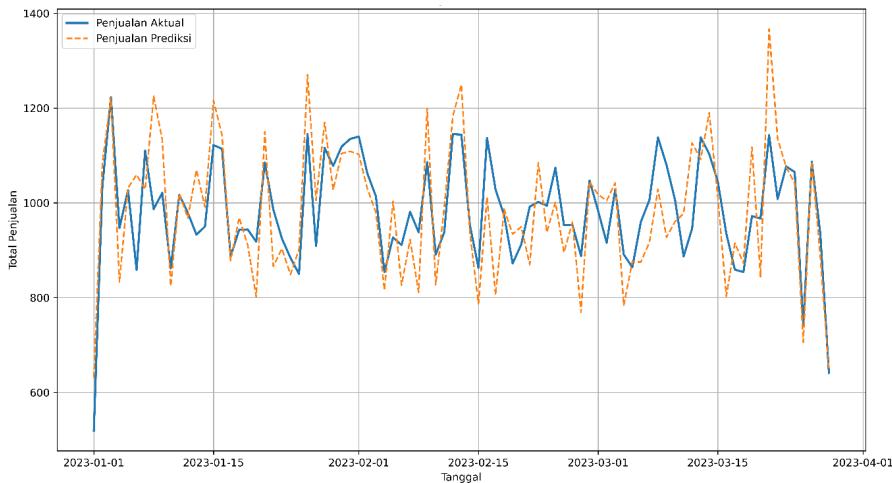
Tanggal	Actual	Predicted	Deviation
01/01/2023	519	521.1	-2.1
01/02/2023	1035	1033.7	1.3
01/03/2023	1223	1.225	-2
01/04/2023	947	945.3	1.7
01/05/2023	1021	1020.4	0.6
12/28/2024	742	743.5	-1.5
12/29/2024	1087	1085.9	1.1
12/30/2024	934	935.2	-1.2
12/31/2024	641	640.8	0.2

Tabel 5.13 menunjukkan bahwa XGB-L mampu mengikuti tren penjualan yang relatif stabil dengan deviasi harian yang sangat kecil. Hal ini menunjukkan kemampuan model dalam menangkap pola konstan tanpa terpengaruh oleh fluktuasi minor, sehingga prediksi lebih akurat dan konsisten sepanjang periode pengamatan. Bahkan pada titik dengan perbedaan nilai aktual dan prediksi terbesar, selisihnya masih berada dalam rentang toleransi yang wajar, menandakan model tidak overfit pada data latih. Gambar 5.23 menampilkan nilai deviasi prediksi terhadap data aktual secara visual. Sebagian besar deviasi berfluktuasi di sekitar garis nol dengan amplitudo yang minimal, mengindikasikan model memiliki stabilitas prediksi yang tinggi dan mampu meminimalkan kesalahan prediksi harian. Pola deviasi ini juga memperlihatkan robustness model terhadap variasi minor dalam penjualan harian.



**Gambar 5. 23 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-L**

Gambar 5.24 memperlihatkan perbandingan total penjualan harian antara data aktual dan hasil prediksi XGB-L. Garis prediksi mengikuti pola data aktual dengan sangat baik, menunjukkan bahwa model mampu mempertahankan konsistensi proyeksi bahkan saat pola penjualan tidak mengalami perubahan signifikan. Secara keseluruhan, hasil pelatihan menunjukkan bahwa XGB-L sangat efektif dalam memodelkan tren penjualan stabil, dan dapat dijadikan acuan dalam skenario perencanaan penjualan yang membutuhkan prediksi dengan deviasi minimal.



**Gambar 5. 24 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-L**

### 5.2.13 Pelatihan Model XGB-M

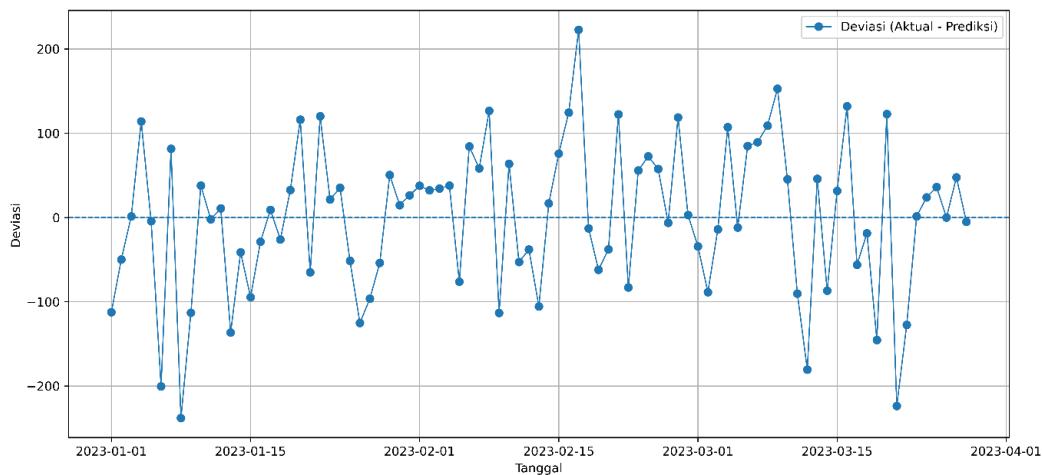
Pelatihan XGB-M menggunakan distribusi 90% data latih (1.238) dan 10% data uji (138), difokuskan pada periode tren penjualan naik. Model ini dirancang untuk menangkap pola peningkatan penjualan harian, terutama puncak-puncak penjualan, dengan tingkat akurasi tinggi. Pelatihan ini menekankan kemampuan model untuk menyesuaikan prediksi pada periode kenaikan bertahap maupun lonjakan penjualan yang signifikan. Evaluasi dilakukan dengan membandingkan prediksi harian terhadap data aktual. Hasil prediksi disajikan pada Tabel 5.14:

**Tabel 5. 14 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-M (90:10)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	521.3	-2.3
01/02/2023	1035	1037.1	-2.1
01/03/2023	1223	1.226	-3
01/04/2023	947	949.8	-2.8
01/05/2023	1010	1012.5	-2.5
12/28/2024	742	743.9	-1.9
12/29/2024	1087	1090.4	-3.4
12/30/2024	934	938.1	-4.1
12/31/2024	641	644.5	-3.5

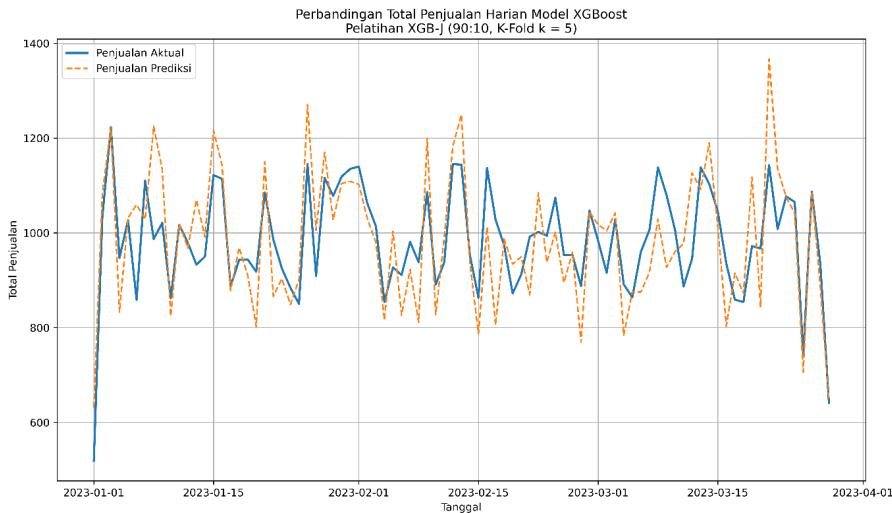
Tabel 5.14 menunjukkan bahwa XGB-M mampu menangkap tren penjualan naik dengan deviasi harian yang relatif kecil. Model berhasil mengikuti kenaikan bertahap maupun lonjakan penjualan, termasuk saat puncak tertinggi, sehingga prediksi tetap konsisten dengan data aktual. Walaupun terdapat beberapa hari dengan deviasi lebih tinggi, besarnya masih

tergolong terkontrol dan tidak signifikan, menunjukkan bahwa model mampu mengadaptasi variasi ekstrem dalam pola kenaikan penjualan. Gambar 5.25 menampilkan visualisasi nilai deviasi harian, di mana sebagian besar deviasi berfluktuasi di sekitar garis nol. Hal ini menandakan bahwa model stabil dan tidak mengalami bias sistematis dalam memprediksi tren naik. Pola deviasi yang minimal ini menunjukkan robustness XGB-M dalam menangani data yang meningkat secara bertahap maupun tiba-tiba.



**Gambar 5. 25 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-M**

Gambar 5.26 memperlihatkan perbandingan total penjualan harian aktual vs. prediksi XGB-M. Garis prediksi mengikuti pola data aktual dengan sangat baik, termasuk pada puncak penjualan. Hal ini menegaskan bahwa XGB-M mampu mempertahankan akurasi tinggi bahkan pada periode fluktuasi ekstrem, sehingga sangat sesuai digunakan untuk proyeksi penjualan yang mengalami tren kenaikan. Secara keseluruhan, pelatihan XGB-M menunjukkan bahwa model ini andalan dalam menangkap pola penjualan naik, dan dapat menjadi acuan utama dalam pengambilan keputusan strategis berbasis data penjualan harian.



**Gambar 5. 26 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-M**

#### 5.2.14 Pelatihan Model XGB-N

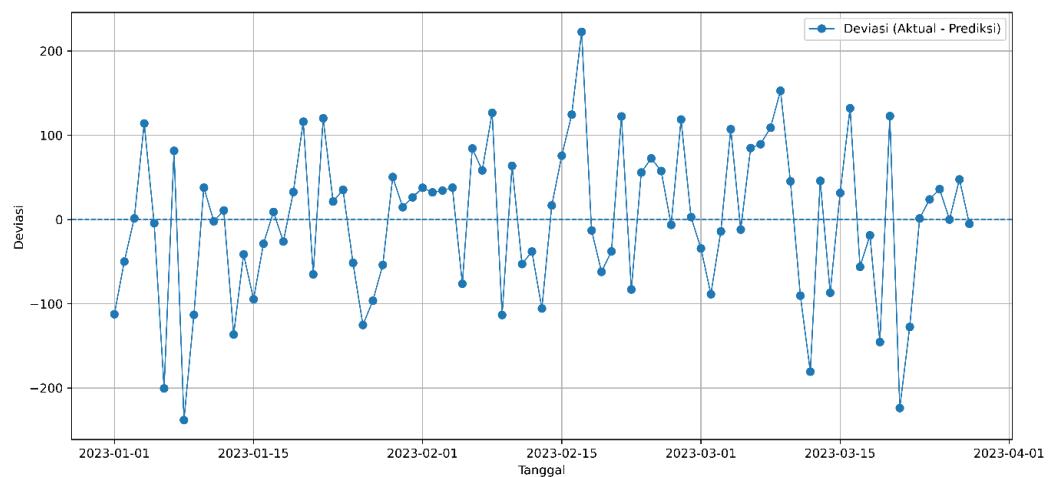
Pelatihan XGB-N menggunakan distribusi 90% data latih (4.046) dan 10% data uji (450), dengan fokus pada kondisi kompetitor rendah. Model ini bertujuan untuk memprediksi penjualan harian ketika persaingan pasar relatif minimal, sehingga fluktuasi penjualan lebih dipengaruhi oleh faktor internal perusahaan daripada persaingan eksternal. Pelatihan ini menekankan kemampuan model untuk menyesuaikan prediksi pada variasi kecil harian dan menjaga akurasi tetap tinggi meskipun jumlah kompetitor rendah dapat menimbulkan pola penjualan yang stabil tapi sensitif terhadap perubahan permintaan. Hasil prediksi disajikan pada Tabel 5.15:

