

TESIS

**PEMODELAN PERAMALAN PENJUALAN BUKU
MENGUNAKAN LSTM BIDIRECTIONAL: PENANGANAN
OUTLIER DAN PENGURANGAN OVERFITTING DENGAN
TEKNIK CLIPPING DAN EARLY STOPPING**

(Studi Kasus: PT Intan Pariwara)



Disusun oleh:

Nama : HENDRIANSYAH SANTOSA
NIM : 24.55.1566
Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2025

TESIS

**PEMODELAN PERAMALAN PENJUALAN BUKU
MENGUNAKAN LSTM BIDIRECTIONAL: PENANGANAN
OUTLIER DAN PENGURANGAN OVERFITTING DENGAN
TEKNIK CLIPPING DAN EARLY STOPPING
(Studi Kasus: PT Intan Pariwara)**

**BOOK SALES FORECASTING MODELING USING
BIDIRECTIONAL LSTM: OUTLIER HANDLING AND
OVERFITTING REDUCTION WITH CLIPPING AND EARLY
STOPPING TECHNIQUES**

(A Case Study of PT Intan Pariwara)

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana

Program Studi PJJ Informatika



Disusun oleh:

Nama : HENDRIANSYAH SANTOSA
NIM : 24.55.1566
Konsentrasi : Business Intelligence

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

**PEMODELAN PERAMALAN PENJUALAN BUKU MENGGUNAKAN LSTM
BIDIRECTIONAL: PENANGANAN OUTLIER DAN PENGURANGAN
OVERFITTING DENGAN TEKNIK CLIPPING DAN EARLY STOPPING (Studi
Kasus: PT Intan Pariwara)**

**BOOK SALES FORECASTING MODELING USING BIDIRECTIONAL LSTM:
OUTLIER HANDLING AND OVERFITTING REDUCTION WITH CLIPPING AND
EARLY STOPPING TECHNIQUES (A Case Study of PT Intan Pariwara)**

Dipersiapkan dan Disusun oleh

HENDRIANSYAH SANTOSA

24.55.1566

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 02 Februari 2026

Dosen Pembimbing,



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PENGESAHAN

**PEMODELAN PERAMALAN PENJUALAN BUKU MENGGUNAKAN LSTM
BIDIRECTIONAL: PENANGANAN OUTLIER DAN PENGURANGAN
OVERFITTING DENGAN TEKNIK CLIPPING DAN EARLY STOPPING (Studi
Kasus: PT Intan Pariwara)**

**BOOK SALES FORECASTING MODELING USING BIDIRECTIONAL LSTM:
OUTLIER HANDLING AND OVERFITTING REDUCTION WITH CLIPPING AND
EARLY STOPPING TECHNIQUES (A Case Study of PT Intan Pariwara)**

Dipersiapkan dan Disusun oleh

HENDRIANSYAH SANTOSA

24.55.1566

Telah dipertahankan di depan Dewan Penguji
pada tanggal 02 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Tonny Hidayat, S.Kom., M.Kom., Ph.D.
NIK. 190302182



Dhani Ariatmanto, S.Kom., M.Kom., Ph.D.
NIK. 190302197



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106



Tesis ini telah diterima sebagai salah satu persyaratan untuk memperoleh
gelar Magister Komputer Tanggal 02 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : HENDRIANSYAH SANTOSA

NIM : 24.55.1566

Pilih salah satu

Menyatakan bahwa Tesis dengan judul berikut:

PEMODELAN PERAMALAN PENJUALAN BUKU MENGGUNAKAN LSTM BIDIRECTIONAL: PENANGANAN OUTLIER DAN PENGURANGAN OVERFITTING DENGAN TEKNIK CLIPPING DAN EARLY STOPPING (Studi Kasus: PT Intan Pariwara)

Dosen Pembimbing

: Prof. Dr. Kusriani, M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini **SAYA** buat dengan **sesungguhnya**, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 02 Februari 2026

Yang Menyatakan,



HENDRIANSYAH SANTOSA

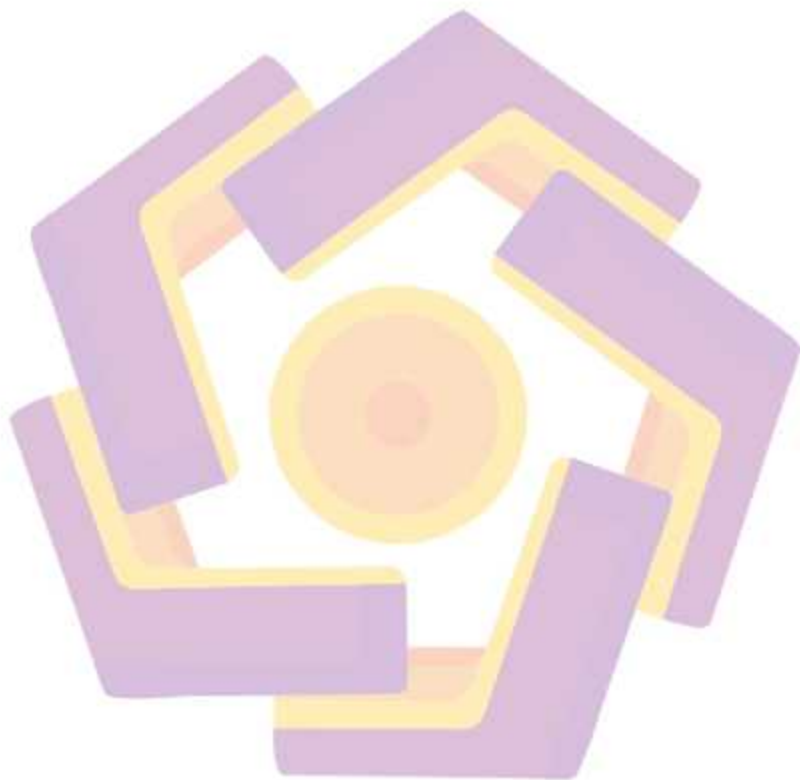
HALAMAN PERSEMBAHAN

Tesis ini dipersembahkan kepada para pendidik dan peneliti di bidang forecasting, serta kepada PT Intan Pariwara yang turut bermitra dengan Kemendikbud dalam mencerdaskan kehidupan bangsa. Dalam kapasitas sebagai Database Administrator dan Business Intelligence, penulis mempersembahkan karya ini sebagai ikhtiar kecil memperkuat praktik berbasis data di dunia pendidikan.



HALAMAN MOTTO

Kamu tidak akan tau sebelum mencobanya. Maka, jangan kalah sebelum perang.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, atas limpahan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tesis dengan judul "Pemodelan Peramalan Penjualan Buku Menggunakan LSTM Bidirectional: Penanganan Outlier dan Pengurangan Overfitting dengan Teknik Clipping dan Early Stopping".

Tesis ini disusun untuk memenuhi salah satu syarat memperoleh gelar Magister pada Program Studi S2 Teknik Informatika, Program Pascasarjana Universitas Amikom Yogyakarta. Dalam penyusunan tesis ini, penulis menyadari bahwa tanpa dukungan, bimbingan, serta doa dari berbagai pihak, tesis ini tidak dapat terselesaikan dengan baik.

Oleh karena itu, dengan segala kerendahan hati penulis menyampaikan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. **Prof. Dr. M. Suyanto, M.M.**, selaku Rektor Universitas Amikom Yogyakarta.
2. **Dr. Kusriani, M.Kom.**, selaku Direktur Program Pascasarjana Universitas Amikom Yogyakarta dan sekaligus Dosen Pembimbing, atas bimbingan, arahan, serta motivasi yang telah diberikan selama proses penelitian dan penulisan tesis ini.
3. **Tim Dosen Penguji** yang telah memberikan masukan berharga demi kesempurnaan penelitian ini.

4. **Seluruh dosen dan staf Program Studi S2 Teknik Informatika Universitas Amikom Yogyakarta** yang telah membekali penulis dengan ilmu, wawasan, serta pelayanan akademik yang baik.
5. **Orang tua, keluarga, dan istri tercinta** yang senantiasa memberikan doa, dorongan, serta semangat yang tidak pernah putus.
6. Semua pihak yang telah membantu secara langsung maupun tidak langsung dalam penyusunan tesis ini.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga tesis ini dapat memberikan manfaat, baik secara akademis maupun praktis, khususnya di bidang Business Intelligence dan penerapan deep learning dalam peramalan penjualan.

Yogyakarta, 02 Februari 2026

HENDRIANSYAH SANTOSA

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah	6
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	8
2.1. Tinjauan Pustaka	8

2.2. Keaslian Penelitian	Error! Bookmark not defined.
2.3. Landasan Teori	23
BAB III METODE PENELITIAN	27
3.1. Jenis, Sifat, dan Pendekatan Penelitian	27
3.2. Metode Pengumpulan Data	28
3.3. Metode Analisis Data	31
3.4. Skenario Pengujian Model	35
3.5. Alur Penelitian	36
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	46
BAB V PENUTUP	66
5.1. Kesimpulan	66
5.2. Saran	67
DAFTAR PUSTAKA	68
LAMPIRAN	70

DAFTAR TABEL

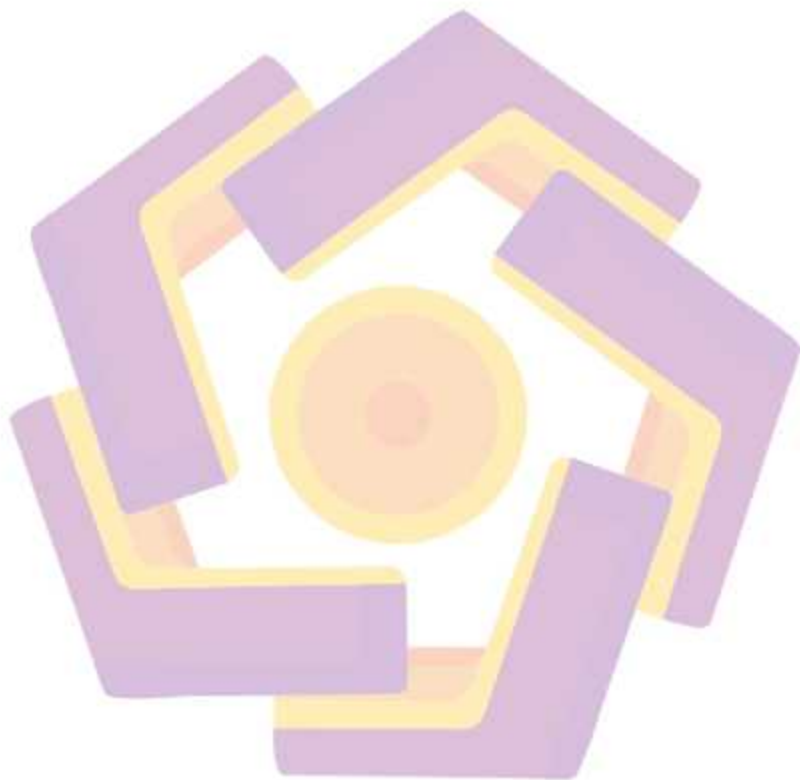
Tabel 2.1. Matriks literatur review dan posisi penelitian	11
Tabel 3.1. Keterkaitan Rumusan Masalah dengan Teknik Analisis	32
Tabel 3.2. Skenario Uji	34
Tabel 3.3. Penyusunan Skenario Pengujian Ilmiah	38
Tabel 3.4. Tabel Proses Training	40
Tabel 4.1. Fitur dan Deskripsi	44
Tabel 4.2. Parameter Pelatihan	45
Tabel 4.3. Perbandingan Jumlah Penjualan Produk Pada Beberapa Cabang	50
Tabel 4.4. Skenario Pengujian	52
Tabel 4.6 Hasil Pengujian Berulang Model Bi-LSTM dengan Clipping dan Early Stopping pada Epoch 100	55
Tabel 4.7 Perbandingan Model Bi-LSTM vs Bi-LSTM + Early Stopping vs Clipping vs Full	56
Tabel 4.6 Perbandingan Model Bi-LSTM vs Bi-LSTM + Early Stopping vs Clipping vs Full	52