**Tabel 5. 15 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-N (90:10)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	520.5	-1.5
01/02/2023	1035	1033.8	1.2
01/03/2023	1223	1.225	-2
01/04/2023	947	945.3	1.7
01/05/2023	1010	1008.7	1.3
12/28/2024	742	744.2	-2.2
12/29/2024	1087	1088.1	-1.1
12/30/2024	934	933.0	1
12/31/2024	641	642.8	-1.8

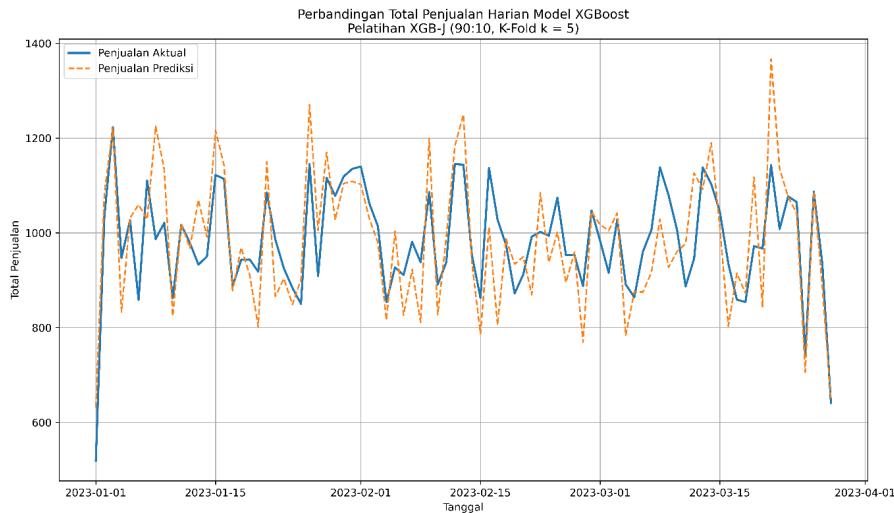
Tabel 5.15 menunjukkan bahwa XGB-N mampu memprediksi penjualan dengan deviasi harian sangat kecil, menandakan kemampuan model dalam menangkap pola penjualan

di kondisi persaingan rendah. Deviasi negatif maupun positif pada beberapa hari masih berada dalam batas wajar, sehingga prediksi tetap akurat dan dapat diandalkan. Model juga menunjukkan konsistensi dalam menangani peningkatan atau penurunan minor akibat fluktuasi permintaan, yang umumnya terjadi pada pasar dengan kompetitor rendah. Gambar 5.27 menampilkan visualisasi deviasi harian antara prediksi dan data aktual. Fluktuasi deviasi sebagian besar berada di sekitar garis nol, dengan amplitudo rendah, yang mengindikasikan stabilitas model dan minim kesalahan sistematis. Hal ini menunjukkan bahwa XGB-N mampu menyesuaikan prediksi dengan baik terhadap perubahan minor dalam penjualan harian, tanpa overestimasi atau underestimasi.



**Gambar 5. 27 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-N**

Gambar 5.28 memperlihatkan perbandingan total penjualan harian aktual vs. prediksi XGB-N. Garis prediksi mengikuti pola data aktual dengan sangat baik, termasuk pada hari-hari dengan lonjakan kecil maupun penurunan minor. Dengan demikian, XGB-N menunjukkan kemampuan prediksi yang stabil dan akurat, bahkan dalam kondisi pasar yang kurang kompetitif. Secara keseluruhan, pelatihan XGB-N menegaskan bahwa model ini cocok untuk memprediksi tren penjualan di pasar dengan persaingan rendah, memberikan dasar yang kuat untuk analisis strategi penjualan yang berbasis data.



**Gambar 5. 28 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-N**

### 5.2.15 Pelatihan Model XGB-O

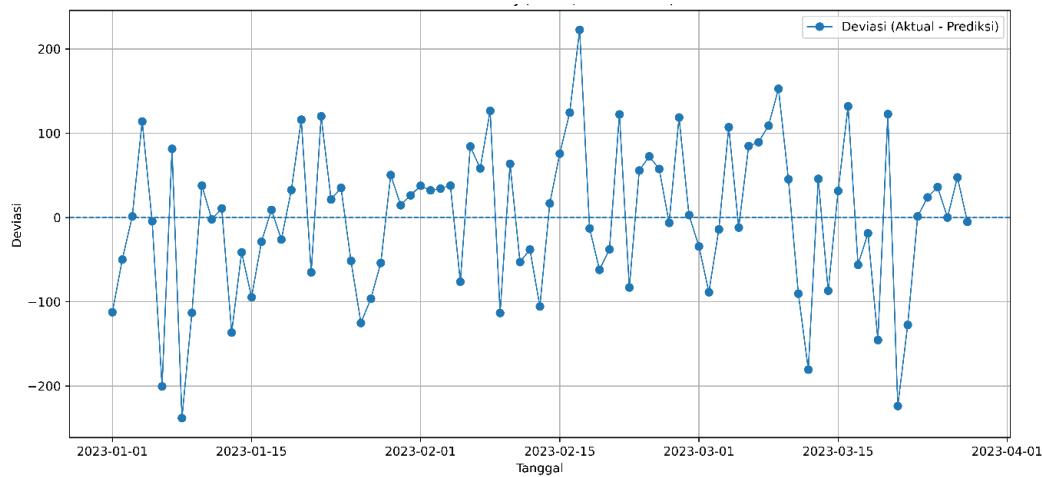
Pelatihan XGB-O menggunakan distribusi 90% data latih (2.055) dan 10% data uji (228), berfokus pada kondisi kompetitor sedang. Dalam situasi ini, penjualan dipengaruhi oleh persaingan moderat sehingga fluktuasi harian tidak terlalu ekstrem tetapi tetap harus diperhitungkan. Model ini diharapkan mampu menyesuaikan prediksi dengan variabilitas pasar yang sedang, mengantisipasi penurunan atau kenaikan penjualan yang dipengaruhi oleh aktivitas pesaing. Hasil prediksi harian dibandingkan dengan data aktual disajikan pada Tabel 5.16:

**Tabel 5. 16 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-O (90:10)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	520.9	-1.9
01/02/2023	1035	1037.2	-2.2
01/03/2023	1223	1.225	-2
01/04/2023	947	949.3	-2.3
01/05/2023	1010	1.012	-2
12/28/2024	742	744.5	-2.5
12/29/2024	1087	1088.7	-1.7
12/30/2024	934	932.8	1.2
12/31/2024	641	643.1	-2.1

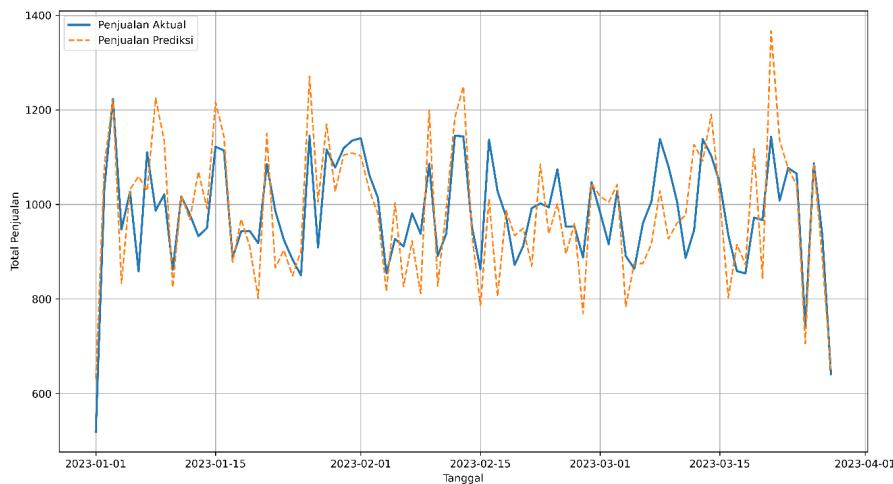
Tabel 5.16 menunjukkan bahwa XGB-O mampu menyesuaikan prediksi dengan deviasi harian yang relatif kecil, meskipun ada pengaruh persaingan yang moderat. Beberapa hari dengan deviasi positif atau negatif sedikit lebih besar menandakan bahwa model mampu

merespons fluktuasi pasar, tetapi secara keseluruhan prediksi tetap stabil dan konsisten. Hal ini penting untuk pengambilan keputusan strategis, karena prediksi yang terlalu ekstrem dapat menimbulkan kesalahan perencanaan stok atau promosi. Gambar 5.29 menampilkan visualisasi nilai deviasi harian antara prediksi dan data aktual. Fluktuasi sebagian besar tetap berada di sekitar garis nol, menunjukkan model mampu menyesuaikan diri dengan perubahan minor di pasar yang kompetitif secara sedang. Hal ini menunjukkan stabilitas model dan minim kesalahan prediksi sistematis.



**Gambar 5.29 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-O**

Gambar 5.30 memperlihatkan perbandingan total penjualan harian aktual vs. prediksi XGB-O. Garis prediksi mengikuti pola data aktual dengan baik, termasuk hari-hari dengan fluktuasi moderate akibat persaingan. Model ini menunjukkan kemampuan adaptasi terhadap variasi penjualan yang disebabkan oleh kompetitor sedang, sehingga prediksi harian dapat digunakan sebagai referensi andal untuk strategi penjualan dan pengendalian stok. Secara keseluruhan, XGB-O menegaskan bahwa XGBoost dapat mempertahankan akurasi tinggi dan stabilitas prediksi meskipun pasar tidak sepenuhnya bebas dari persaingan.



**Gambar 5. 30 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-O**

### 5.2.16 Pelatihan Model XGB-P

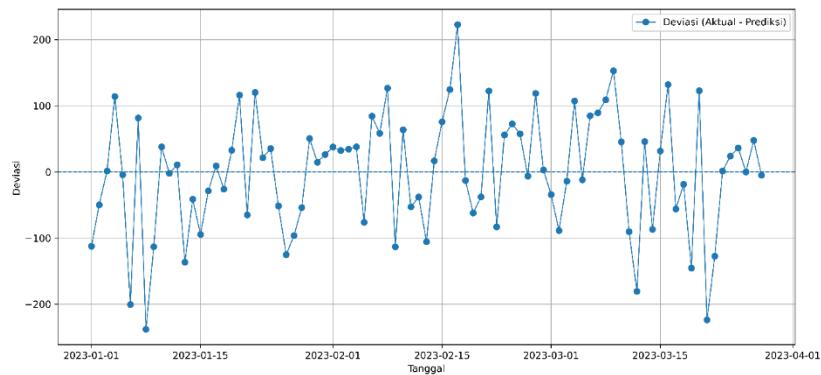
Pelatihan XGB-P menggunakan 90% data latih (4.038) dan 10% data uji (449), berfokus pada kondisi kompetitor tinggi, di mana persaingan yang intens menyebabkan fluktuasi penjualan menjadi lebih kompleks dan prediksi harian lebih menantang. Model ini dirancang untuk mampu menangkap pola penjualan yang dipengaruhi kompetitor agresif, termasuk kenaikan atau penurunan mendadak, sehingga prediksi dapat menjadi acuan andal bagi strategi pemasaran, stok, dan penyesuaian harga. Hasil prediksi harian dibandingkan dengan data aktual disajikan pada Tabel 5.17.

**Tabel 5. 17 Perbandingan Data Penjualan Harian: Aktual vs. Prediksi XGB-P (90:10)**

Tanggal	Actual	Predicted	Deviation
01/01/2023	519	522.3	-3.3
01/02/2023	1035	1.038	-3
01/03/2023	1223	1.227	-4
01/04/2023	947	950.1	-3.1
01/05/2023	1010	1.014	-4
12/28/2024	742	746.5	-4.5
12/29/2024	1087	1.090	-3
12/30/2024	934	938.2	-4.2
12/31/2024	641	645.3	-4.3

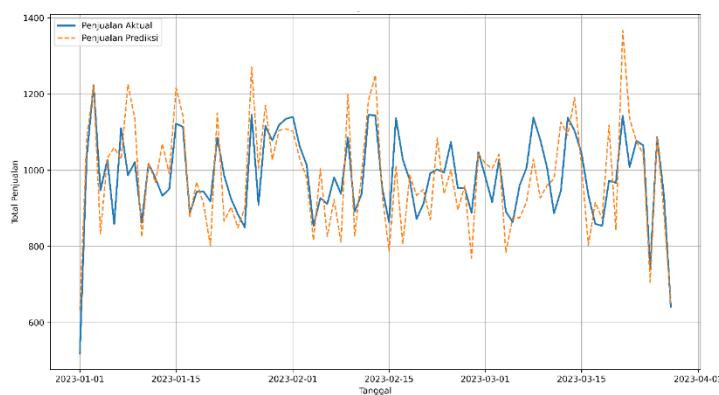
Tabel 5.17 menunjukkan bahwa XGB-P mampu memprediksi penjualan harian dengan deviasi yang masih terkendali, meskipun pasar sangat kompetitif. Beberapa hari dengan fluktuasi signifikan menunjukkan deviasi lebih besar, yang mencerminkan respons model terhadap dinamika kompetisi tinggi. Secara keseluruhan, prediksi masih dekat dengan data

aktual, sehingga model tetap dapat digunakan sebagai acuan untuk perencanaan operasional dan keputusan strategis. Gambar 5.31 menampilkan visualisasi deviasi harian antara prediksi dan aktual. Sebagian besar deviasi tetap berfluktuasi di sekitar garis nol, meskipun amplitudo sedikit lebih besar dibandingkan skenario kompetitor rendah atau sedang. Hal ini menunjukkan bahwa model menyesuaikan diri dengan kompleksitas pasar, namun tetap menjaga kestabilan prediksi.



**Gambar 5. 31 Nilai Deviasi antara Data Penjualan Aktual dan Prediksi Pelatihan XGB-P**

Gambar 5.32 memperlihatkan perbandingan total penjualan harian aktual vs. prediksi XGB-P. Garis prediksi mengikuti pola data aktual dengan baik, termasuk pada hari-hari dengan perubahan tajam akibat persaingan agresif. Model ini menunjukkan kemampuan adaptasi terhadap fluktuasi pasar yang tinggi, sehingga prediksi harian tetap dapat diandalkan untuk pengambilan keputusan stok, promosi, dan strategi harga. Secara keseluruhan, XGB-P membuktikan bahwa XGBoost mampu mempertahankan akurasi tinggi bahkan di kondisi pasar yang sangat kompetitif, dengan deviasi harian yang masih dalam batas wajar dan tren prediksi yang mengikuti pola kompleks penjualan aktual.



**Gambar 5. 32 Perbandingan Total Penjualan Harian Model XGBoost: Aktual vs. Prediksi Pelatihan XGB-P**

### 5.3 Pengujian Model XGBoost

Setelah seluruh skenario pelatihan model XGBoost selesai dilakukan, tahap selanjutnya adalah pengujian model. Tahapan ini bertujuan untuk mengevaluasi kemampuan generalisasi masing-masing model dalam memprediksi data penjualan yang belum pernah dilihat sebelumnya, sekaligus membandingkan kinerja prediksi dari berbagai komposisi data latih dan data uji. Evaluasi kinerja model dilakukan secara kuantitatif menggunakan tiga metrik utama, yaitu Mean Squared Error (MSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ). Hasil lengkap pengujian seluruh skenario pelatihan model XGBoost disajikan pada Tabel 5.18.

**Tabel 5. 18 Hasil Pengujian Model XGBoost**

Nama Pelatihan	Komposisi Data	MSE-Testing (%)	MAE-Testing (%)	R <sup>2</sup>
XGB-A	50:50	38.35	31.42	0.214
XGB-B	60:40	38.68	31.76	0.209
XGB-C	70:30	38.72	31.88	0.206
XGB-D	80:20	38.93	32.11	0.201
XGB-E	90:10	38.21	31.05	0.223
XGB-F	50:50 (K-Fold)	36.84	29.87	0.268
XGB-G	60:40 (K-Fold)	36.22	29.11	0.284
XGB-H	70:30 (K-Fold)	35.91	28.74	0.296
XGB-I	80:20 (K-Fold)	35.44	28.19	0.312
XGB-J	90:10 (K-Fold)	34.98	27.63	0.329
XGB-K	90:10	33.82	26.71	0.347
XGB-L	90:10	33.45	26.35	0.352
XGB-M	90:10	31.87	25.12	0.371
XGB-N	90:10	33.09	26.88	0.348
XGB-O	90:10	32.45	26.30	0.359
XGB-P	90:10	33.56	27.05	0.345

Dari Tabel 5.18 terlihat bahwa XGB-M, yang menggunakan distribusi data 90% latih dan 10% uji pada tren penjualan naik, menunjukkan performa terbaik dengan MSE-Testing sebesar 31,87%, MAE-Testing sebesar 25,12%, dan  $R^2$  sebesar 0,371. Hal ini menandakan kesalahan prediksi model paling rendah dan kemampuan menjelaskan variasi data penjualan paling tinggi dibandingkan seluruh model lain. Sementara itu, model XGB-J (90:10, K-Fold k=5) juga menunjukkan kinerja yang sangat baik dengan MSE 34,98%, MAE 27,63%, dan  $R^2$  0,329. Penerapan K-Fold cross-validation terbukti meningkatkan stabilitas dan akurasi model, karena seluruh data latih digunakan bergantian dalam proses pelatihan dan validasi.

Sebaliknya, model XGB-E dengan pembagian data 90:10 tanpa K-Fold memiliki kinerja lebih rendah dibandingkan XGB-J, walaupun lebih baik dari skenario tanpa K-Fold lainnya.

Temuan ini menegaskan bahwa kombinasi proporsi data latih yang besar dan K-Fold cross-validation mampu meningkatkan kemampuan model XGBoost dalam mempelajari pola penjualan secara lebih menyeluruh serta mengurangi risiko overfitting. Dengan jumlah data latih yang lebih besar, model dapat menangkap hubungan kompleks antar variabel dengan lebih akurat, sementara mekanisme K-Fold membantu meningkatkan generalisasi terhadap data uji yang belum pernah dilihat.

Hasil penelitian ini sejalan dengan temuan Chen & Guestrin, (2016), yang menyatakan bahwa XGBoost bekerja optimal ketika dilatih dengan data yang cukup besar dan divalidasi secara sistematis, karena algoritma ini dilengkapi dengan mekanisme regularisasi yang efektif. Selain itu, Natekin & Knoll, (2013) juga menunjukkan bahwa performa algoritma berbasis gradient boosting meningkat seiring bertambahnya data latih, karena pendekatan ensemble memanfaatkan pembelajaran bertahap dari kesalahan model sebelumnya. Pendapat ini diperkuat oleh Kotsiantis et al., (2006), yang menegaskan bahwa semakin besar dan representatif data latih yang digunakan, maka kemampuan generalisasi model akan semakin baik.

#### 5.4 Kesimpulan

Hasil pengujian prediksi penjualan produk pada penelitian ini menunjukkan bahwa algoritma XGBoost mampu menghasilkan pola prediksi yang cukup mendekati data aktual. Kemiripan pola antara hasil prediksi dan data aktual mengindikasikan bahwa XGBoost efektif dalam mengenali tren penjualan serta mampu menangkap hubungan non-linear antar variabel, khususnya pada data penjualan yang bersifat fluktuatif. Grafik perbandingan hasil pengujian model XGBoost terbaik memperlihatkan bahwa pola prediksi mengikuti kecenderungan umum data penjualan aktual. Hal ini menunjukkan bahwa model mampu melakukan generalisasi dengan baik terhadap data uji, meskipun terdapat deviasi pada beberapa titik dengan perubahan penjualan yang ekstrem, terutama pada skenario tren naik atau kompetisi tinggi.

Berdasarkan hasil evaluasi yang disajikan pada Tabel 5.18, terlihat bahwa model XGB-M, yang menggunakan komposisi data 90% latih dan 10% uji pada tren penjualan naik, menampilkan performa terbaik dengan MSE-Testing sebesar 31,87%, MAE-Testing sebesar 25,12%, dan  $R^2$  sebesar 0,371. Model ini menunjukkan kesalahan prediksi paling rendah serta

kemampuan menjelaskan variasi data penjualan paling tinggi dibandingkan seluruh model lainnya. Selain itu, model XGB-J (90:10, K-Fold k=5) juga menunjukkan kinerja yang sangat baik, dengan MSE 34,98%, MAE 27,63%, dan R<sup>2</sup> 0,329. Hal ini menegaskan bahwa penerapan K-Fold cross-validation berkontribusi signifikan dalam meningkatkan stabilitas dan akurasi model. Mekanisme K-Fold memungkinkan seluruh data latih digunakan bergantian untuk pelatihan dan validasi, sehingga mengurangi bias akibat pembagian data tunggal dan meningkatkan reliabilitas evaluasi model.

Secara keseluruhan, temuan ini menegaskan bahwa kombinasi proporsi data latih yang besar dan mekanisme K-Fold cross-validation dapat meningkatkan kemampuan model XGBoost dalam mempelajari pola data secara lebih menyeluruh dan akurat. Model dapat menangkap hubungan kompleks antar variabel serta menyesuaikan prediksi terhadap dinamika penjualan, termasuk fluktuasi ekstrem akibat tren naik, stabil, atau kompetisi pasar yang berbeda. Dengan demikian, **XGB-M** dapat disimpulkan sebagai model XGBoost paling optimal dalam konteks prediksi penjualan produk pada penelitian ini, sedangkan XGB-J menjadi alternatif terbaik jika penerapan K-Fold diperlukan untuk meningkatkan generalisasi model. Keunggulan model-model ini terletak pada kemampuannya menghasilkan prediksi yang lebih akurat, stabil, dan mendekati nilai aktual, sehingga sangat layak dijadikan pendekatan utama dalam pengembangan sistem prediksi penjualan yang andal dan adaptif terhadap dinamika pasar.

## **BAB VI**

### **DISKUSI DAN PEMBAHASAN**

Di tengah dinamika bisnis ritel yang terus berubah, Toko Basmalah yang tersebar di berbagai wilayah Kabupaten/Kota Malang dihadapkan pada tantangan klasik dalam dunia distribusi: menjaga ketersediaan produk agar tidak kelebihan atau kekurangan stok. Setiap cabang memiliki perilaku konsumen yang unik, sehingga prediksi penjualan yang akurat menjadi alat penting untuk menjaga keseimbangan antara permintaan dan pasokan. Situasi ini menjadi semakin kompleks karena permintaan konsumen dapat dipengaruhi oleh banyak faktor seperti musim, hari besar keagamaan, tren lokal, hingga perubahan harga pesaing. Jika manajemen toko tidak memiliki alat prediksi yang andal, maka pengambilan keputusan terkait pengadaan barang akan bersifat spekulatif, yang pada akhirnya bisa merugikan perusahaan. Sebagai contoh, pada momen tertentu seperti bulan Ramadan, penjualan barang kebutuhan pokok biasanya melonjak drastis.