DAFTAR GAMBAR

Gambar 3.1. Diagram Alur Penelitian	37
Gambar 3.2. Flowchart Proses Clipping dan Early Stoppin	44
Gambar 4.1. Tren Evaluasi MSE, MAE, RSME dan R^2	58
Gambar 4.2. Kurva Perbandingan Penjualan Aktual dan Hasil Prediksi	59



DAFTAR ISTILAH



INTISARI

Tesis ini berjudul “Pemodelan Peramalan Penjualan Buku Menggunakan LSTM Bidirectional: Penanganan Outlier dan Pengurangan Overfitting dengan Teknik Clipping dan Early Stopping”. Penelitian ini bertujuan untuk merancang model peramalan penjualan buku berbasis deep learning dengan pendekatan Bidirectional Long Short-Term Memory (Bi-LSTM). Fokus penelitian diarahkan pada penanganan data penjualan yang bersifat fluktuatif dengan keberadaan outlier serta potensi overfitting. Data yang digunakan berupa catatan penjualan harian dengan atribut waktu dan kategori. Tahapan pra-proses mencakup *feature engineering*, transformasi logaritmik, standarisasi, dan penerapan clipping pada variabel target. Dataset dibentuk dalam format deret waktu dengan pendekatan *sliding window*, sedangkan evaluasi model dilakukan menggunakan metrik MSE, MAE, RMSE, dan koefisien determinasi (R^2).

Hasil penelitian menunjukkan bahwa penerapan teknik clipping dan *early stopping* pada model Bi-LSTM mampu meningkatkan kinerja prediksi secara signifikan. Model mencapai nilai R^2 sebesar 0,87 dan RMSE sebesar 0,44, yang menandakan tingkat akurasi dan kemampuan generalisasi yang baik. Bi-LSTM terbukti efektif dalam menangkap pola nonlinier serta tren musiman pada data penjualan, sementara clipping berperan dalam mereduksi pengaruh nilai ekstrem dan *early stopping* berfungsi menjaga stabilitas proses pelatihan.

Penelitian ini menyimpulkan bahwa Bi-LSTM yang dipadukan dengan teknik clipping dan *early stopping* merupakan pendekatan yang efektif dan efisien dalam peramalan penjualan buku. Temuan ini diharapkan dapat memberikan kontribusi praktis bagi industri penerbitan dan perdagangan buku dalam optimalisasi perencanaan persediaan serta pengambilan keputusan bisnis, sekaligus memberikan kontribusi akademik pada pengembangan kajian *time-series multivariate forecasting* berbasis *deep learning*.

Kata kunci: LSTM, Bi-LSTM, clipping, early stopping, book sales forecasting.

ABSTRACT

This thesis is entitled "Book Sales Forecasting Modeling Using Bidirectional LSTM: Handling Outliers and Reducing Overfitting with Clipping and Early Stopping Techniques." This research aims to design a deep learning-based book sales forecasting model using the Bidirectional Long Short-Term Memory (Bi-LSTM) approach. The research focuses on handling fluctuating sales data with the presence of outliers and the potential for overfitting. The data used consists of daily sales records with time and category attributes. Preprocessing stages include feature engineering, logarithmic transformation, standardization, and the application of clipping to the target variable. The dataset is formatted in time series format using a sliding window approach, while model evaluation is performed using MSE, MAE, RMSE, and the coefficient of determination (R^2) metrics.

The results show that the application of clipping and early stopping techniques to the Bi-LSTM model significantly improves predictive performance. The model achieved an R^2 value of 0.87 and an RMSE of 0.44, indicating good accuracy and generalization capabilities. Bi-LSTM has proven effective in capturing nonlinear patterns and seasonal trends in sales data, while clipping helps reduce the influence of extreme values, and early stopping maintains the stability of the training process.

This study concludes that Bi-LSTM, combined with clipping and early stopping techniques, is an effective and efficient approach for book sales forecasting. These findings are expected to provide practical contributions to the publishing and book trade industries in optimizing inventory planning and business decision-making, while also contributing academically to the development of deep learning-based time-series multivariate forecasting studies.

Keyword: satu, dua, tiga, empat, lima

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Peramalan (Forecasting), sebagai proses membuat prediksi tentang kejadian di masa depan berdasarkan data masa lalu dan data saat ini, serta memainkan peran utama dalam pengambilan keputusan di berbagai domain, misalnya: bisnis, cuaca, energi, transportasi, dan hiburan (Mahmoud & Mohammed, 2024). Peramalan deret waktu (time series) digunakan untuk memprediksi nilai di masa depan. Dalam konteks ritel, peramalan penjualan membantu memprediksi penjualan sebelum terjadi, yang sangat penting untuk mendukung keberhasilan (Helmini et al., 2019). Meskipun telah banyak diteliti oleh para ahli statistik, meningkatnya kompleksitas dan volume data deret waktu dari berbagai aplikasi nyata menjadi tantangan besar bagi metode statistik tradisional (Brockwell & Davis, 2002). Oleh karena itu, adanya kebutuhan yang semakin meningkat untuk mengembangkan metode pemodelan dan peramalan deret waktu yang mampu menangani data besar dengan laju tinggi seperti yang berasal dari lingkungan dinamis, melibatkan banyak variabel.

Seiring dengan perkembangan pendekatan machine learning (ML) yang telah di adopsi untuk memodelkan hubungan yang kompleks dalam data multivariate, masih meninggalkan tantangan serius yang sepenuhnya belum terpecahkan. Salah satu isu utama adalah kompleksitas hubungan antar variabel. di mana keterkaitan temporal dan kausal antar fitur input sulit dimodelkan secara

efektif oleh algoritma ML konvensional. Selain itu, masalah seperti curse of dimensionality, perbedaan skala dan frekuensi data, serta ketidaklengkapan (missing values) memperbesar risiko overfitting dan penurunan akurasi prediksi.

Meskipun demikian, beberapa metode peramalan deret waktu telah diusulkan, termasuk pendekatan metode statistik tradisional dan pendekatan ML dalam bidang ini, seperti Algoritma Random Forest yang sudah ditingkatkan. Sayangnya algoritma sebelumnya banyak meninggalkan masalah. Seperti kemungkinan overfitting akibat penggunaan algoritma random forest yang ditingkatkan, terutama jika data pelatihan tidak cukup representatif atau beragam. Selain itu, kompleksitas model yang tinggi dapat menyebabkan waktu komputasi yang lebih lama dan sulit dalam interpretasi hasil secara menyeluruh. Algoritma ini juga mungkin kurang optimal untuk data yang sangat besar atau memiliki fitur yang sangat berkorelasi tinggi, serta kurangnya penyesuaian terhadap dinamika pasar yang cepat berubah dalam e-commerce, sehingga performa prediksi dapat menurun saat diterapkan di kondisi yang berbeda dari data pelatihan (S. Li, 2022).

Dalam konteks objek penelitian, PT Intan Pariwara sebagai perusahaan penerbit dan distributor buku menghadapi tantangan dalam melakukan peramalan penjualan yang akurat dan konsisten. Pola penjualan buku menunjukkan fluktuasi yang signifikan antar periode waktu, khususnya pada awal dan akhir semester akademik, serta pada masa menjelang tahun ajaran baru. Kondisi ini menyebabkan ketidakpastian dalam perencanaan stok, distribusi, dan pengambilan keputusan operasional. Selain itu, keberadaan nilai penjualan ekstrem pada periode tertentu berpotensi menimbulkan bias pada proses evaluasi model

peramalan, sehingga hasil prediksi yang dihasilkan kurang stabil dan sulit digeneralisasi. Oleh karena itu, diperlukan pendekatan pemodelan peramalan yang tidak hanya mampu menangkap pola temporal penjualan buku, tetapi juga memiliki mekanisme untuk menangani outlier dan mengendalikan overfitting. Penelitian ini dirancang sebagai upaya untuk mengatasi permasalahan tersebut melalui penerapan model Bidirectional Long Short-Term Memory (Bi-LSTM) yang dipadukan dengan teknik clipping dan early stopping, sehingga diharapkan dapat menghasilkan prediksi penjualan yang lebih akurat, stabil, dan aplikatif bagi kebutuhan PT Intan Pariwara.

Dalam penelitian ini, kami mengusulkan pendekatan pemodelan berbasis Long Short-Term Memory (LSTM) dengan penekanan pada strategi penerapannya, bukan pada pengembangan arsitektur baru, sebagai model utama untuk peramalan deret waktu. LSTM merupakan jenis recurrent neural network (RNN) yang sangat efektif dalam menangani data urutan (sequential data) dan mampu menangkap ketergantungan temporal yang kompleks antar variabel dalam deret waktu. Kemampuan LSTM untuk memproses dan memprediksi data time series dengan informasi historis dalam jangka panjang menjadikannya sangat relevan untuk aplikasi dalam e-commerce, yang memiliki dinamika pasar yang cepat berubah. Salah satu keunggulan utama dari LSTM adalah kemampuannya untuk belajar dari hubungan non-linear dan interaksi fitur dalam data yang kompleks, serta mengatasi masalah yang sering muncul pada algoritma machine learning (ML) konvensional, seperti curse of dimensionality dan ketidaklengkapan data. Untuk meningkatkan akurasi model, penelitian ini juga menggunakan

Bidirectional LSTM, yang memungkinkan model untuk menganalisis data dari dua arah, yaitu masa lalu dan masa depan, memberikan konteks yang lebih kaya dalam memprediksi tren penjualan. Dengan memanfaatkan teknik Dropout dan Early Stopping, risiko overfitting dapat diminimalkan, sementara performa model tetap optimal meskipun berhadapan dengan data yang besar dan bervariasi. Improvisasi yang dimaksud dalam penelitian ini dilakukan melalui integrasi teknik regulasi data dan proses pelatihan, yaitu clipping untuk mereduksi pengaruh nilai ekstrem serta early stopping untuk mengendalikan overfitting, yang dievaluasi secara sistematis melalui beberapa skenario pengujian.

Dalam konteks peramalan penjualan buku, tantangan utama tidak hanya terletak pada kemampuan model dalam menghasilkan nilai prediksi, tetapi juga pada bagaimana kinerja model tersebut dievaluasi secara konsisten. Data penjualan yang mengandung nilai ekstrem dan perubahan pola antar periode waktu berpotensi menghasilkan nilai error yang bias, terutama ketika model dievaluasi menggunakan metrik regresi yang sensitif terhadap outlier.

Oleh karena itu, diperlukan pendekatan pemodelan yang tidak hanya berfokus pada pemilihan arsitektur, tetapi juga pada proses evaluasi kinerja model yang mempertimbangkan stabilitas error dan kemampuan generalisasi. Penelitian ini memposisikan kontribusinya pada pengujian empiris penerapan teknik clipping sebagai mekanisme penanganan nilai ekstrem serta early stopping sebagai strategi pengendalian proses pelatihan, yang dievaluasi menggunakan metrik regresi seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2).

Melalui pendekatan tersebut, penelitian ini berupaya memberikan pemahaman yang lebih terstruktur dan menyeluruh mengenai perilaku error prediksi dan konsistensi hasil evaluasi model Bidirectional LSTM pada data penjualan buku, sehingga evaluasi kinerja model dapat dilakukan secara lebih objektif dan dapat dipertanggungjawabkan.

Data penjualan buku dalam penelitian ini menunjukkan variasi temporal dan pola musiman yang cukup signifikan antar periode waktu, meskipun tidak dimaksudkan untuk merepresentasikan seluruh spektrum data yang sangat dinamis dan ekstrem. Oleh karena itu, pendekatan pemodelan yang digunakan difokuskan pada kemampuan model dalam menangkap perubahan pola penjualan dan menjaga stabilitas evaluasi pada data yang fluktuatif, khususnya pada periode-periode dengan perubahan permintaan yang relatif cepat. Penggunaan data penjualan multivariat dengan berbagai fitur turunan berpotensi meningkatkan kompleksitas ruang fitur, yang dalam literatur dikenal sebagai curse of dimensionality. Kondisi ini dapat memperbesar risiko overfitting, khususnya ketika model deep learning dilatih pada dataset dengan rentang waktu terbatas.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, penelitian ini mengidentifikasi beberapa permasalahan sebagai berikut :

- Bagaimana pola penjualan buku PT Intan Pariwara berdasarkan data historis?
- Berapa tingkat evaluasi MSE, MAE, R^2 , dan RMSE dalam memprediksi jumlah penjualan buku di masa mendatang?

- c. Faktor apa saja yang memengaruhi kinerja model dalam menghasilkan evaluasi yang baik?

1.3. Batasan Masalah

Penelitian ini memiliki beberapa batasan agar lebih terfokus dan terarah :

- Data yang digunakan terbatas pada data historis penjualan buku PT Intan Pariwara selama periode Januari 2024 hingga April 2025.
- Penelitian ini memanfaatkan model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping untuk meminimalkan risiko overfitting.
- Penelitian dilakukan berdasarkan data penjualan bulanan yang dikelompokkan berdasarkan tanggal, semester, cabang, kategori, dan nama produk.
- Faktor eksternal, seperti promosi, libur nasional, atau tren pendidikan, tidak disertakan sebagai variabel dalam model peramalan.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk:

- Memberikan solusi terhadap keterbatasan akurasi dan generalisasi pada model prediksi penjualan buku yang digunakan pada penelitian sebelumnya, melalui penerapan model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping untuk meminimalkan risiko overfitting.
- Mengembangkan model prediktif yang mampu menangkap kompleksitas pola penjualan buku dengan lebih baik melalui integrasi keunggulan dari model yang diterapkan.

- c. Mengevaluasi kinerja model yang diusulkan dengan menggunakan ukuran MSE, MAE, RMSE, dan R^2 sebagai dasar untuk menetapkan keberhasilan model dalam memprediksi penjualan buku.
- d. Memiliki hasil penilaian yang lebih efisien dibandingkan dengan prosedur sebelumnya, dan menjadi komponen strategi analisis data penjualan buku yang lebih akurat dan layak untuk industri penerbitan atau distribusi.

1.5. Manfaat Penelitian

Bagian ini memuat penjelasan tentang:

- a. Berkontribusi pada pengembangan metode prediksi penjualan buku yang memanfaatkan model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping untuk meminimalkan risiko overfitting.
- b. Sistem ini mampu memprediksi penjualan buku secara efektif, sehingga memungkinkan informasi yang lebih tepat untuk membuat keputusan.
- c. Memungkinkan perusahaan untuk mengoptimalkan pemesanan dan pemasaran buku dengan lebih baik dengan presisi peramalan yang ditingkatkan, mencegah risiko kelebihan stok atau kekurangan stok.
- d. Membantu pengambilan keputusan dalam merencanakan stok buku dan meningkatkan efektivitas operasional dan profitabilitas perusahaan.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian yang dilakukan oleh Li, Zhang, Sun, dan Zhu (2024) membahas pemanfaatan algoritma ensemble, yaitu XGBoost dan Random Forest, untuk meramalkan penjualan sayuran guna mengoptimalkan perencanaan pengadaan. Pendekatan analisis dimulai dengan model analisis faktor dan korelasi Spearman untuk mengukur hubungan antar kelompok sayuran, yang menunjukkan adanya korelasi tinggi antara golongan sayuran tertentu seperti solanaceae dan akar akuatik. Hasil penelitian menunjukkan bahwa model XGBoost memiliki performa lebih unggul dengan nilai R-squared sebesar 0,945 dibandingkan Random Forest yang hanya mencapai 0,576. Selain itu, digunakan pula model deret waktu ARIMA untuk memprediksi harga grosir di masa depan dengan hasil yang cukup baik. Penelitian ini memberikan gambaran bahwa kombinasi teknik prediktif modern berbasis data dapat meningkatkan efisiensi pengambilan keputusan dalam pengelolaan stok dan pengadaan barang. Relevansinya dengan penelitian ini terletak pada penggunaan model XGBoost dan Random Forest, meskipun dalam konteks produk yang berbeda, di mana penelitian ini mengembangkan pendekatan lebih lanjut melalui teknik stacking regressor untuk meningkatkan akurasi dalam memprediksi penjualan buku (J. Li et al., 2024).

Peramalan penjualan dalam konteks e-commerce menjadi aspek krusial untuk meningkatkan efektivitas manajemen bisnis, khususnya dalam menghadapi berbagai aktivitas seperti promosi dan penggunaan konsumen yang sangat dinamis. Berbagai penelitian menyoroti pentingnya penggunaan metode data mining dan algoritma pembelajaran mesin, terutama algoritma Random Forest, untuk memproses dan menganalisis data besar yang berhubungan dengan aktivitas e-commerce. Algoritma Random Forest yang ditingkatkan melalui optimisasi hyperparameter dan pemodelan segmentasi deret waktu terbukti mampu meningkatkan akurasi prediksi penjualan, terutama untuk kategori produk yang

dipengaruhi secara signifikan oleh aktivitas seperti makanan dan pakaian. Hal ini didukung oleh pemilihan fitur yang cermat serta teknik optimisasi seperti Bayesian Optimization yang memperbaiki proses clustering dan pengelolaan data deret waktu secara lebih efisien. Selain itu, pemodelan yang mengintegrasikan data mining dengan pengolahan big data mobile e-commerce mampu mengatasi kompleksitas dan keanekaragaman data, sehingga menghasilkan model peramalan yang lebih handal dan aplikatif di dunia nyata. Dengan demikian, pendekatan berbasis Random Forest yang dioptimalkan memberikan kontribusi positif dalam meningkatkan performa prediksi yang berorientasi pada kebutuhan praktis industri e-commerce saat ini (S. Li, 2022).

Prediksi penjualan restoran merupakan tugas penting yang dipengaruhi oleh berbagai faktor kompleks seperti informasi kalender, kondisi cuaca, dan karakteristik lokasi toko, dimana interaksi antar fitur tersebut memiliki peran signifikan dalam akurasi prediksi. Metode tradisional dan model neural network konvensional seperti MLP dan LSTM sering kurang efektif dalam menangkap efek interaksi multiplikatif antar fitur, sehingga akurasi prediksi terbatas. Dalam bidang pembelajaran mesin, teknik seperti Factorization Machines, self-attention, dan graph neural networks (GNN) telah terbukti efektif untuk mempelajari interaksi fitur tingkat rendah hingga tinggi, meskipun penerapannya dalam prediksi penjualan restoran masih terbatas. Penelitian terkini mengusulkan mekanisme pembelajaran interaksi fitur berbasis message passing neural networks yang mengelompokkan fitur menjadi yang berdampak langsung dan tidak langsung terhadap penjualan, sehingga meningkatkan kemampuan model dalam menangkap interaksi dan menghasilkan prediksi yang lebih akurat. Evaluasi pada data sebenarnya menunjukkan bahwa mekanisme ini unggul dibandingkan model lain seperti MLP biasa, LSTM, GCN, dan Self-Attention, khususnya dalam konteks jumlah fitur yang terbatas dan interaksi fitur yang kompleks namun sederhana secara struktural. Dengan demikian, pembelajaran interaksi fitur menjadi pendekatan penting untuk meningkatkan performa prediksi penjualan restoran (Yang et al., 2022).

2.2. Paradigma Pembelajaran Mesin

Pembelajaran mesin (machine learning) secara umum dapat diklasifikasikan ke dalam beberapa paradigma utama, yaitu supervised learning, unsupervised learning, ensemble learning, dan reinforcement learning.

Supervised learning merupakan pendekatan pembelajaran mesin yang menggunakan data berlabel, di mana setiap data input memiliki target atau output yang telah diketahui. Pendekatan ini banyak digunakan dalam permasalahan regresi dan klasifikasi, termasuk peramalan deret waktu, karena memungkinkan model mempelajari hubungan langsung antara variabel input dan target. Model Bidirectional LSTM yang digunakan dalam penelitian ini termasuk dalam kategori supervised learning dengan pendekatan regresi time series.

Unsupervised learning adalah pendekatan pembelajaran tanpa label target, yang bertujuan menemukan pola tersembunyi dalam data, seperti clustering dan dimensionality reduction. Pendekatan ini umumnya digunakan untuk eksplorasi data, namun tidak secara langsung digunakan untuk peramalan numerik seperti pada penelitian ini.

Ensemble learning merupakan pendekatan yang menggabungkan beberapa model untuk meningkatkan akurasi dan stabilitas prediksi. Contoh pendekatan ini meliputi bagging, boosting, dan stacking. Meskipun ensemble learning terbukti efektif pada berbagai studi peramalan, penelitian ini berfokus pada optimalisasi satu arsitektur deep learning melalui teknik regulasi data dan proses pelatihan.