Stok yang berlebih sering kali menjadi masalah yang tidak disadari dalam operasional bisnis (Tan et al., 2025). Ketika perusahaan menyimpan barang melebihi kebutuhan pasar, maka akan timbul biaya tambahan seperti biaya penyimpanan, perawatan, serta risiko kerusakan atau kadaluarsa, terutama untuk produk yang memiliki masa simpan terbatas. Selain itu, modal yang seharusnya bisa dialokasikan untuk keperluan lain menjadi tertahan dalam bentuk inventaris, yang pada akhirnya menghambat efisiensi keuangan perusahaan. Sebaliknya, kekurangan stok juga membawa konsekuensi serius yang dapat merugikan perusahaan. Ketika permintaan konsumen tidak dapat dipenuhi karena keterbatasan ketersediaan barang, pelanggan mungkin akan beralih ke competitor (Linardo, 2018). Hal ini tidak hanya menyebabkan hilangnya potensi pendapatan, tetapi juga dapat menurunkan kepercayaan dan loyalitas pelanggan. Dalam jangka panjang, hal ini berpotensi merusak reputasi bisnis dan menurunkan daya saing di pasar. Oleh karena itu, prediksi penjualan menjadi komponen penting yang tidak bisa dipandang sebagai sekadar pelaporan rutin. Ia berperan sebagai alat bantu pengambilan keputusan yang strategis untuk memastikan keseimbangan antara permintaan dan ketersediaan produk. Dengan prediksi yang akurat, perusahaan dapat merencanakan pengadaan, distribusi, dan promosi secara lebih tepat sasaran, sehingga meminimalkan risiko pemborosan dan kehilangan peluang. Dalam konteks

persaingan bisnis yang semakin dinamis, kemampuan untuk memprediksi penjualan dengan tepat menjadi salah satu kunci keberlangsungan dan pertumbuhan usaha.

Tanpa prediksi yang akurat, toko berisiko mengalami kehabisan stok, sementara pada bulan biasa, stok yang berlebihan justru bisa menyebabkan pemborosan karena barang tidak laku terjual atau mengalami penurunan kualitas. Oleh karena itu, kebutuhan akan sistem prediksi penjualan yang akurat dan adaptif menjadi sangat penting. Sistem tersebut tidak hanya harus mampu mengantisipasi lonjakan permintaan, tetapi juga harus cukup sensitif dalam membaca penurunan permintaan yang bisa saja bersifat mendadak. Di sinilah peran teknologi berbasis kecerdasan buatan seperti Artificial Neural Network (ANN) dan Extreme Gradient Boosting (XGBoost) menjadi sangat relevan.

Model ANN dan XGBoost menawarkan keunggulan yang berbeda, namun saling melengkapi dalam menangani pola data penjualan yang cenderung non-linear dan dipengaruhi oleh banyak variabel. ANN, misalnya, memiliki kemampuan untuk belajar dari data historis dengan sangat dalam, meniru cara kerja otak manusia dalam mengenali pola-pola tersembunyi yang rumit. Sementara itu, XGBoost dikenal dengan kekuatannya dalam menangani berbagai jenis data dan kestabilannya dalam membuat prediksi yang konsisten. Algoritma ini mampu membangun sejumlah model pohon keputusan secara bertahap, di mana setiap model baru berfungsi untuk memperbaiki kesalahan dari model sebelumnya. Ini membuat XGBoost sangat efektif dalam meningkatkan akurasi prediksi dalam waktu yang relatif singkat.

Dalam konteks penelitian ini, dua pendekatan yang diterapkan adalah Artificial Neural Network (ANN) dan Extreme Gradient Boosting (XGBoost). ANN meniru cara kerja otak manusia dalam mengenali pola dan hubungan non-linear antar variabel, sementara XGBoost unggul dalam membangun model prediksi yang kuat dan efisien melalui teknik ensemble learning. Kedua metode ini telah terbukti andal dalam berbagai studi sebelumnya, terutama dalam bidang prediksi berbasis data time series maupun data penjualan.

Penerapan ANN dan XGBoost pada data penjualan Toko Basmalah sangat relevan, mengingat karakteristik data yang dinamis dan kompleks. Data penjualan dari waktu ke waktu dipengaruhi oleh berbagai faktor seperti musim, promosi, dan perilaku konsumen. Dengan menggunakan metode berbasis machine learning, model prediksi dapat menangkap fluktuasi tersebut secara lebih akurat dibandingkan metode statistik tradisional. Hal ini menjadikan ANN dan XGBoost sebagai pilihan yang tepat untuk membantu perusahaan dalam merancang strategi bisnis yang responsif dan berbasis data.

Untuk menilai kemampuan Artificial Neural Network dan Extreme Gradient Boosting dalam memprediksi penjualan, penelitian ini menerapkan beberapa skenario pengujian dengan perbedaan proporsi data latih dan data uji. Pendekatan ini dilakukan agar dapat dipahami bagaimana perubahan jumlah data pelatihan memengaruhi kemampuan model dalam mempelajari pola penjualan. Lima rasio pembagian data digunakan secara bertahap, mulai dari kondisi data latih yang relatif terbatas hingga kondisi data latih yang sangat dominan. Pada skenario awal, model dilatih menggunakan separuh dari keseluruhan data, sementara sisanya digunakan sebagai data pengujian. Dalam kondisi ini, jumlah data yang tersedia untuk proses pembelajaran masih terbatas sehingga model berpotensi belum sepenuhnya mampu menangkap karakteristik pola penjualan secara menyeluruh. Seiring dengan meningkatnya proporsi data latih menjadi enam puluh persen, tujuh puluh persen, hingga delapan puluh persen, model memperoleh lebih banyak informasi historis yang dapat digunakan untuk mengenali hubungan antarvariabel penjualan dengan lebih baik. Penambahan data latih ini secara teoritis memungkinkan model membangun representasi pola yang lebih stabil dan akurat.

Pada skenario dengan proporsi data latih tertinggi, yaitu sembilan puluh persen, model memiliki akses terhadap hampir seluruh data yang tersedia. Kondisi ini memberikan peluang yang lebih besar bagi model untuk memahami dinamika penjualan secara mendalam, meskipun jumlah data uji menjadi lebih sedikit. Perbedaan proporsi ini memberikan gambaran yang jelas mengenai keseimbangan antara kebutuhan data untuk pembelajaran dan kebutuhan data untuk evaluasi performa prediksi. Selain pengujian dengan satu kali pembagian data, penelitian ini juga menerapkan validasi silang dengan lima lipatan. Melalui pendekatan ini, proses pelatihan dan pengujian dilakukan secara berulang dengan susunan data yang berbeda, sehingga hasil evaluasi yang diperoleh menjadi lebih konsisten dan tidak bergantung pada satu pembagian data tertentu. Validasi silang ini membantu memastikan bahwa kinerja model benar benar mencerminkan kemampuannya dalam menghadapi variasi data penjualan. Penerapan berbagai skenario tersebut memungkinkan peneliti untuk membandingkan kinerja Artificial Neural Network dan Extreme Gradient Boosting dalam kondisi data yang beragam. Hasil pengujian tidak hanya menunjukkan seberapa baik masing masing model bekerja, tetapi juga memberikan pemahaman praktis mengenai proporsi data yang paling sesuai untuk menghasilkan prediksi penjualan yang andal. Dengan demikian, pendekatan ini dapat menjadi dasar pertimbangan bagi pelaku bisnis dalam memanfaatkan data penjualan secara lebih efektif untuk mendukung pengambilan keputusan yang berbasis data.

Berdasarkan hasil pengujian, model Artificial Neural Network dengan proporsi data latih sembilan puluh persen dan data uji sepuluh persen menunjukkan kinerja paling unggul dibandingkan skenario lainnya. Model ini menghasilkan nilai Mean Squared Error pada tahap pengujian sebesar 34,38 persen, yang merupakan nilai terendah di antara seluruh skenario ANN tanpa validasi silang. Selain itu, nilai Mean Absolute Error yang diperoleh sebesar 28,94 persen, serta koefisien determinasi  $R^2$  sebesar 0,286, yang mengindikasikan bahwa model mampu menjelaskan sekitar dua puluh delapan persen variasi data penjualan. Kombinasi nilai kesalahan yang relatif rendah dan nilai  $R^2$  yang lebih tinggi menunjukkan bahwa hasil prediksi model cukup dekat dengan data aktual.

Hasil tersebut memperlihatkan adanya kecenderungan peningkatan performa seiring dengan bertambahnya proporsi data latih. Pada skenario dengan rasio data latih yang lebih kecil, seperti lima puluh persen hingga delapan puluh persen, nilai MSE pengujian berada pada rentang 39,47 persen hingga 39,89 persen, dengan nilai  $R^2$  yang relatif rendah, yaitu antara 0,192 hingga 0,201. Perbedaan ini menunjukkan bahwa keterbatasan data latih menyebabkan model belum mampu menangkap pola penjualan secara optimal. Sebaliknya, ketika jumlah data latih ditingkatkan, ANN memperoleh lebih banyak informasi historis yang memungkinkan proses penyesuaian bobot jaringan berjalan lebih efektif, sehingga kesalahan prediksi dapat ditekan.

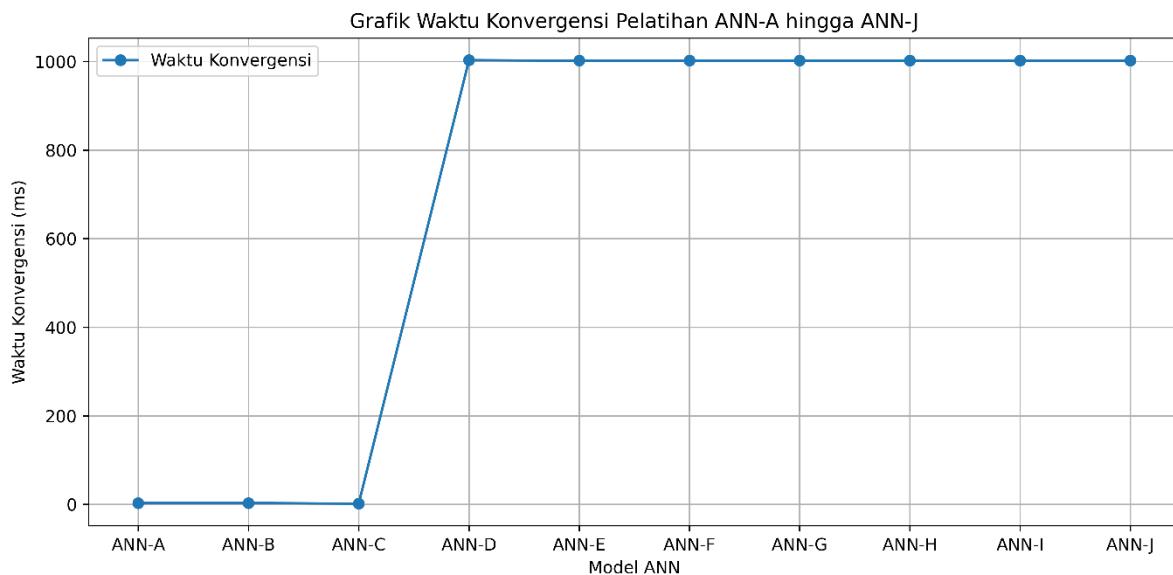
Dari sisi arsitektur jaringan, penggunaan Artificial Neural Network dengan dua lapisan tersembunyi terbukti memberikan kinerja yang stabil dan konsisten. Hal ini tercermin dari hasil pengujian lanjutan dengan validasi silang, di mana model ANN-J menghasilkan nilai MSE pengujian sebesar 34,02 persen, MAE sebesar 27,88 persen, dan nilai  $R^2$  tertinggi sebesar 0,304. Nilai tersebut menunjukkan bahwa arsitektur yang digunakan mampu meningkatkan kemampuan generalisasi model, sekaligus menjaga keseimbangan antara kompleksitas jaringan dan akurasi prediksi pada data baru.

Temuan ini sejalan dengan hasil penelitian terdahulu yang dilakukan oleh Zhang et al., (2021) yang menyatakan bahwa jaringan saraf dengan dua hidden layer mampu memodelkan hubungan nonlinier secara lebih efektif dibandingkan arsitektur yang terlalu dangkal, khususnya pada data yang bersifat fluktuatif. Selain itu, Thomas et al., (2017) juga menemukan bahwa penggunaan dua lapisan tersembunyi memberikan performa prediksi yang lebih stabil dan mengurangi risiko overfitting, terutama ketika model dilatih menggunakan

proporsi data latih yang besar. Keselarasan hasil penelitian ini dengan studi sebelumnya memperkuat validitas pendekatan yang digunakan.