Reinforcement learning merupakan paradigma pembelajaran berbasis interaksi dengan lingkungan melalui mekanisme reward dan punishment. Pendekatan ini lebih umum digunakan pada permasalahan pengambilan keputusan berurutan dan kontrol, sehingga tidak diterapkan dalam konteks peramalan penjualan pada penelitian ini.

2.3. Keaslian Penelitian

Untuk memperkuat argumen mengenai keaslian dan kontribusi penelitian ini, diperlukan pemetaan komprehensif terhadap penelitian-penelitian terdahulu yang relevan. Literatur yang dikaji tidak hanya mencakup studi mengenai peramalan penjualan dengan pendekatan deep learning, tetapi juga penelitian yang membahas strategi penanganan outlier serta teknik pengurangan overfitting pada model Long Short-Term Memory (LSTM). Melalui kajian tersebut, dapat diidentifikasi celah penelitian (research gap) yang belum banyak disentuh, khususnya pada integrasi metode *clipping* dan *early stopping* dalam arsitektur **Bidirectional LSTM** untuk kasus peramalan penjualan buku.

Sebagai dasar untuk menegaskan perbedaan dan kontribusi spesifik penelitian ini, dilakukan penyusunan matriks literatur yang merangkum tujuan, metode, variabel, teknik preprocessing, serta hasil utama dari studi-studi sebelumnya. Melalui matriks ini, posisi penelitian dapat diuraikan secara sistematis sehingga terlihat jelas aspek kebaruan dalam pendekatan yang digunakan.

Selengkapnya, pemetaan tersebut disajikan dalam **Tabel 2.1 Matriks Literatur Review dan Posisi Penelitian**, yang menjadi rujukan untuk menunjukkan bagaimana penelitian ini mengisi kekosongan ilmiah yang belum terjawab pada riset-riset sebelumnya.

Tabel 2.1.Matriks literatur review dan posisi penelitian
PEMODELAN PERAMALAN PENJUALAN BUKU MENGGUNAKAN LSTM BIDIRECTIONAL: PENANGANAN OUTLIER
DAN PENGURANGAN OVERFITTING DENGAN TEKNIK CLIPPING DAN EARLY STOPPING

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Restaurant Sales Forecasting with Feature Interaction-learning Mechanism-based Neural Network Model	Xin Yang, Tetsuya Tsuboi, Keisuke Kimura, Keitarou Hangai, Maiko Shigeno, 2022 IEEE International Conference on Big Data (Big Data) , 2022	Mengembangkan model peramalan penjualan restoran yang dapat mempelajari efek interaksi fitur secara efektif menggunakan mekanisme pembelajaran interaksi fitur dalam neural network, serta mampu memanfaatkan data dari banyak toko secara bersama-sama untuk	Efek interaksi antar fitur secara signifikan mempengaruhi akurasi peramalan penjualan restoran, namun sulit dipelajari secara efisien oleh jaringan neural non-linear seperti MLP. Dengan menerapkan mekanisme pembelajaran interaksi fitur yang diusulkan, model neural network dapat meningkatkan kemampuan belajar efek interaksi	Mekanisme pembelajaran interaksi fitur yang digunakan kurang efektif ketika mencoba memasukkan interaksi antar semua fitur, terutama karena banyak fitur tidak saling berinteraksi, sehingga model bisa berkinerja lebih buruk dibanding model MLP sederhana. Untuk mengatasi hal ini, penelitian menyarankan	Penelitian yang dilakukan oleh Xing Yang dan rekan-rekannya berfokus pada eksplorasi efektivitas interaksi fitur dalam model prediktif. Sementara itu, penelitian dalam tesis ini menitikberatkan pada evaluasi kinerja model melalui metrik kuantitatif seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), dan Root Mean Squared Error (RMSE). Perbedaan fokus ini menunjukkan keaslian dan kontribusi khas dari penelitian yang dilakukan.

			meningkatkan akurasi ramalan penjualan harian.	tersebut, sehingga menghasilkan peningkatan akurasi yang signifikan dibandingkan model lain pada data penjualan restoran dari 100 toko.	penggunaan metode berbasis machine learning yang dapat secara otomatis mendeteksi dan memanfaatkan efek interaksi fitur, terutama saat jumlah fitur bertambah banyak dan membangun graf interaksi secara manual menjadi tidak praktis.	
2	Sales Forecasting Model of E-commerce Activities Based on Improved Random Forest Algorithm	Suang Li, 2022 2nd International Conference on Computer Graphics, Image and Virtualization (ICCGIV), 2022	Mengembangkan model prediksi penjualan mobile e-commerce yang lebih akurat dengan memanfaatkan teknologi data mining, khususnya algoritma random forest yang telah	model prediksi penjualan mobile e-commerce yang menggunakan algoritma random forest yang diperbaiki dapat secara efektif meningkatkan akurasi prediksi penjualan dengan mempertimbangkan parameter dan data	Penelitian ini menyarankan pengembangan lebih lanjut pada model dengan mengintegrasikan parameter dan data yang lebih beragam serta pengujian pada skala dan konteks e-commerce yang lebih luas untuk	Penelitian yang dilakukan oleh Suang Li berfokus pada analisis variasi volume penjualan yang dipengaruhi oleh aktivitas e-commerce secara signifikan. Berbeda dengan penelitian tersebut, penelitian dalam tesis ini memusatkan perhatian pada evaluasi performa model prediksi menggunakan metrik Mean Squared Error

			diperbaiki. Penelitian ini fokus pada analisis karena variasi volume penjualan terkait aktivitas e-commerce yang signifikan, dan bertujuan menyediakan model prediksi yang dapat memfasilitasi pemahaman lebih baik terhadap pola konsumsi konsumen serta membantu dalam pengambilan keputusan bisnis yang lebih tepat sasaran.	yang relevan. Model ini mampu menangani variabilitas data dan ketergantungan fitur yang kompleks, sehingga hasil prediksi lebih dapat diandalkan untuk membantu analisis dampak aktivitas e-commerce terhadap volume penjualan dan rekomendasi pengelolaan penjualan yang lebih baik.	meningkatkan generalisasi hasil. Kelemahan penelitian terletak pada keterbatasan data historis dan sampel yang digunakan, serta kemungkinan kurangnya penerapan pada berbagai jenis produk dan kondisi pasar yang berbeda, sehingga model perlu diuji dan disesuaikan lebih lanjut agar optimal dalam berbagai situasi nyata.	(MSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), dan Root Mean Squared Error (RMSE) pada dataset berukuran besar. Perbedaan fokus ini menegaskan kontribusi orisinal dari penelitian yang dilakukan.
3	Vegetable Sales Forecasting Based on	- Jun Li - Jinyang Zhang - Mengxin Sun	Memprediksi penjualan sayuran dan	model regresi XGBOOST menunjukkan kinerja	Meskipun model XGBOOST memberikan	Penelitian oleh Jun Li dan rekan-rekan berfokus pada analisis hubungan

XGBOOST Algorithm and Random Forest	- Rongyu Zhu Media Publikasi : 2024 International Conference on Computers, Information Processing and Advanced Education (CIPAE), Tahun : 2024	mengoptimalkan strategi pengadaan menggunakan algoritma XGBOOST dan random forest. Studi ini menganalisis hubungan penjualan antar berbagai jenis sayuran melalui analisis faktor dan korelasi Spearman, kemudian membangun model regresi untuk memprediksi jumlah penjualan sayuran. Dengan demikian, penelitian ini memberikan wawasan bagi supermarket dan	yang lebih baik dibandingkan model random forest dalam memprediksi penjualan sayuran, dengan nilai R-kuadrat sebesar 0,945. Selain itu, analisis deret waktu juga mampu memberikan prediksi harga grosir sayuran yang akurat. Secara keseluruhan, penelitian ini menyediakan kerangka kerja yang berguna bagi supermarket dan pengecer dalam pengambilan keputusan yang lebih informasi untuk penjualan dan pengadaan bahan.	prediksi yang sangat akurat, nilai R-kuadrat untuk model ARIMA yang digunakan dalam analisis deret waktu relatif sedang (0,619), menunjukkan bahwa ada ruang untuk peningkatan dalam pemodelan harga grosir. Oleh karena itu, penelitian selanjutnya bisa mempertimbangkan penggunaan metode atau fitur tambahan untuk meningkatkan akurasi model prediksi dan memperluas cakupan variabel yang dianalisis guna mengatasi ketidakpastian	penjualan antar jenis sayuran menggunakan analisis faktor dan korelasi Spearman, serta membangun model regresi untuk prediksi penjualan. Sebaliknya, penelitian tesis ini menitikberatkan pada prediksi penjualan buku serta evaluasi kinerja model menggunakan metrik MSE, MAE, R^2, dan RMSE dengan model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping untuk meminimalkan risiko overfitting.
--	---	--	---	---	---

			pengecer untuk pengambilan keputusan yang lebih tepat dalam perencanaan penjualan dan pengadaan, sehingga dapat meningkatkan efektivitas respons terhadap permintaan pasar dan keuntungan bisnis.		dalam data penjualan dan harga.	
4	A Drug Sales Forecasting Method Based on Multiple Models and Parameter Optimization	Dawei Liu Media Publikasi: Computers & Chemical Engineering: An International Journal of Computer Applications in Chemical Engineering. Tahun : 2023	mengembangkan metode peramalan penjualan obat yang efektif bagi jaringan apotek dengan menggunakan struktur optimasi dua lapis, yaitu pemilihan model dan optimasi parameter.	metode peramalan penjualan obat berbasis struktur optimasi dua lapis—menggunakan multi-model deep learning dengan integrasi Bagging dan optimasi parameter melalui algoritma genetika—dapat secara efektif meningkatkan	Penelitian ini menyarankan agar penelitian lanjutan fokus pada pengembangan ekstraksi fitur untuk memungkinkan sampling dinamis dari karakteristik obat dan mengeksplorasi metode optimasi parameter yang	Penelitian oleh Dawei Liu menggunakan struktur optimasi dua lapis, yaitu pemilihan model dan optimasi parameter. Metode ini menggabungkan multi-model berbasis deep learning serta algoritma genetika yang ditingkatkan untuk mengatasi keragaman data akibat perbedaan daerah,