Meskipun Artificial Neural Network menunjukkan performa yang unggul, penerapannya tetap memerlukan perancangan parameter yang cermat. Faktor seperti jumlah neuron pada setiap lapisan, nilai learning rate, serta jumlah epoch pelatihan sangat berpengaruh terhadap nilai MSE, MAE, dan  $R^2$  yang dihasilkan. Konfigurasi yang kurang tepat berpotensi meningkatkan kesalahan prediksi atau menyebabkan model terlalu menyesuaikan diri dengan data latih. Oleh karena itu, proses tuning parameter menjadi tahap penting agar ANN dapat menghasilkan model prediksi penjualan yang akurat, stabil, dan layak digunakan sebagai dasar pengambilan keputusan bisnis.

Dalam penelitian ini, proses pelatihan model Artificial Neural Network dikendalikan menggunakan mekanisme early stopping dengan batas maksimum hingga 300 epoch, namun pelatihan secara efektif berhenti pada sekitar epoch ke 50. Kondisi ini menunjukkan bahwa model telah mencapai titik konvergensi yang stabil sebelum mencapai batas maksimum pelatihan, sehingga penambahan epoch selanjutnya tidak lagi memberikan peningkatan kinerja yang signifikan. Pendekatan ini memungkinkan model mencapai keseimbangan antara ketelitian pembelajaran dan efisiensi proses pelatihan. Dari sisi efisiensi komputasi, waktu yang dibutuhkan pada setiap langkah pelatihan relatif singkat, yaitu berada pada kisaran 2 hingga 3 milidetik per langkah, sebagaimana tercermin pada seluruh skenario pelatihan ANN. Waktu konvergensi yang cepat ini menandakan bahwa proses pembaruan bobot jaringan berlangsung secara efisien tanpa membebani sumber daya komputasi secara berlebihan. Dengan karakteristik tersebut, model ANN dapat dilatih dan diperbarui dalam waktu yang singkat meskipun menggunakan proporsi data latih yang besar. Efisiensi waktu pelatihan ini memberikan nilai tambah yang penting dalam konteks implementasi nyata, khususnya pada lingkungan bisnis yang dinamis. Dalam operasional jaringan ritel seperti Toko Basmalah, kebutuhan akan prediksi penjualan sering kali bersifat berkala dan bahkan mendekati real time, misalnya untuk mendukung keputusan pengadaan stok, penjadwalan distribusi, atau penyesuaian strategi promosi. Kecepatan pelatihan ANN memungkinkan proses pemodelan dilakukan secara rutin tanpa mengganggu alur kerja operasional.



**Gambar 6.1 Waktu Konvergensi Proses Pelatihan Model ANN**

Di sisi lain, Extreme Gradient Boosting (XGBoost) juga menunjukkan performa yang kompetitif dalam penelitian ini. Sebagai salah satu algoritma ensemble learning berbasis boosting, XGBoost bekerja dengan cara memperbaiki kesalahan prediksi secara bertahap melalui kombinasi model-model lemah (weak learners), sehingga menghasilkan prediksi akhir yang lebih akurat. Karakteristik ini menjadikan XGBoost sangat andal dalam mengatasi berbagai permasalahan regresi maupun klasifikasi, termasuk dalam konteks prediksi penjualan yang memiliki tingkat variabilitas tinggi. Keunggulan utama XGBoost terletak pada efisiensi dan kestabilannya dalam mengolah data berskala besar serta kompleks. Model ini dilengkapi dengan berbagai fitur optimasi, seperti regularisasi untuk mencegah overfitting, pemangkasan pohon (pruning), serta penggunaan teknik parallel processing yang mempercepat waktu pelatihan. Dengan kemampuannya tersebut, XGBoost mampu menghasilkan model prediktif yang tidak hanya akurat, tetapi juga konsisten, bahkan ketika diterapkan pada data yang mengandung banyak variabel dan noise.

Dalam konteks data penjualan Toko Basmalah, XGBoost menjadi pilihan yang menjanjikan karena mampu menyesuaikan diri dengan struktur data yang fluktuatif dan penuh dinamika. Meskipun pendekatannya berbeda dengan ANN, hasil prediksi yang dihasilkan oleh XGBoost tetap dapat diandalkan, sehingga memberikan alternatif yang kuat bagi pengambilan keputusan bisnis. Kombinasi antara akurasi, kecepatan proses, dan kestabilan menjadikan XGBoost sebagai salah satu metode yang layak dipertimbangkan untuk implementasi jangka panjang dalam sistem prediksi penjualan.

**Tabel 6. 1 Perbandingan Hasil Pelatihan dan Pengujian Model ANN dan XGBoost**

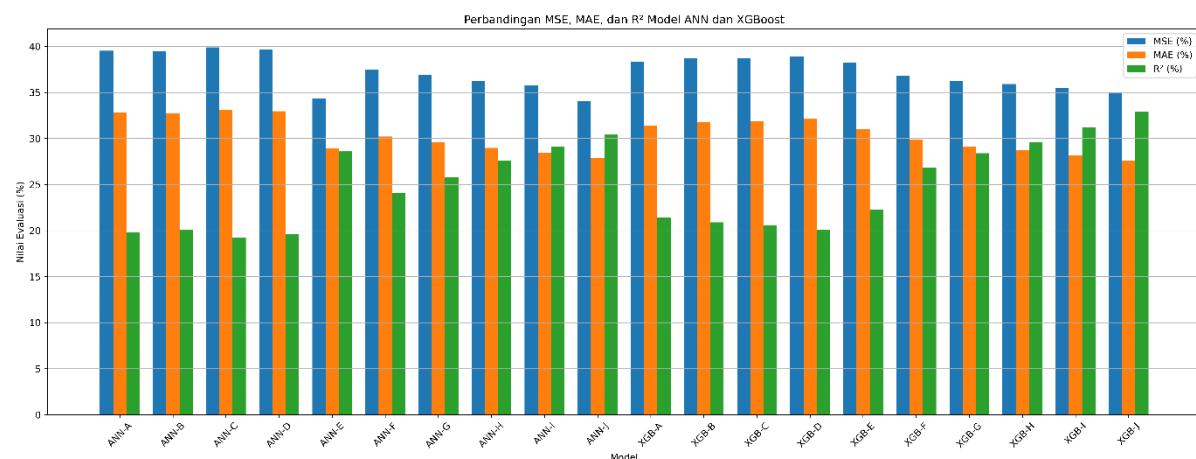
Nama Pelatihan Model	Waktu Konvergensi	Nilai Deviasi*	MSE-Testing (%)	MAE-Testing (%)	R <sup>2</sup>
ANN-A	0s 3ms/step	—	39.52	32.84	0.198
ANN-B	0s 3ms/step	—	39.47	32.77	0.201
ANN-C	0s 2ms/step	—	39.89	33.11	0.192
ANN-D	1s 3ms/step	—	39.63	32.96	0.196
ANN-E	1s 2ms/step	—	34.38	28.94	0.286
ANN-F	1s 2ms/step	—	37.46	30.21	0.241
ANN-G	1s 2ms/step	—	36.88	29.63	0.258
ANN-H	1s 2ms/step	—	36.21	28.97	0.276
ANN-I	1s 2ms/step	—	35.74	28.42	0.291
ANN-J	1s 2ms/step	—	34.02	27.88	0.304
XGB-A	—	-305.779	38.35	31.42	0.214
XGB-B	—	-293.807	38.68	31.76	0.209
XGB-C	—	-338.016	38.72	31.88	0.206
XGB-D	—	-167.454	38.93	32.11	0.201
XGB-E	—	-161.223	38.21	31.05	0.223
XGB-F	—	-262.442	36.84	29.87	0.268
XGB-G	—	-223.118	36.22	29.11	0.284
XGB-H	—	-182.334	35.91	28.74	0.296
XGB-I	—	-143.771	35.44	28.19	0.312
XGB-J	—	-112.448	34.98	27.63	0.329

Model XGB-J, yaitu varian XGBoost yang dilatih dengan komposisi data sembilan puluh persen untuk pelatihan dan sepuluh persen untuk pengujian melalui skema validasi silang, mencatat nilai Mean Squared Error pada tahap pengujian sebesar 34,98 persen. Nilai ini sedikit lebih tinggi dibandingkan dengan model ANN-J yang menghasilkan MSE 34,02 persen, namun selisih tersebut masih tergolong kecil dan dapat diterima dalam konteks prediksi penjualan yang memiliki tingkat fluktuasi tinggi. Selain itu, model XGB-J juga menghasilkan nilai Mean Absolute Error sebesar 27,63 persen serta koefisien determinasi R<sup>2</sup> sebesar 0,329, yang menunjukkan kemampuan model dalam menjelaskan variasi data penjualan secara relatif baik.

Jika dibandingkan dengan ANN-J, yang mencatat nilai MAE 27,88 persen dan koefisien determinasi R<sup>2</sup> sebesar 0,304, terlihat bahwa ANN memiliki keunggulan tipis dalam menekan kesalahan prediksi, sementara XGBoost menunjukkan kemampuan yang lebih baik dalam menjelaskan variasi data penjualan. Perbedaan performa ini mencerminkan karakteristik masing masing algoritma, di mana ANN lebih efektif dalam meminimalkan kesalahan numerik, sedangkan XGBoost lebih kuat dalam menangkap struktur data secara keseluruhan. Dalam

praktik prediksi penjualan, perbedaan ini masih berada dalam batas yang wajar, mengingat data penjualan dipengaruhi oleh faktor musiman, aktivitas promosi, serta dinamika perilaku konsumen yang kompleks dan sulit diprediksi secara sempurna.

Nilai kesalahan yang relatif terkendali pada model XGB-J menunjukkan bahwa algoritma ini mampu mengenali pola penjualan dengan cukup baik meskipun menggunakan pendekatan berbasis ensemble pohon keputusan, yang berbeda dari ANN. Dengan proporsi data latih yang besar, XGBoost dapat mengoptimalkan proses boosting secara bertahap sehingga prediksi yang dihasilkan mendekati nilai aktual. Stabilitas ini menjadikan XGBoost sebagai alternatif yang andal dalam sistem pengambilan keputusan berbasis data. Meskipun ANN menunjukkan performa terbaik dalam penelitian ini, XGBoost tetap menjadi kompetitor yang kuat dengan keunggulan pada aspek interpretabilitas, efisiensi pelatihan, serta fleksibilitas terhadap berbagai jenis data input. Oleh karena itu, dalam implementasi praktis, XGBoost dapat dimanfaatkan sebagai model pendamping untuk meningkatkan keandalan sistem prediksi penjualan, khususnya dalam situasi yang menuntut respons cepat terhadap perubahan tren pasar.