			Metode ini menggabungkan multi-model berbasis deep learning serta algoritma genetika yang ditingkatkan untuk mengatasi keragaman data akibat perbedaan daerah, karakteristik penyakit, pengguna, dan jenis obat. Selain itu, penelitian ini memperkenalkan metode SS-KSMOTE untuk mengatasi masalah sampel tidak seimbang dan sparsity, sehingga dapat meningkatkan akurasi peramalan	akurasi prediksi di jaringan apotek dengan kondisi data yang kompleks dan tidak seimbang. Metode SS-KSMOTE juga terbukti efisien dalam mengatasi masalah sparsity sampel, sehingga keseluruhan pendekatan ini memenuhi kebutuhan prediksi penjualan obat secara akurat dalam berbagai scenario.	lebih baik guna meningkatkan performa model. Kelemahan yang diidentifikasi adalah bahwa model yang berbeda memiliki keunggulan yang bervariasi tergantung pada wilayah dan jenis obat, sehingga masih diperlukan penyesuaian model yang lebih spesifik serta eksplorasi lebih lanjut dalam optimasi parameter untuk kondisi yang lebih kompleks.	karakteristik penyakit, pengguna, dan jenis obat. Sedangkan penelitian tesis ini menggunakan model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping dalam memprediksi jumlah penjualan buku di masa mendatang.
--	--	--	--	--	---	---

			penjualan obat pada skenario kompleks jaringan apotek.			
5	Big Data Modeling Based on KNN-RF-SVM and Its Application in Product Sales Forecasting Field	Peneliti : Chen Zhang, Hongru Ren, Xiuhua Lu, Qunyao Yuan, Renquan Lu Media Publikasi : International Conference on Intelligent Manufacturing and Industrial Big Data (ICIMIBD) Tahun : 2022	Penelitian ini bertujuan untuk meningkatkan akurasi prediksi penjualan mobil dengan mengembangkan model gabungan KNN-RF-SVM menggunakan metode soft voting, sehingga dapat memberikan hasil prediksi yang lebih akurat dan terpercaya dibandingkan model algoritma tunggal, serta membantu perusahaan otomotif dalam	Model gabungan KNN-RF-SVM yang dibangun menggunakan metode soft voting memiliki performa yang lebih baik dibandingkan model algoritma tunggal dalam hal akurasi, presisi, recall, dan skor F1 untuk prediksi penjualan mobil. Model gabungan ini lebih cocok untuk rekomendasi penjualan mobil dan memiliki potensi untuk diterapkan di bidang lain, sehingga meningkatkan	Penelitian ini menyarankan agar akurasi prediksi dapat ditingkatkan dengan mengubah bobot masing-masing algoritma dalam kombinasi melalui metode soft voting. Selain itu, kelemahan model gabungan yang saat ini adalah rendahnya akurasi prediksi serta potensi kehilangan informasi selama pelatihan akibat banyaknya algoritma yang digabungkan, yang dapat diatasi dengan mengganti	Penelitian oleh Chen Zhang dan rekan-rekan berfokus pada peningkatan akurasi prediksi penjualan mobil melalui pengembangan model gabungan KNN-RF-SVM dengan metode soft voting. Sementara itu, penelitian tesis ini bertujuan memperoleh hasil evaluasi terbaik berdasarkan metrik MSE, MAE, R^2, dan RMSE, sehingga menunjukkan perbedaan fokus dan kontribusi orisinal dalam konteks pemodelan prediktif.

			menyesuaikan strategi penjualan secara efektif dan memperluas penerapan model prediksi di berbagai bidang terkait.	cakupan aplikasi dan kepraktisan model prediksi.	atau menyesuaikan sub-algoritma gabungan agar model menjadi lebih efisien dan mengurangi waktu proses.	
6	New Energy Vehicle Sales Forecasting Based on MICMAC Algorithm and Gray Forecasting Model	Peneliti : Jinyi Xie, Zhongming Cai, Jiawei Xia, dan Haoyuan Xue Media Publikasi : 2024 International Conference on Telecommunications and Power Electronics (TELEPE) Tahun : 2024	Mengidentifikasi dan menganalisis faktor-faktor kunci yang mempengaruhi perkembangan kendaraan listrik energi baru di Tiongkok, serta memprediksi tren masa depan produksi kendaraan dan jumlah infrastruktur pengisian daya menggunakan model prediksi	Dengan menggunakan analisis literatur, algoritma MICMAC, dan model struktur interpretatif (ISM), telah diidentifikasi 14 faktor utama yang memengaruhi perkembangan kendaraan listrik energi baru di Tiongkok, yang dibagi dalam lima lapisan hierarki berdasarkan pengaruh dan ketergantungan	Pengembangan industri kendaraan listrik di Tiongkok perlu memperhatikan desain kebijakan tingkat atas, riset teknologi inti, dan peningkatan infrastruktur sebagai kunci untuk mencapai perkembangan berkualitas tinggi. Namun, kelemahan penelitian ini mungkin terletak pada	Penelitian yang dilakukan Jinyi Xie dkk menggunakan model prediksi gray yang dikombinasikan dengan algoritma MICMAC dan model struktur interpretatif (ISM). Sedangkan penelitian tesis ini menggunakan model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping.

			gray yang dikombinasikan dengan algoritma MICMAC dan model struktur interpretatif (ISM). Dengan pendekatan ini, penelitian berupaya memberikan pemahaman sistematis tentang interaksi antar faktor serta memberikan prediksi jangka panjang yang akurat untuk mendukung perkembangan industri kendaraan listrik energi baru.	mereka. Data produksi kendaraan listrik dan jumlah stasiun pengisian publik dari periode sebelumnya telah berhasil dimodelkan menggunakan model prediksi gray, yang menunjukkan bahwa pada tahun 2032, produksi kendaraan listrik dan jumlah stasiun pengisian akan tumbuh secara moderat mencapai masing-masing sekitar 16,200,180 unit dan 2,982,354 unit. Penelitian ini menegaskan pentingnya perancangan top-level, riset teknologi inti, dan peningkatan infrastruktur untuk mendukung perkembangan	ketergantungan data historis terbatas dan asumsi linearitas dalam model prediksi gray, yang dapat membatasi akurasi prediksi di masa depan mengingat dinamika cepat dan kompleksitas faktor eksternal yang memengaruhi industri kendaraan Listrik.	
--	--	--	---	--	---	--

				berkualitas tinggi industri kendaraan Listrik.		
7	TSFRN: Integrated Time and Spatial-Frequency domain based on Triple-links Residual Network for Sales Forecasting	Peneliti : Yi Xiang, Haoran Sun, Wenting Tu, dan Zejin Tian. Media Publikasi : Proceedings of the 2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI) Tahun : 2023	Mengembangkan model deep learning baru yang disebut TSFRN untuk peramalan penjualan, dengan memanfaatkan mekanisme attention yang menggabungkan informasi interaksi dalam domain frekuensi dan spasial. Pendekatan ini bertujuan untuk meningkatkan akurasi prediksi dengan menangkap hubungan kompleks dalam	Model TSFRN yang mengintegrasikan informasi domain waktu dan frekuensi melalui jaringan residual triple-link serta mekanisme spatial-frequency attention secara efektif meningkatkan akurasi peramalan penjualan dibandingkan model-model lain seperti LSTM dan DeepAR. Model ini mampu menangkap pola interaksi produk baik dalam domain waktu maupun frekuensi, dan hasil eksperimen pada dua dataset nyata	Meningkatkan interpretabilitas dan akurasi prediksi TSFRN dengan menggabungkannya menggunakan model peramalan probabilistik serta memperluas model agar mampu mempertimbangkan variabel eksternal seperti harga, kondisi cuaca, dan tingkat promosi guna mendapatkan pendekatan peramalan penjualan yang lebih komprehensif dan akurat. Kelemahan yang diidentifikasi adalah penurunan akurasi	Penelitian Yi Xiang dkk berfokus mengembangkan model deep learning TSFRN. Sedangkan penelitian tesis ini berfokus pada kolaborasi model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping.

		data waktu dan ruang yang sebelumnya kurang diperhatikan dalam metode-metode tradisional maupun berbasis domain waktu saja.	menunjukkan keunggulannya. Penelitian juga menunjukkan bahwa akurasi prediksi menurun seiring bertambahnya delay waktu, dan rencana pengembangan ke depan meliputi peningkatan interpretabilitas serta penggabungan dengan model probabilistik dan variabel eksternal untuk prediksi yang lebih komprehensif.	prediksi seiring bertambahnya delay waktu (jangka panjang), serta adanya keterbatasan dalam mempertimbangkan faktor eksternal yang dapat mempengaruhi penjualan.	
--	--	--	--	---	--

2.4. Landasan Teori

Penjualan buku merupakan aktivitas ekonomi yang dipengaruhi oleh berbagai faktor seperti musim pembelajaran, tren literasi, kampanye pemasaran, harga, serta ketersediaan stok. Dalam konteks industri penerbitan dan distribusi buku, pola permintaan dapat bersifat fluktuatif dan sulit diprediksi. Buku memiliki karakteristik produk yang spesifik karena bisa bersifat musiman (misalnya buku ajar sekolah), memiliki umur produk yang panjang, dan terkadang bergantung pada kebijakan pendidikan. Oleh karena itu, pendekatan prediksi penjualan buku harus mampu mengakomodasi kompleksitas data dan pola historis yang tidak selalu linier (Makridakis et al., 2008).

Data mining adalah proses menemukan pola atau informasi bermakna dari kumpulan data besar menggunakan teknik statistik, machine learning, dan algoritma komputasional. Salah satu aplikasi penting dari data mining adalah prediksi, yaitu memperkirakan kejadian atau nilai di masa depan berdasarkan pola data historis. Dalam konteks bisnis, prediksi penjualan menjadi alat strategis untuk mendukung perencanaan produksi, pengelolaan stok, serta pengambilan keputusan pemasaran (Berry & Linoff, 2009). Teknik prediksi modern memanfaatkan model statistik hingga algoritma pembelajaran mesin, yang semakin relevan untuk mengelola data penjualan berskala besar dan kompleks.

Random Forest merupakan algoritma pembelajaran ansambel berbasis decision tree yang digunakan baik untuk klasifikasi maupun regresi. Algoritma ini bekerja dengan membuat sejumlah pohon keputusan secara acak pada subset data latih dan menggabungkan hasil prediksi dari seluruh pohon untuk menghasilkan

output akhir. Pada regresi, hasilnya diperoleh dari rata-rata seluruh prediksi pohon (Yu, 2020). Keunggulan Random Forest terletak pada kemampuannya mengurangi overfitting dan menangani data berdimensi tinggi. Namun, kelemahannya antara lain adalah kebutuhan komputasi yang cukup besar dan interpretasi model yang kurang transparan dibanding model linier.

Extreme Gradient Boosting (XGBoost) adalah algoritma boosting yang sangat efisien dan akurat, dikembangkan berdasarkan prinsip Gradient Boosted Decision Trees (GBDT). XGBoost bekerja dengan membangun pohon keputusan secara berurutan, di mana setiap pohon baru dibuat untuk memperbaiki kesalahan dari pohon sebelumnya. Keunggulan utama XGBoost terletak pada efisiensinya dalam pemrosesan data, regularisasi untuk mencegah overfitting, serta kemampuan menangani missing value dan data yang tidak teratur (Wade & Glynn, 2020). Algoritma ini sering digunakan dalam berbagai kompetisi data science karena performanya yang tinggi dalam tugas prediksi dan klasifikasi.

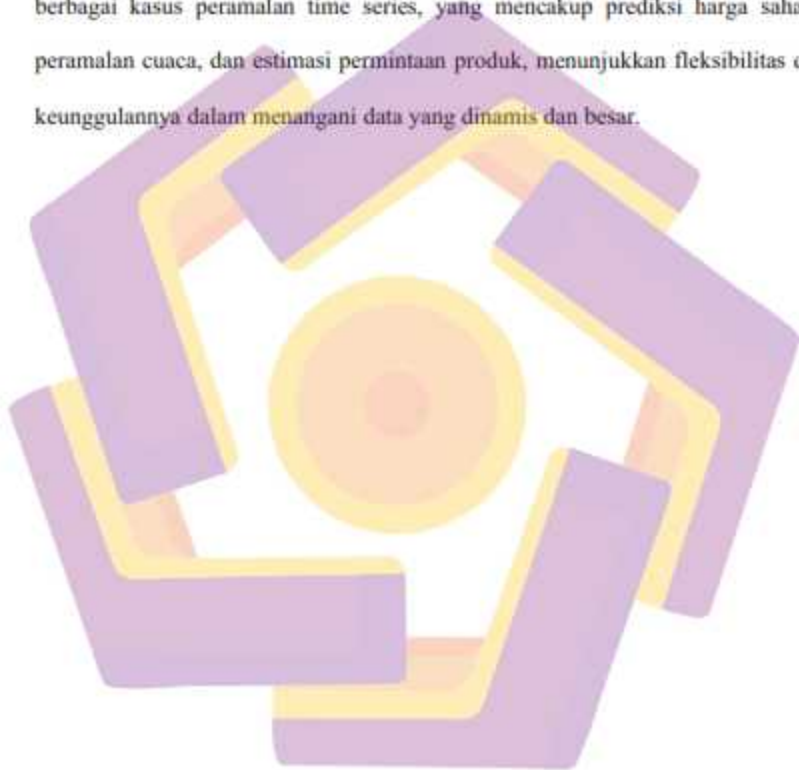
Ensemble learning adalah pendekatan dalam machine learning yang menggabungkan beberapa model pembelajaran dasar (base learners) untuk meningkatkan akurasi prediksi. Gagasan dasarnya adalah bahwa kombinasi beberapa model biasanya menghasilkan kinerja yang lebih baik dibandingkan satu model tunggal. Metode ensemble populer mencakup bagging (misalnya Random Forest), boosting (misalnya XGBoost), dan stacking. Ensemble dapat mengurangi varians, bias, dan meningkatkan generalisasi model terhadap data baru (Zhou-ISBN, 2012).

Stacking (stacked generalization) adalah metode ensemble yang menggabungkan berbagai model dasar (base models) dengan cara melatih model tingkat kedua (meta-model) untuk mempelajari kombinasi terbaik dari prediksi-prediksi model dasar tersebut. Dalam konteks regresi, Stacking Regressor memungkinkan integrasi berbagai algoritma regresi seperti Random Forest dan XGBoost sehingga keunggulan masing-masing model dapat saling melengkapi. Pendekatan ini terbukti meningkatkan akurasi prediksi, terutama pada data kompleks dan tidak linear, dibandingkan dengan menggunakan satu model Tunggal (Zhou-ISBN, 2012).

Evaluasi model prediksi dilakukan untuk mengukur seberapa akurat model dalam menghasilkan nilai yang mendekati data aktual. Beberapa metrik umum dalam regresi adalah Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R-squared (R^2). MAE mengukur rata-rata kesalahan absolut, RMSE memberikan penalti lebih besar terhadap kesalahan besar, sedangkan R-squared menunjukkan seberapa besar variasi data yang mampu dijelaskan oleh model (Ruppert, 2004). Pemilihan metrik tergantung pada konteks bisnis dan karakteristik data yang digunakan.

Long Short-Term Memory (LSTM) merupakan jenis recurrent neural network (RNN) yang dirancang untuk mengatasi keterbatasan pada model RNN konvensional, seperti vanishing gradient problem, yang sering muncul pada data urutan panjang. LSTM memiliki kemampuan untuk mengingat informasi jangka panjang dalam time series data, menjadikannya sangat efektif dalam peramalan data yang memiliki ketergantungan temporal yang kuat. Dalam Deep Learning for

Time Series Forecasting (Brownlee, 2018), dijelaskan bahwa LSTM mampu mempelajari hubungan non-linear dan interaksi kompleks antar fitur dalam data deret waktu, yang membuatnya lebih unggul dibandingkan model tradisional seperti ARIMA. Selain itu, buku ini juga menyoroti penerapan LSTM dalam berbagai kasus peramalan time series, yang mencakup prediksi harga saham, peramalan cuaca, dan estimasi permintaan produk, menunjukkan fleksibilitas dan keunggulannya dalam menangani data yang dinamis dan besar.



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini termasuk dalam kategori penelitian eksperimen dan penelitian tindakan (action research), yang melibatkan serangkaian percobaan untuk membuktikan suatu konsep melalui implementasi model prediksi penjualan menggunakan LSTM Bidirectional dengan teknik clipping dan early stopping. Eksperimen dilakukan untuk menguji hubungan sebab-akibat antara input variabel (data penjualan historis) dan output (peramalan penjualan). Penelitian ini bertujuan untuk meningkatkan akurasi model prediksi penjualan buku dan mengurangi risiko overfitting. Oleh karena itu, penelitian ini termasuk dalam jenis penelitian eksperimen yang menguji efektivitas teknik pengolahan data dan penerapan model deep learning.

Penelitian ini bersifat deskriptif dan eksplanatif. Sifat deskriptif digunakan untuk menggambarkan karakteristik data penjualan buku yang digunakan dalam model, serta untuk menganalisis pola dan tren yang muncul dari data historis. Di sisi lain, sifat eksplanatif digunakan untuk menjelaskan hubungan kausal antara berbagai variabel input (misalnya: tanggal, kategori buku, semester) dan variabel output (penjualan buku). Dengan kata lain, penelitian ini tidak hanya menggambarkan fenomena yang ada, tetapi juga menjelaskan bagaimana variabel-variabel tersebut saling berinteraksi dalam menghasilkan prediksi yang lebih baik menggunakan model Bi-LSTM.

Pendekatan yang digunakan dalam penelitian ini adalah kuantitatif dengan pendekatan pemodelan prediktif berbasis deep learning. Penelitian ini mengandalkan metode kuantitatif karena melibatkan analisis data numerik untuk memprediksi penjualan buku di masa mendatang berdasarkan data historis yang telah dikumpulkan. Data yang digunakan dianalisis menggunakan model Long Short-Term Memory (LSTM) Bidirectional yang diimprovisasi dengan teknik dropout dan early stopping untuk meningkatkan akurasi serta mencegah overfitting. Penelitian ini berfokus pada penggunaan algoritma machine learning dalam meramalkan nilai penjualan berdasarkan variabel yang bersifat temporal.

Berdasarkan paradigma machine learning, penelitian ini termasuk dalam supervised learning karena proses pelatihan model dilakukan menggunakan data berlabel dengan target jumlah penjualan buku sebagai variabel dependen.

Berdasarkan jenis permasalahan yang diselesaikan, penelitian ini termasuk dalam supervised learning dengan pendekatan regresi. Hal ini dikarenakan variabel target yang diprediksi berupa nilai numerik kontinu, yaitu jumlah penjualan buku pada periode waktu tertentu. Berbeda dengan permasalahan klasifikasi yang bertujuan mengelompokkan data ke dalam kelas-kelas diskrit, penelitian ini berfokus pada estimasi nilai kuantitatif sehingga metode regresi time series menjadi pendekatan yang relevan.

3.2. Metode Pengumpulan Data

Untuk pengumpulan data difokuskan untuk menjawab pertanyaan pada rumusan masalah, yang dijabarkan sebagai berikut:

- a. Untuk mengetahui pola penjualan buku PT Intan Pariwara berdasarkan data historis, dibutuhkan data penjualan dalam bentuk time series multivariat. Data ini diperoleh dari sistem basis data internal perusahaan menggunakan query SQL. Selanjutnya dilakukan proses pembersihan (cleaning) dan transformasi data untuk keperluan analisis.
- b. Untuk mengevaluasi tingkat akurasi model prediktif dan faktor yang memengaruhinya, dilakukan pengumpulan data tambahan melalui teknik rekayasa fitur. Fitur yang digunakan meliputi waktu (hari, minggu, bulan), kategori buku, wilayah penjualan, dan variabel lain yang berkaitan dengan kinerja model.

Untuk mengetahui pola penjualan buku berdasarkan data historis dilakukan langkah sebagai berikut :

- a. Melakukan pembersihan data untuk mengatasi nilai kosong dan duplikasi.
- b. Melakukan transformasi data waktu (time series) agar sesuai dengan kebutuhan analisis.
- c. Melakukan eksplorasi data (EDA) guna mengidentifikasi pola musiman dan tren penjualan.
- d. Menyusun visualisasi berupa grafik time series untuk memperjelas hasil temuan.

Untuk mengevaluasi performa model prediktif berdasarkan data yang telah diolah, dilakukan langkah-langkah sebagai berikut:

- a. Membagi dataset menjadi data latih dan uji berdasarkan pembagian waktu (time-based split).

- b. Membangun model LSTM yang diimprovisasi dengan teknik Dropout dan Early Stopping untuk meminimalkan risiko overfitting.
- c. Mengukur performa model dengan menggunakan metrik MAE, MSE, RMSE, dan koefisien determinasi (R^2).
- d. Melakukan perbandingan kinerja model LSTM dan LSTM yang diimprovisasi.

Untuk menganalisis faktor-faktor yang memengaruhi performa model prediktif berdasarkan data yang tersedia, dilakukan langkah-langkah sebagai berikut:

- a. Melakukan analisis tingkat kepentingan fitur (feature importance).
- b. Melakukan tuning parameter menggunakan metode grid search untuk melihat sensitivitas performa model.
- c. Melakukan eksperimen dengan penghapusan atau penambahan fitur tertentu untuk mengukur dampaknya terhadap performa model.

Untuk menguji performa model secara menyeluruh, penelitian ini menerapkan beberapa skenario pengujian dengan tahapan sebagai berikut:

- a. Membangun dan menguji model LSTM yang diimprovisasi dengan Dropout dan Early Stopping.
- b. Melakukan pengujian kombinasi fitur berdasarkan hasil feature selection.
- c. Melakukan tuning parameter untuk mengevaluasi pengaruh terhadap performa.

- d. Melakukan variasi rentang waktu data latih dan uji untuk melihat kestabilan model.

3.3. Metode Analisis Data

Metode analisis data dalam penelitian ini disusun untuk menjawab setiap rumusan masalah secara terarah dan sistematis. Analisis dilakukan dalam tiga bagian utama sesuai fokus penelitian, yaitu: (1) analisis pola penjualan buku, (2) analisis evaluasi kinerja model prediksi, dan (3) analisis faktor-faktor yang memengaruhi performa model.