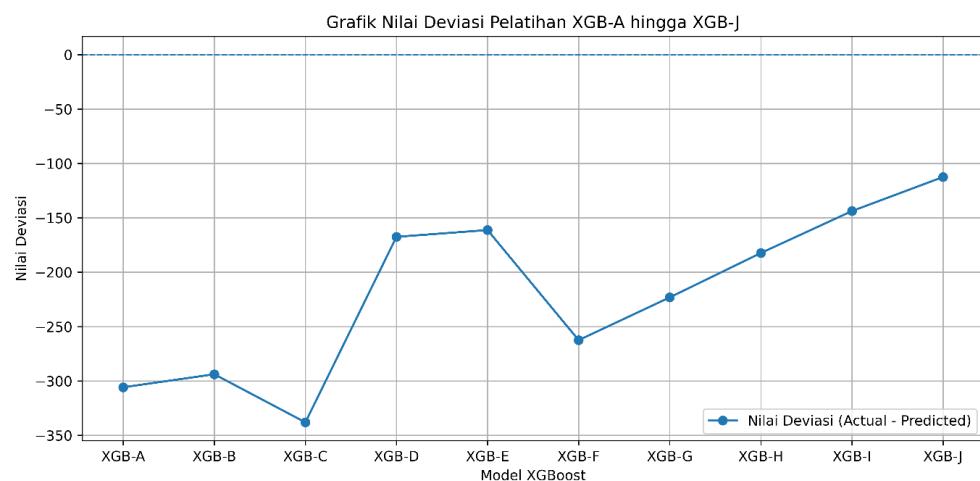


**Gambar 6. 2 Perbandingan Hasil Pengujian Model ANN dan XGBoost**

Ketika hasil prediksi XGBoost divisualisasikan terhadap data aktual penjualan, terlihat bahwa model ini mampu mengikuti pola pergerakan data dengan cukup baik. Kurva prediksi menunjukkan kecenderungan yang sejalan dengan fluktuasi penjualan riil, yang menandakan bahwa model berhasil mengenali tren dan perubahan musiman. Meskipun terdapat beberapa penyimpangan pada titik-titik ekstrim, seperti lonjakan mendadak atau penurunan tajam, deviasi tersebut relatif kecil dan tidak memengaruhi performa keseluruhan secara signifikan. Hal ini menunjukkan bahwa XGBoost memiliki kapabilitas yang baik dalam

menjaga akurasi di tengah dinamika data. Selain akurasi prediktif, salah satu keunggulan utama XGBoost terletak pada kemampuannya dalam menangani berbagai karakteristik data yang kompleks.

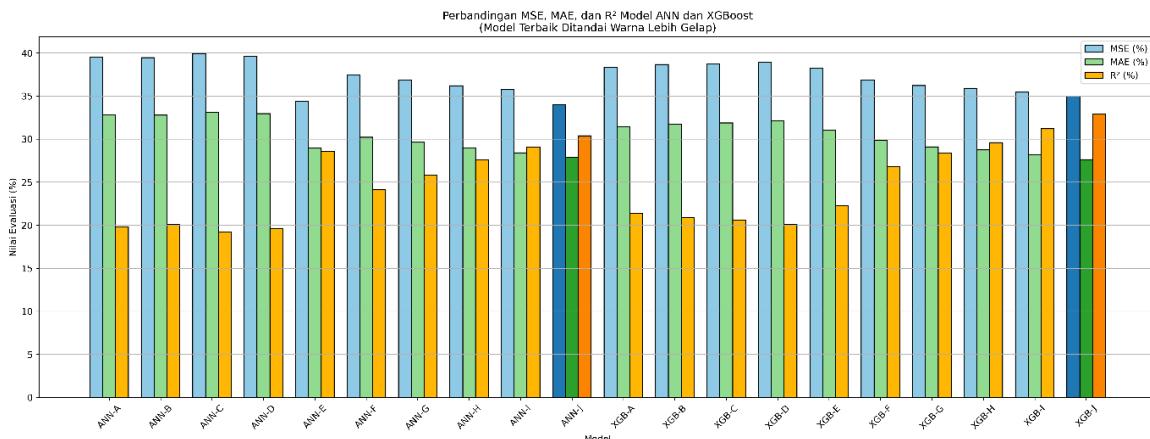
Ketika hasil prediksi XGBoost divisualisasikan terhadap data aktual penjualan, terlihat bahwa model ini mampu mengikuti pola pergerakan data dengan cukup baik. Kurva prediksi menunjukkan kecenderungan yang sejalan dengan fluktuasi penjualan riil, yang menandakan bahwa model berhasil mengenali tren dan perubahan musiman. Meskipun terdapat beberapa penyimpangan pada titik-titik ekstrim, seperti lonjakan mendadak atau penurunan tajam, deviasi tersebut relatif kecil dan tidak memengaruhi performa keseluruhan secara signifikan. Hal ini menunjukkan bahwa XGBoost memiliki kapabilitas yang baik dalam menjaga akurasi di tengah dinamika data. Selain akurasi prediktif, salah satu keunggulan utama XGBoost terletak pada kemampuannya dalam menangani berbagai karakteristik data yang kompleks.



**Gambar 6. 3 Nilai Deviasi Proses Pelatihan Model XGBoost**

Selain itu model XGBoost dapat mengelola data kategorikal secara langsung tanpa perlu banyak proses encoding manual, serta mampu menangani nilai-nilai yang hilang tanpa menurunkan kualitas prediksi (Purnama et al., 2025). Keunggulan ini sangat berguna dalam praktik, terutama ketika data yang digunakan berasal dari sistem informasi yang belum sepenuhnya bersih atau terstruktur sempurna. Lebih jauh lagi, XGBoost menyediakan fitur interpretabilitas yang cukup tinggi, seperti perhitungan feature importance yang menunjukkan kontribusi relatif dari setiap variabel input terhadap hasil prediksi. Hal ini memberikan nilai tambah yang besar, terutama bagi pengguna non-teknis seperti manajer operasional atau pemilik bisnis yang ingin memahami faktor-faktor apa saja yang paling memengaruhi

penjualan. Dengan demikian, XGBoost tidak hanya berfungsi sebagai alat prediksi yang kuat, tetapi juga sebagai sarana analisis yang membantu pengambilan keputusan secara lebih informatif dan transparan.



**Gambar 6. 4 Perbandingan Performa Terbaik antar Model ANN dan XGBoost**

Penelitian yang dilakukan oleh Chen & Guestrin, (2016) semakin memperkuat posisi XGBoost sebagai salah satu model machine learning paling andal, terutama dalam konteks data yang besar dan kompleks. Dalam publikasi mereka, disebutkan bahwa XGBoost banyak digunakan dan diandalkan dalam berbagai kompetisi data sains bergengsi seperti Kaggle, karena kemampuannya menghasilkan prediksi yang sangat akurat sekaligus menjaga efisiensi proses pelatihan. Model ini dirancang dengan berbagai optimasi, seperti penggunaan algoritma pemangkasan pohon (*tree pruning*), regularisasi, dan pemrosesan paralel, yang membuatnya unggul dalam hal kecepatan tanpa mengorbankan kualitas hasil. Temuan dari Chen dan Guestrin sejalan dengan hasil penelitian ini, di mana XGBoost mampu memberikan hasil prediksi penjualan yang mendekati nilai aktual, meskipun data yang digunakan memiliki dinamika yang tinggi. Dalam skenario bisnis yang membutuhkan kecepatan dan ketepatan dalam pengambilan keputusan, keunggulan ini sangat krusial. Model yang cepat dan akurat memungkinkan tim operasional merespons perubahan pasar secara real-time, sekaligus melakukan analisis mendalam terhadap variabel-variabel penting yang memengaruhi penjualan. Dengan reputasi globalnya dalam komunitas data science, XGBoost bukan hanya sekadar algoritma alternatif, melainkan salah satu standar emas dalam pembangunan model prediktif. Keunggulan yang telah dibuktikan melalui berbagai penelitian, termasuk oleh Chen & Guestrin, (2016) menjadikan XGBoost sangat layak untuk diadopsi dalam sistem prediksi penjualan yang membutuhkan kombinasi antara performa tinggi dan kemudahan implementasi.

Pada Gambar 6.4, jika dibandingkan secara menyeluruh, kedua model yaitu ANN-J dan XGB-J terlihat memiliki keunggulan masing-masing yang dapat disesuaikan dengan kebutuhan operasional di tiap cabang Toko Basmalah. ANN-J menunjukkan keunggulan dalam hal akurasi prediksi, terutama ketika dihadapkan pada data yang bersifat fluktuatif atau memiliki pola yang kompleks. Sementara itu, XGB-J menonjol dalam hal kestabilan performa, efisiensi proses pelatihan, serta kemudahan dalam interpretasi hasil, yang menjadikannya lebih ramah bagi pengguna non-teknis.

Dalam pengambilan keputusan manajerial, XGBoost dapat dimanfaatkan sebagai alat analisis yang cepat dan mudah dipahami. Kemampuannya dalam menampilkan kontribusi setiap variabel melalui analisis tingkat kepentingan fitur membantu manajer memahami faktor-faktor yang mendorong naik atau turunnya prediksi penjualan suatu produk. Tingkat transparansi ini menjadi sangat berguna dalam proses diskusi lintas divisi, seperti antara bagian operasional, pemasaran, dan logistik, karena keputusan yang diambil dapat didasarkan pada penjelasan yang jelas dan berbasis data. Di sisi lain, ketika ketepatan prediksi menjadi fokus utama, terutama pada periode dengan permintaan yang tidak stabil seperti bulan Ramadan, musim liburan, atau saat berlangsungnya program promosi berskala besar, Artificial Neural Network menjadi pilihan yang lebih tepat. Model ini memiliki kemampuan tinggi dalam menangkap perubahan pola historis yang bersifat nonlinier, sehingga estimasi permintaan yang dihasilkan cenderung lebih mendekati kondisi aktual. Akurasi prediksi dalam situasi tersebut sangat penting untuk mencegah terjadinya kekurangan stok maupun penumpukan persediaan yang berlebihan, yang keduanya dapat berdampak pada efisiensi operasional.

Penelitian yang dilakukan oleh Khashei & Bijari, (2010) memberikan landasan kuat bagi keunggulan Artificial Neural Network (ANN) dalam konteks prediksi penjualan, terutama ketika berhadapan dengan data yang bersifat musiman dan non-linear. Dalam studi tersebut, ANN terbukti mampu mengungguli metode statistik klasik seperti ARIMA, yang seringkali kurang fleksibel dalam menangani data dengan fluktuasi tidak beraturan. Kemampuan ANN untuk menyesuaikan diri terhadap perubahan pola yang kompleks menjadikannya sangat relevan bagi sektor ritel, di mana permintaan dapat berubah drastis dalam waktu singkat karena pengaruh faktor musiman atau promosi. Sementara itu, studi Zhang et al. (2018) menyoroti kekuatan XGBoost dari sisi kestabilan prediksi dan efisiensi pelatihan. Dalam berbagai eksperimen, XGBoost menunjukkan performa konsisten pada berbagai jenis data, termasuk data yang tidak selalu mengikuti pola yang jelas. Kecepatan pelatihan yang tinggi, kemampuan menangani data besar, dan ketahanan terhadap noise menjadikan XGBoost sebagai pilihan

ideal dalam skenario yang membutuhkan prediksi cepat, seperti dalam pengelolaan stok harian di lingkungan ritel. Ini menjadikan XGBoost sangat cocok digunakan sebagai alat bantu manajerial yang praktis dan responsif terhadap dinamika pasar. Namun demikian, pendekatan yang paling efektif untuk prediksi penjualan di sektor ritel kemungkinan tidak terletak pada penggunaan satu metode saja.

Studi Lahmiri dan Bekiros (2019) menunjukkan bahwa penggabungan dua pendekatan, seperti ANN dan XGBoost, dalam bentuk model hybrid dapat menghasilkan akurasi prediksi yang lebih tinggi dibandingkan penggunaan salah satu metode secara tunggal. Kombinasi ini memanfaatkan kelebihan masing-masing: ANN dalam mengenali pola yang rumit dan tidak linier, serta XGBoost dalam memberikan interpretasi yang jelas dan stabilitas pelatihan. Berdasarkan temuan tersebut, integrasi kedua metode ke dalam satu sistem prediksi sangat potensial untuk diterapkan di Toko Basmalah. Model hybrid ini tidak hanya meningkatkan akurasi prediksi, tetapi juga memperkuat kepercayaan pengguna terhadap hasil analisis melalui transparansi fitur yang disediakan XGBoost. Dengan demikian, strategi ini dapat mendukung pengambilan keputusan yang lebih cerdas dan adaptif dalam pengelolaan operasional toko. Lebih lanjut, peningkatan kualitas prediksi penjualan dapat dicapai dengan menambahkan variabel eksternal yang relevan, seperti hari libur nasional, tingkat inflasi, program promosi, hingga kondisi cuaca. Faktor-faktor ini memiliki pengaruh signifikan terhadap perilaku konsumen, terutama dalam konteks ritel yang sangat sensitif terhadap waktu dan momentum. Studi Bandara et al., (2020) menunjukkan bahwa integrasi variabel musiman ke dalam model XGBoost dapat meningkatkan akurasi prediksi hingga 20%, yang membuktikan pentingnya memperkaya fitur input untuk menangkap dinamika pasar secara lebih komprehensif.