A. Analisis untuk Menjawab Rumusan Masalah 1

"Bagaimana pola penjualan buku PT Intan Pariwara berdasarkan data historis?"

Langkah analisis yang dilakukan:

1. Exploratory Data Analysis (EDA)

- a) Menghitung statistik deskriptif (mean, median, varian, dan distribusi penjualan).
- b) Mengidentifikasi musiman (seasonality) berdasarkan bulan, semester, dan kategori produk.

2. Visualisasi Time Series

- a) Plot tren penjualan harian/bulanan
- b) Analisis perubahan tren sebelum dan setelah periode akademik tertentu

3. Analisis Pola Perilaku Pasar

- a) Mengamati lonjakan permintaan pada waktu tertentu

- b) Identifikasi cabang/kategori dengan kontribusi dominan

Hasil dari tahap ini digunakan sebagai dasar rekayasa fitur dan pemodelan prediktif.

B. Analisis untuk Menjawab Rumusan Masalah 2

"Berapa tingkat akurasi prediksi penjualan yang dicapai model?"

Langkah analisis:

1. **Pembentukan data time series untuk supervised learning**
 - a) Sliding window (5 time-steps)
2. **Pelatihan Model**
 - a) Bidirectional LSTM + Dropout + Early Stopping
3. **Evaluasi Model Menggunakan Metrik Regresi**
 - a) MSE, MAE, RMSE, R^2
4. **Validasi Hasil**
 - a) Perbandingan prediksi vs data aktual (plot kurva)
 - b) Identifikasi periode deviasi terbesar
5. **Interpretasi Hasil**
 - a) Menyimpulkan tingkat akurasi dan generalisasi model berdasarkan nilai evaluasi terbaik

C. Analisis untuk Menjawab Rumusan Masalah 3

"Faktor apa saja yang memengaruhi kinerja model prediksi?"

Tahapan analisis:

1. Analisis Feature Contribution

- a) Evaluasi pengaruh fitur waktu & kategori melalui:
 - 1) Korelasi fitur-target
 - 2) Evaluasi performa saat fitur ditambah/dihapus

2. Evaluasi Dampak Data Preprocessing

- a) Perbandingan model:
 - 1) tanpa clipping vs dengan clipping
 - 2) tanpa early stopping vs dengan early stopping

3. Analisis Parametrik Model

- a) Pengaruh jumlah epoch & dropout terhadap overfitting

4. Generalization Behavior Test

- a) Uji dengan rentang waktu data latih–uji berbeda

Untuk memastikan bahwa langkah-langkah penelitian disusun secara sistematis dan selaras dengan tujuan yang ingin dicapai, penelitian ini terlebih dahulu merumuskan beberapa pertanyaan penelitian yang menjadi dasar alur analisis. Setiap rumusan masalah kemudian dipetakan dengan teknik analisis yang digunakan, jenis data atau proses yang terlibat, serta skenario uji yang dilakukan untuk memperoleh jawaban empiris yang valid. Pendekatan ini penting agar rancangan metodologis tidak hanya bersifat deskriptif, tetapi juga menunjukkan keterkaitan logis antara permasalahan, metode, dan hasil yang diharapkan.

Pemetaan tersebut dilakukan untuk memastikan bahwa setiap aspek permasalahan memperoleh perlakuan analitis yang tepat, mulai dari upaya

memahami pola data historis melalui *Exploratory Data Analysis* (EDA), hingga evaluasi kinerja model Bidirectional LSTM menggunakan berbagai metrik akurasi. Selain itu, analisis komparatif seperti *feature ablation*, penggunaan teknik *clipping*, serta *early stopping* dirancang untuk mengidentifikasi faktor-faktor yang memengaruhi performa model secara lebih mendalam.

Keseluruhan hubungan ini disusun secara ringkas dan terstruktur dalam **Tabel**

3.1 Keterkaitan Rumusan Masalah dengan Teknik Analisis, sehingga pembaca dapat melihat alur pemikiran penelitian secara menyeluruh, mulai dari pertanyaan awal hingga keluaran yang dihasilkan dari setiap tahap analisis.

Tabel 3.1. Keterkaitan Rumusan Masalah dengan Teknik Analisis

Rumusan Masalah	Teknik Analisis/Data	Skenario Uji	Output/Hasil
Bagaimana pola penjualan buku berdasarkan data historis?	Exploratory Data Analysis (EDA); statistik deskriptif, analisis tren, seasonality	Plot time series per kategori & cabang	Informasi pola dasar: periode puncak/penurunan penjualan → digunakan untuk rekayasa fitur
Berapa tingkat akurasi model prediksi penjualan?	Bidirectional LSTM dengan loss MSE; evaluasi MSE, MAE, RMSE, R^2	Variasi epoch (10, 15, 20, 50, 100)	Nilai akurasi model → menunjukkan keberhasilan peramalan
Faktor apa saja yang memengaruhi kinerja model?	Feature ablation, clipping-no clipping comparison, early stopping-no early stopping comparison	Perbandingan tabel hasil setiap skenario	Insight faktor penting & regulasi model

3.4. Skenario Pengujian Model

Untuk memperoleh pemahaman yang komprehensif mengenai pengaruh teknik regulasi dan penanganan outlier terhadap performa model Bidirectional LSTM, penelitian ini menetapkan sejumlah skenario uji yang dirancang secara sistematis. Setiap skenario mewakili kombinasi tertentu dari teknik preprocessing dan mekanisme regularisasi yang diterapkan, sehingga memungkinkan analisis terstruktur terhadap perubahan akurasi, stabilitas pelatihan, serta kecenderungan model mengalami overfitting atau underfitting.

Perancangan skenario uji ini tidak hanya bertujuan untuk membandingkan efektivitas masing-masing teknik secara terpisah, tetapi juga untuk mengevaluasi interaksi antar-teknik, seperti penerapan *clipping* bersamaan dengan *early stopping*, atau perbedaan hasil ketika outlier dipertahankan vs. dibersihkan. Dengan pendekatan ini, penelitian dapat mengidentifikasi kombinasi optimal yang memberikan performa prediktif terbaik sekaligus menjaga generalisasi model.

Seluruh pengaturan dan variasi tersebut dirangkum secara jelas dalam **Tabel 3.2 Skenario Uji**, yang menyajikan struktur uji secara ringkas sehingga pembaca dapat memahami alur eksperimental dan rasionalisasi di balik pemilihan setiap skenario.

Tabel 3.2.Skenario Uji

Kode Skenario	Konfigurasi Model	Tujuan Evaluasi	Indikator Keberhasilan
S1 – Baseline	Bi-LSTM tanpa clipping & tanpa early stopping	Mengetahui performa awal model	Perbandingan dasar akurasi
S2 – Outlier Handling	Bi-LSTM + clipping	Mengurangi distorsi akibat nilai ekstrem	Penurunan MSE/RMSE

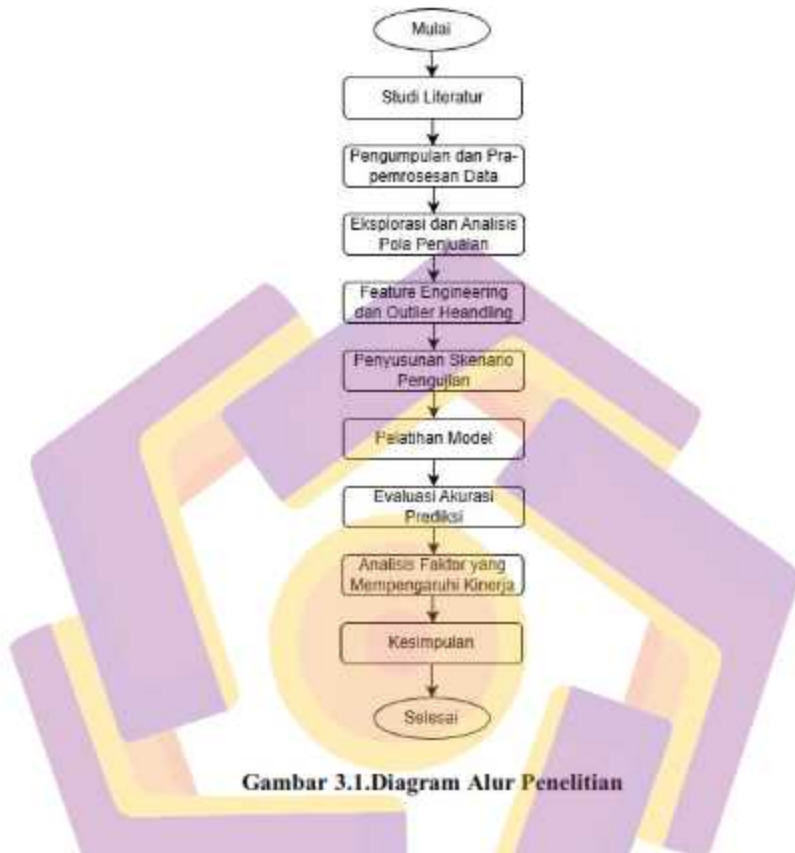
S3 – Overfitting Reduction	Bi-LSTM + early stopping	Meningkatkan generalisasi model	Gap train-test mengecil
S4 – Full Optimization	Bi-LSTM + clipping + early stopping	Konfigurasi terbaik	Pencapaian evaluasi terbaik

3.5. Alur Penelitian

Untuk memberikan pemahaman yang lebih komprehensif mengenai proses metodologis yang diterapkan dalam penelitian ini, diperlukan visualisasi alur kerja yang menggambarkan setiap tahapan secara runtut. Diagram alur penelitian berfungsi sebagai representasi sistematis dari langkah-langkah yang dilakukan, mulai dari pengumpulan dan pengolahan data, analisis eksploratori, perancangan model, hingga evaluasi performa dan penarikan kesimpulan. Penyajian diagram ini bertujuan agar pembaca dapat melihat hubungan logis antar-tahapan serta memahami bagaimana setiap proses memberikan kontribusi terhadap tujuan penelitian.

Selain itu, diagram alur berperan untuk memastikan bahwa keseluruhan metodologi dapat direplikasi dan diverifikasi, baik oleh peneliti lain maupun pihak yang berkepentingan. Dengan struktur visual tersebut, alur penelitian menjadi lebih jelas, terukur, dan transparan, sehingga mendukung integritas ilmiah dari penelitian ini.

Secara lengkap, tahapan metodologis yang dimaksud ditampilkan dalam **Gambar 3.1 Diagram Alur Penelitian**, yang menggambarkan urutan proses dari awal hingga akhir penelitian secara ringkas namun informatif.