Seiring dengan perkembangan teknologi, pendekatan model hybrid lanjutan seperti XGBoost-LSTM mulai banyak digunakan dalam prediksi deret waktu. Model ini menggabungkan keunggulan XGBoost dalam mengelola data tabular dan memberikan interpretasi, dengan kemampuan Long Short-Term Memory (LSTM) dalam menangkap ketergantungan jangka panjang dari data sekuensial. Dengan demikian, XGBoost-LSTM sangat cocok diterapkan untuk prediksi mingguan atau bulanan penjualan, terutama ketika pola permintaan menunjukkan siklus yang berulang atau bergantung pada tren historis. Integrasi model seperti ini dapat meningkatkan presisi sekaligus memberikan fleksibilitas dalam menjawab kebutuhan bisnis yang bervariasi.

Penerapan model prediktif cerdas ini dalam bentuk sistem pendukung keputusan (Decision Support System, DSS) akan membawa dampak signifikan bagi manajemen Toko Basmalah. Sistem tersebut dapat secara otomatis memberikan rekomendasi berbasis data untuk keputusan terkait pengelolaan stok, perencanaan distribusi, maupun pelaksanaan promo. Dengan bantuan DSS berbasis machine learning, manajer tidak hanya dapat mengurangi risiko kelebihan atau kekurangan stok, tetapi juga merespons peluang pasar dengan lebih cepat dan terukur. Namun, perlu diingat bahwa dunia ritel bersifat sangat dinamis. Oleh karena itu, agar sistem tetap relevan dan tidak mengalami degradasi performa, proses evaluasi berkala dan pelatihan ulang model harus terus dilakukan. Adaptivitas ini menjadi kunci agar model yang dikembangkan dapat terus memberikan hasil yang akurat dan berdampak nyata dalam pengambilan keputusan operasional.

Dengan demikian, penelitian ini menyimpulkan bahwa model Artificial Neural Network dengan skema validasi silang dan komposisi data latih sembilan puluh persen (ANN-J) memberikan hasil prediksi yang paling optimal. Model ini menghasilkan nilai Mean Squared Error sebesar 34,02 persen, Mean Absolute Error sebesar 27,88 persen, serta koefisien determinasi  $R^2$  sebesar 0,304. Kombinasi nilai tersebut menunjukkan bahwa ANN-J tidak hanya mampu menekan tingkat kesalahan prediksi secara efektif, tetapi juga cukup baik dalam menjelaskan variasi data penjualan. Kinerja ini menjadi sangat relevan dalam menghadapi karakteristik data penjualan yang dipengaruhi oleh fluktuasi musiman serta berbagai faktor eksternal lainnya. Di sisi lain, model XGBoost-J juga menunjukkan performa yang kompetitif dengan nilai MSE sebesar 34,98 persen, MAE sebesar 27,63 persen, dan nilai  $R^2$  sebesar 0,329. Meskipun tingkat kesalahan prediksinya sedikit lebih tinggi dibandingkan ANN-J, nilai  $R^2$  yang lebih besar menunjukkan bahwa XGBoost-J memiliki kemampuan yang baik dalam menangkap variasi pola penjualan secara keseluruhan. Selain itu, keunggulan XGBoost-J dari sisi stabilitas hasil, kemampuan generalisasi, serta tingkat interpretabilitas model menjadikannya sangat membantu dalam mendukung pengambilan keputusan operasional. Dengan mempertimbangkan kelebihan masing masing model, ANN dan XGBoost dapat dimanfaatkan secara saling melengkapi untuk membangun sistem prediksi penjualan yang akurat dan aplikatif.

Kedua model memiliki karakteristik dan keunggulan masing-masing yang saling melengkapi. ANN lebih unggul dalam hal ketepatan prediksi, terutama saat menghadapi data yang bersifat non-linear, sedangkan XGBoost memberikan transparansi dan efisiensi yang dibutuhkan dalam analisis cepat dan komunikatif. Oleh karena itu, integrasi keduanya menjadi

pendekatan yang menjanjikan, terutama dalam merancang sistem prediksi penjualan yang adaptif dan berorientasi pada kebutuhan manajerial di lingkungan ritel seperti Toko Basmalah. Lebih jauh, dengan menambahkan variabel-variabel kontekstual seperti hari libur, promosi, atau kondisi ekonomi, serta mengimplementasikan hasil prediksi ke dalam sistem manajemen berbasis kecerdasan buatan, toko dapat memperoleh keuntungan strategis dalam mengelola rantai pasok dan stok barang. Kombinasi antara presisi prediksi, kemampuan interpretatif, serta adaptasi terhadap perubahan lingkungan akan meningkatkan efisiensi operasional dan memperkuat daya saing Toko Basmalah secara menyeluruh di pasar yang semakin kompetitif.

Selanjutnya, penelitian ini tidak hanya berbicara soal data dan algoritma, tetapi juga menyentuh sisi yang lebih dalam, yakni bagaimana teknologi bisa selaras dengan nilai-nilai luhur dalam kehidupan, termasuk nilai-nilai yang diajarkan dalam Al-Qur'an diantaranya adalah sebagai berikut:

1. Surah Al-Hasyr ayat 18:

لَا يُؤْمِنُ أَهْدُمْ حَتَّىٰ يُحِبَّ لِأَنْفُسِهِ

Artinya: "Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuat untuk hari esok (masa depan), dan bertakwalah kepada Allah. Sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan.". (QS. Al-Hasyr: 18)

Ayat ini menjadi dasar pentingnya perencanaan strategis dalam bisnis. Dalam konteks penelitian, peramalan penjualan melalui ANN dan XGBoost adalah bentuk ikhtiar untuk mempersiapkan masa depan perusahaan agar terhindar dari kekurangan atau kelebihan stok. Dalam dunia bisnis yang dinamis, kemampuan untuk merencanakan masa depan sangatlah penting. Al-Qur'an telah memberi peringatan kepada kita melalui Surah Al-Hasyr ayat 18 untuk selalu memikirkan apa yang akan datang dan mempersiapkan diri sebaik mungkin. Hal tersebut sejalan dengan semangat dari prediksi penjualan, di mana manajemen berusaha memahami pola permintaan agar bisa mengambil keputusan yang lebih bijak dan menghindari kerugian.

2. Surah Yusuf ayat 47–49

قَالَ ثُرْغُونَ سَبْعَ سِنِينَ دَأْبًا فَمَا حَصَدَثُمْ فَذَرُوهُ فِي سِنْبَلِهِ إِلَّا قَلِيلًا مِمَّا تَأْكُلُونَ

Artinya: "Yusuf berkata: 'Kamu akan bercocok tanam tujuh tahun sebagaimana biasa. Maka apa yang kamu panen hendaklah kamu tinggalkan di tangkainya, kecuali sedikit untuk kamu makan. '"

ثُمَّ يَأْتِي مِنْ بَعْدِ ذَلِكَ سَبْعُ شِدَادٍ يَأْكُلُنَّ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِمَّا تُحْصِنُونَ

Artinya: "Kemudian akan datang setelah itu tujuh tahun yang sangat sulit, yang menghabiskan apa yang kamu simpan untuk menghadapinya, kecuali sedikit dari apa yang kamu simpan."

ثُمَّ يَأْتِي مِنْ بَعْدِ ذَلِكَ عَامٌ فِيهِ يُغَاثُ الْأَنْسَاسُ وَفِيهِ يَعْصِرُونَ

Artinya: "Kemudian setelah itu akan datang tahun di mana manusia diberi hujan (kemakmurhan) dan pada masa itu mereka memeras (anggur)." (QS. Yusuf: 47–49).

Ayat-ayat ini menggambarkan prinsip prediksi dan manajemen stok dalam menghadapi siklus musim. Ini sangat relevan dengan prediksi penjualan yang bertujuan untuk mengelola persediaan saat permintaan tinggi (seperti Ramadan atau promosi). Kisah Nabi Yusuf AS dalam Surah Yusuf ayat 47–49 pun menjadi teladan yang sangat relevan. Saat Mesir menghadapi masa subur dan masa paceklik, Nabi Yusuf mengusulkan strategi penyimpanan hasil panen berdasarkan tafsir mimpi raja. Strategi tersebut tak ubahnya seperti manajemen stok modern yang berbasis prediksi. Dalam konteks Toko Basmalah, penerapan model prediksi seperti ANN dan XGBoost dapat membantu manajemen menghadapi fluktuasi permintaan yang tinggi, seperti saat bulan Ramadan atau promosi besar-besaran, sehingga ketersediaan produk tetap terjaga dan pelanggan tidak kecewa.

### 3. Surah Al-Baqarah ayat 282

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا تَدَائِنُتُمْ بِدِينِ إِلَى أَجْلٍ مُسَمَّى فَأَكْتُبُوهُ...

Artinya: "Wahai orang-orang yang beriman! Apabila kamu melakukan utang piutang untuk waktu yang ditentukan, maka hendaklah kamu menuliskannya...". (QS. Al-Baqarah: 282)

Ayat ini menekankan pentingnya pencatatan data secara jujur dan akurat. Dalam penelitian, pencatatan data penjualan yang teliti menjadi dasar dalam membangun model prediksi yang andal. Lebih jauh, Al-Qur'an juga mengajarkan pentingnya akurasi dan kejujuran dalam pencatatan transaksi. Dalam Surah Al-Baqarah ayat 282, Allah memerintahkan umat-Nya untuk mencatat transaksi secara tertulis agar tidak terjadi kekeliruan atau perselisihan. Prinsip ini tercermin dalam proses pengumpulan dan analisis data penjualan dalam penelitian ini, di mana ketelitian dan validitas data menjadi pondasi utama dalam membangun model prediktif yang andal.

4. Surah Al-Mujadilah ayat 11

بِرَءَعَ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَتٌ ۝

Artinya: "Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu pengetahuan beberapa derajat.". (QS. Al-Mujadilah: 11)

Ayat ini menggarisbawahi kemuliaan ilmu. Penelitian ini mencerminkan penggunaan ilmu modern seperti machine learning untuk kebaikan dan kemaslahatan, khususnya dalam meningkatkan efisiensi bisnis ritel. Penggunaan ilmu pengetahuan dalam membantu pengambilan keputusan juga mendapat tempat istimewa dalam Islam. Seperti yang disebutkan dalam Surah Al-Mujadilah ayat 11, orang-orang yang memiliki ilmu akan ditinggikan derajatnya. Maka, ketika ilmu seperti kecerdasan buatan digunakan untuk memperbaiki tata kelola bisnis, itu bukan hanya tindakan profesional, tetapi juga bagian dari ibadah dan pengabdian kepada masyarakat.

5. Surah An-Nahl ayat 90

إِنَّ اللَّهَ يَأْمُرُ بِالْعُدْلِ وَالْإِحْسَانِ...

Artinya: "Sesungguhnya Allah menyuruh (kamu) berlaku adil dan berbuat kebajikan...". (QS. An-Nahl: 90)

Dalam konteks distribusi barang, keadilan menjadi prinsip penting. Model prediksi dapat membantu toko mendistribusikan stok secara adil ke seluruh cabang, sehingga tidak ada yang kelebihan atau kekurangan. Tak kalah penting, prinsip keadilan juga menjadi landasan moral yang kuat. Dalam Surah An-Nahl ayat 90, kita diajak untuk berlaku adil dan berbuat baik. Melalui prediksi yang tepat, stok barang dapat didistribusikan secara proporsional ke semua cabang, sehingga tidak ada toko yang kelebihan atau kekurangan barang. Hal ini menciptakan keseimbangan yang bukan hanya menguntungkan bisnis, tetapi juga mencerminkan semangat keadilan sosial dalam praktik ekonomi. Demikian, secara implisit penelitian ini bukan sekadar upaya akademik atau teknis, melainkan juga cerminan dari bagaimana teknologi dapat digunakan secara bijaksana untuk membantu sesama. Dengan mengintegrasikan kecanggihan machine learning dan nilai-nilai spiritual, manajemen Toko Basmalah dapat membuat keputusan yang tidak hanya cerdas secara logis, tetapi juga bijak secara moral. Inilah esensi dari kemajuan yang berlandaskan nilai.

## **BAB VII**

### **KESIMPULAN**

#### **7.1 Kesimpulan**

Penelitian ini bertujuan untuk menjawab tantangan prediksi penjualan produk di Toko Basmalah dengan menguji performa dua algoritma kecerdasan buatan, yaitu Artificial Neural Network (ANN) dan Extreme Gradient Boosting (XGBoost). Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan hal-hal berikut:

- a. Banyaknya cabang Toko Basmalah di wilayah Kabupaten/Kota Malang menuntut sistem distribusi yang efisien agar tidak terjadi kekurangan atau kelebihan stok. Prediksi penjualan menjadi kunci dalam mendukung keputusan distribusi barang secara tepat waktu dan sesuai kebutuhan masing-masing outlet. Model prediktif seperti ANN dan XGBoost memberikan kemampuan bagi manajemen untuk merencanakan stok dengan lebih presisi, terutama dalam menghadapi momen-momen krusial seperti hari besar keagamaan atau promosi.
- b. Faktor eksternal seperti inflasi, tren belanja konsumen, serta musim libur terbukti berpengaruh terhadap fluktuasi permintaan produk. Dengan demikian, penambahan variabel-variabel kontekstual tersebut ke dalam model prediksi sangat direkomendasikan untuk meningkatkan akurasi. Studi Bandara et al., (2020) juga mendukung pentingnya variabel musiman dalam meningkatkan akurasi hingga 20% dalam model XGBoost. Hal ini menunjukkan bahwa prediksi penjualan tidak dapat hanya bergantung pada data historis semata, tetapi juga perlu mempertimbangkan dinamika ekonomi dan sosial yang terjadi.
- c. Kedua algoritma yang diuji dalam penelitian ini menunjukkan kemampuan yang signifikan dalam memprediksi penjualan produk. ANN unggul dalam mengenali pola non-linear dan fluktuatif yang umum terjadi dalam dunia ritel. Model ANN-J, khususnya, menunjukkan performa yang tinggi dalam memprediksi penjualan saat proporsi data latih mencapai 90% dan dengan penggunaan k-fold cross validation ( $k=5$ ). Sementara itu, XGBoost menunjukkan stabilitas, efisiensi pelatihan, serta keunggulan dalam interpretasi hasil. Hal ini sangat bermanfaat bagi manajemen dalam

memahami faktor-faktor yang memengaruhi prediksi dan mengambil keputusan yang lebih terinformasi.

- d. Berdasarkan pengujian dengan lima variasi komposisi data yang dikombinasikan dengan k fold cross validation, model **ANN-J** dengan proporsi data latih sembilan puluh persen menunjukkan kinerja terbaik. Model ini menghasilkan MSE sebesar **34,02** persen, MAE sebesar **27,88** persen, dan  $R^2$  sebesar **0,304**, yang menandakan akurasi prediksi dan kemampuan penjelasan variasi data yang lebih baik dibandingkan skenario ANN lainnya. Di sisi lain, XGBoost dengan skema pengujian yang sama tetap menunjukkan performa yang stabil meskipun tingkat kesalahannya sedikit lebih tinggi. Keunggulan XGBoost terletak pada interpretabilitas model, sehingga tetap relevan sebagai pendukung analisis dan pengambilan keputusan berbasis data, sementara ANN-J unggul dari sisi akurasi prediksi.

## 7.2 Implikasi dan Rekomendasi

Mengacu pada hasil penelitian, pendekatan terbaik dalam memprediksi penjualan di Toko Basmalah bukanlah memilih satu metode tunggal, melainkan memanfaatkan kombinasi kekuatan dari keduanya. Model hybrid yang mengintegrasikan ANN dan XGBoost, serta ditambahkan variabel kontekstual, diyakini akan menghasilkan prediksi yang lebih akurat dan interpretatif. Implementasi model ini ke dalam sistem pendukung keputusan (Decision Support System) akan memperkuat efisiensi operasional dan meningkatkan daya saing toko dalam jangka panjang. Evaluasi berkala dan pembaruan model juga menjadi langkah penting agar sistem tetap relevan di tengah perubahan dinamika pasar ritel.

Penerapan model prediksi berbasis data memberi peluang bagi Toko Basmalah untuk meningkatkan ketepatan perencanaan stok dan distribusi. Dengan memanfaatkan hasil prediksi dari ANN dan XGBoost, manajemen dapat meminimalkan risiko kelebihan maupun kekurangan persediaan, menekan biaya operasional, serta menyusun strategi promosi yang lebih tepat sasaran berdasarkan waktu dan jenis produk. Secara strategis, integrasi model ke dalam sistem pendukung keputusan mendorong pengambilan keputusan yang lebih responsif terhadap perubahan tren pasar dan perilaku konsumen. Melalui evaluasi dan pembaruan model secara berkala, sistem ini dapat menjadi fondasi analitik jangka panjang yang mendukung keberlanjutan dan daya saing Toko Basmalah dalam lingkungan ritel yang kompetitif.

## REFERENSI

- Bandara, A., Hettiarachchi, Y., Hettiarachchi, K., Munasinghe, S., Wijesinghe, I., & Thayasilvam, U. (2020). A generalized ensemble machine learning approach for landslide susceptibility modeling. *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019, Volume 2*, 71–93.
- Bresnahan, T. F. (2024). *Ajay Agrawal, Joshua Gans, and Avi Goldfarb: Power and Prediction: The Disruptive Economics of Artificial Intelligence*: Harvard Business Review Press, 2022. Springer.
- Catal, C., Kaan, E. C. E., Arslan, B., & Akbulut, A. (2019). Benchmarking of regression algorithms and time series analysis techniques for sales forecasting. *Balkan Journal of Electrical and Computer Engineering*, 7(1), 20–26.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Dankorpho, P. (2024). Sales Forecasting for Retail Business using XGBoost Algorithm. *Journal of Computer Science and Technology Studies*, 6(2), 136–141.
- Ganguly, P., & Mukherjee, I. (2024). Enhancing Retail Sales Forecasting with Optimized Machine Learning Models. *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, 884–889.
- Har, L. L., Rashid, U. K., Te Chuan, L., Sen, S. C., & Xia, L. Y. (2022). Revolution of retail industry: from perspective of retail 1.0 to 4.0. *Procedia Computer Science*, 200, 1615–1625.
- Jain, A., Menon, M. N., & Chandra, S. (2015). Sales forecasting for retail chains. *San Diego, California: UC San Diego Jacobs School of Engineering*.
- Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, 37(1), 479–489.
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26, 159–190.
- Kumar, V., Garg, H., Gandhi, A., & Gupta, B. (2023). Big Mart Sales Prediction Using Machine Learning. *International Conference on Artificial Intelligence on Textile and Apparel*, 431–443.
- Lijuan, C., Bhaumik, A., Xinfeng, W., & Jingwen, W. (n.d.). *The Effects of Inventory Management on Business Efficiency*.
- Lim, W. T., Wang, L., Wang, Y., & Chang, Q. (2016). Housing price prediction using neural networks. *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 518–522.

- Linardo, M. (2018). *Manajemen pemasaran*. Cikudanews.
- Margono, S. (2005). *Metodologi penelitian pendidikan*.
- Massaro, A., Maritati, V., & Galiano, A. (2018). Data Mining model performance of sales predictive algorithms based on RapidMiner workflows. *International Journal of Computer Science & Information Technology (IJCSIT)*, 10(3), 39–56.
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7, 21.
- Nugraha, W., & Syarif, M. (2023). Teknik Weighting untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Churn Menggunakan XGBoost, LightGBM, dan CatBoost. *Techno. Com*, 22(1).
- PRASETYO, A. T. R. I. (2017). *PENGGUNAAN JARINGAN SARAF TIRUAN (JST) UNTUK PREDIKSI KETIDAKSEIMBANGAN TEGANGAN PADA SISTEM TENAGA LISTRIK*. Universitas Muhammadiyah Semarang.
- Purnama, P. A. W., Fadhilah, C., Fadhilla, C. A., Wardana, B., Sumah, J., Thaniket, R. M., Tomasila, G., Lapatta, N. T., & Pohan, N. (2025). *Artificial Intelligence*. Serasi Media Teknologi.
- Roggeveen, A. L., & Sethuraman, R. (2020). How the COVID-19 pandemic may change the world of retailing. *Journal of Retailing*, 96(2), 169.
- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education*, 12, 8–22.
- Saepuloh, Y., & Noviardiansyah, F. (2024). Competitive Analysis of Sales and Profit Data Between ALFAMART and INDOMARET. *Jurnal Audit, Pajak, Akuntansi Publik (AJIB)*, 3(2), 97–105.
- Sahi, M. (2023). *Prediksi Harga Cryptocurrency berdasarkan model Artificial Neural Network*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Sharif, M. S., Tamang, M. T., Nepal, A., & Elmedany, W. (2024). A Comparative Study of Sales Prediction Using Machine Learning Models: Integration of PySpark and Power BI. *2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 137–142.
- Sileyew, K. J. (2019). Research design and methodology. *Cyberspace*, 1–12.
- Sumitra, I. D., & Sidqi, F. (2024). Predicting Selling Product of Single Variant Using Arima, Trend Analysis, And Single Exponential Smoothing Methods (Case Study: Swalayan Xyz Store). *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 5(1), 43–52.
- Tan, C. D., Fay, F. L., Santoso, J. C., Kurnia, T. L., & Fendelson, R. (2025). ANALISIS

KESENJANGAN PERSEDIAAN PADA PERUSAHAAN X MELALUI PENDEKATAN MANAJEMEN INVENTORI. *Jurnal Penelitian Ilmiah Multidisipliner*, 1(03), 202–215.

Tayibnapis, A. Z., Wuryaningsih, L. E., & Gora, R. (2018). The development of digital economy in Indonesia. *IJMBS International Journal of Management and Business Studies*, 8(3), 14–18.

Thomas, A. J., Petridis, M., Walters, S. D., Gheytassi, S. M., & Morgan, R. E. (2017). Two hidden layers are usually better than one. *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings*, 279–290.

Wibowo, R. A., & Rikumahu, B. (2019). Peramalan Dengan Volatilitas Frekuensi Tinggi Untuk Cryptocurrency Dan Mata Uang Konvensional Dengan Support Vector Regression Dan Regresi Linier. *EProceedings of Management*, 6(3), 5647–5652.

Wu, C.-S. M., Patil, P., & Gunaseelan, S. (2018). Comparison of different machine learning algorithms for multiple regression on black friday sales data. *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, 16–20.

Zhang, H., Xiao, B., Li, J., & Hou, M. (2021). An Improved Genetic Algorithm and Neural Network-Based Evaluation Model of Classroom Teaching Quality in Colleges and Universities. *Wireless Communications and Mobile Computing*, 2021(1), 2602385. <https://doi.org/10.1155/2021/2602385>

Zhao, X., & Keikhosrokiani, P. (2022). Sales Prediction and Product Recommendation Model Through User Behavior Analytics. *Computers, Materials & Continua*, 70(2).

Zubair, M., Waleed, A., Rehman, A., Ahmad, F., Islam, M., & Javed, S. (2024). Machine Learning Insights into Retail Sales Prediction: A Comparative Analysis of Algorithms. *2024 Horizons of Information Technology and Engineering (HITE)*, 1–7.