1. Studi Literatur

Tahapan ini dilakukan untuk memperoleh dasar teoritis dan memperkuat argumentasi metodologis, yaitu mencakup:

- Konsep peramalan deret waktu dalam industri ritel dan penerbitan
- Penerapan deep learning, khususnya **Bi-LSTM**
- Teknik mitigasi **outlier** dan **overfitting**
- Evaluasi model prediksi: MSE, MAE, RMSE, dan R^2

- e. Kajian empiris penelitian sebelumnya sebagai pembanding (state-of-the-art)

Output:

- a. Gap penelitian
- b. Rumusan masalah dan hipotesis metodologis

2. Pengumpulan dan Pra-pemrosesan Data

Sumber data berasal dari sistem database perusahaan.

Proses dilakukan secara sistematis:

- a. Ekstraksi data penjualan (SQL)
- b. Penghapusan duplikasi
- c. Handling missing values
- d. Standardisasi format waktu (datetime indexing)

Output:

Dataset terstruktur dalam format time-series multivariat.

3. Eksplorasi dan Analisis Pola Penjualan

Dilakukan Exploratory Data Analysis (EDA) untuk memahami karakteristik deret waktu:

Analisis mencakup:

- a. Tren umum penjualan
- b. Pola musiman (semester, awal tahun ajaran)
- c. Lonjakan akibat faktor eksternal
- d. Kontribusi berdasarkan cabang/kategori produk

Output:

Insight yang dikonversi menjadi fitur input untuk model prediksi.

4. Feature Engineering dan Outlier Heandling

Transformasi data untuk meningkatkan kualitas pembelajaran model:

Teknik yang digunakan:

- a. Ekstraksi fitur waktu:
dayofweek, month, semester
- b. Agregasi historis:
rolling mean (window 5)
- c. Label encoding untuk data kategorik
- d. Log transform target variabel
- e. Clipping (Winsorization) → menjaga stabilitas distribusi

Output:

Fitur yang relevan & target yang robust terhadap ekstrem.

5. Penyusunan Skenario Pengujian

Dalam rangka memastikan bahwa model yang dikembangkan teruji secara ilmiah dan mampu memberikan hasil yang dapat dipertanggungjawabkan, penelitian ini menyusun serangkaian skenario pengujian yang dirancang secara sistematis. Setiap skenario mewakili pendekatan analitis tertentu yang berkaitan langsung dengan rumusan masalah penelitian, sehingga evaluasi performa model dapat dilakukan secara terarah dan berbasis bukti.

Penyusunan skenario ini bertujuan untuk mengidentifikasi performa dasar model tanpa intervensi, menilai pengaruh teknik penanganan outlier melalui *clipping*, menguji efektivitas *early stopping* dalam mengurangi gejala overfitting, serta mengevaluasi kombinasi kedua teknik untuk memperoleh konfigurasi model yang paling optimal. Dengan demikian, pengujian tidak hanya menilai hasil akhir, tetapi juga memberikan pemahaman yang lebih mendalam mengenai mekanisme internal model dan faktor-faktor yang memengaruhi kinerjanya.

Rangkaian skenario tersebut dirangkum secara terstruktur dalam **Tabel 3.3 Penyusunan Skenario Pengujian Ilmiah**, yang menampilkan tujuan masing-masing skenario serta keterkaitannya dengan rumusan masalah. Tabel ini memudahkan pembaca untuk memahami bagaimana setiap skenario berkontribusi pada pemecahan masalah penelitian secara keseluruhan.

Tabel 3.3. Penyusunan Skenario Pengujian Ilmiah

Skenario	Tujuan	Rumusan Masalah (RM)
S1 – Baseline	Menilai performa dasar Bi-LSTM	RM2
S2 – Clipping	Menganalisis pengaruh outlier	RM3
S3 – Early Stopping	Mengurangi overfitting	RM3
S4 – Kombinasi	Model terbaik (optimal)	RM2 & RM3

Output:

Rancangan eksperimen yang terukur & dapat direplikasi.

6. Pelatihan Model

Untuk memastikan model peramalan yang dibangun memiliki performa yang optimal dan proses pelatihannya berlangsung secara terstruktur, penelitian

ini merancang konfigurasi *training* yang disesuaikan dengan karakteristik data serta tujuan analisis. Setiap komponen dalam proses pelatihan—mulai dari arsitektur model, mekanisme regularisasi, pemilihan algoritma optimasi, hingga strategi pembagian data—ditentukan berdasarkan pertimbangan teoretis dan bukti empiris dari penelitian terdahulu.

Arsitektur model menggunakan susunan *Bidirectional LSTM* diikuti oleh lapisan *LSTM* dan *Dense*, yang memungkinkan model menangkap pola sekuensial jangka pendek maupun jangka panjang. Penerapan *dropout* sebagai teknik regularisasi dimaksudkan untuk mengurangi risiko overfitting, sementara penggunaan *Adam optimizer* dengan *MSE loss* dipilih karena stabilitas dan efektivitasnya dalam memproses data deret waktu. Selain itu, pembagian data secara time-based serta penerapan *early stopping* berperan penting dalam menjaga validitas hasil pelatihan serta mencegah model berlatih terlalu lama pada pola yang tidak relevan.

Seluruh konfigurasi tersebut dirangkum secara ringkas dan sistematis dalam **Tabel 3.4 Tabel Proses Training**, yang memberikan gambaran menyeluruh mengenai parameter dan mekanisme yang digunakan selama proses pelatihan model.

Tabel 3.4. Tabel Proses Training

Komponen	Detail
Arsitektur	Bidirectional LSTM → LSTM → Dense
Regularisasi	Dropout rate 0.2
Optimizer & Loss	Adam + MSE loss
Data Split	Time-based: Train 80% – Test 20%

Stopping rule	Early Stopping (patience 10)
---------------	------------------------------

Output:

Model prediktif untuk memproduksi nilai forecast.

7. Evaluasi Akurasi Prediksi

Evaluasi pada setiap skenario menggunakan metrik:

- MSE: penalti kesalahan besar
- RMSE: interpretasi pada skala asli
- MAE: stabil untuk error linear
- R^2 : tingkat variasi yang dijelaskan model

Visualisasi kurva aktual vs prediksi.

Pemilihan model dengan performa terbaik.

Output:

Kesimpulan akurasi dan generalisasi model

8. Analisis Faktor yang Mempengaruhi Kinerja

Studi komparatif antar skenario:

Analisis yang dilakukan:

- Dampak clipping terhadap nilai ekstrim
- Pengaruh early stopping pada gap train-test
- Perubahan performa ketika fitur ditambah/diablas

Output:

Identifikasi faktor kunci peningkatan performa model

9. Kesimpulan

Berisi kontribusi ilmiah dan saran penelitian lanjutan:

- Bi-LSTM efektif untuk prediksi penjualan buku
- Clipping & early stopping → meningkatkan stabilitas model

- c. Rekomendasi integrasi variabel eksternal (promosi, kebijakan akademik, dsb.)

Output:

Jawaban final atas seluruh rumusan masalah

Untuk memperjelas mekanisme penerapan teknik clipping dan early stopping dalam proses pemodelan peramalan penjualan, diperlukan visualisasi alur proses yang menggambarkan tahapan pelatihan model secara lebih rinci. Visualisasi ini bertujuan untuk menunjukkan secara eksplisit pada tahapan mana clipping diterapkan sebagai mekanisme penanganan nilai ekstrem, serta bagaimana early stopping berperan dalam mengendalikan proses pelatihan model berdasarkan kinerja pada data validasi.

Flowchart proses disusun untuk merepresentasikan alur pemodelan yang menyerupai kondisi implementasi nyata, dimulai dari input data penjualan historis, tahap praproses dan deteksi outlier, hingga proses pelatihan model Bidirectional LSTM yang berlangsung secara iteratif pada setiap epoch. Pada setiap iterasi, kinerja model dievaluasi menggunakan nilai validation loss, yang selanjutnya menjadi dasar pengambilan keputusan untuk melanjutkan atau menghentikan pelatihan melalui mekanisme early stopping.

Dengan adanya flowchart ini, diharapkan pembaca dapat memahami secara sistematis keterkaitan antara proses preprocessing data, pelatihan model, serta mekanisme regulasi yang diterapkan, sehingga alur metodologis penelitian menjadi lebih transparan dan mudah direplikasi. Alur lengkap proses pemodelan ditunjukkan pada Gambar 3.2.

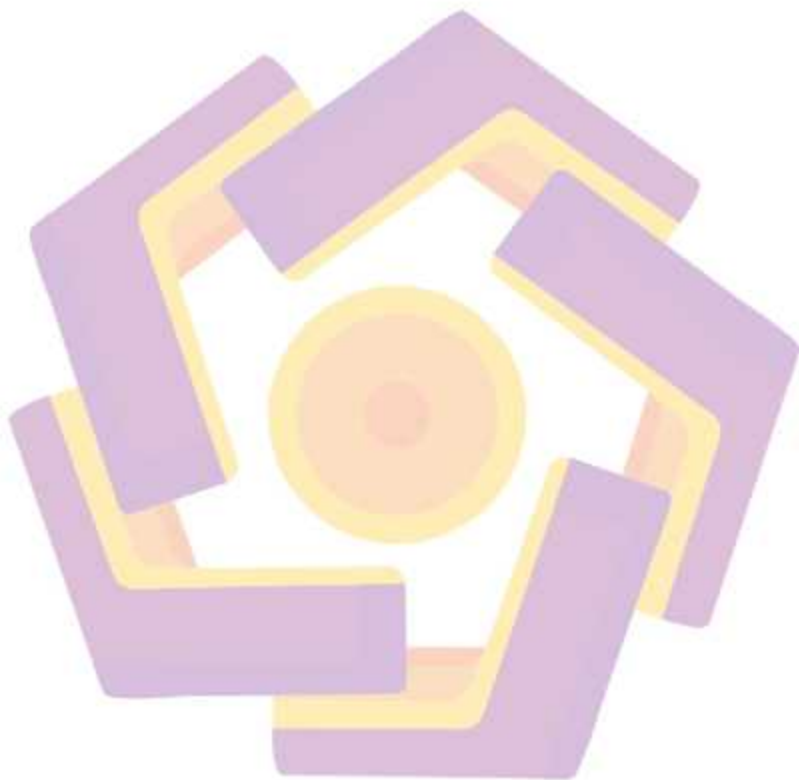


Gambar 3.2. Flowchart Proses Clipping dan Early Stopping

Gambar 3.2 menunjukkan alur proses pemodelan peramalan penjualan menggunakan Bidirectional LSTM yang dilengkapi dengan mekanisme clipping dan early stopping. Proses diawali dengan input data penjualan historis, dilanjutkan dengan tahap preprocessing dan deteksi outlier. Pada tahap ini, clipping diterapkan untuk membatasi nilai target agar tetap berada dalam rentang yang stabil. Selanjutnya, data yang telah diproses digunakan dalam pelatihan model Bi-LSTM.

Selama proses pelatihan, kinerja model dievaluasi pada setiap epoch menggunakan nilai validation loss. Apabila validation loss tidak mengalami perbaikan dalam sejumlah epoch tertentu sesuai parameter patience, maka mekanisme early stopping akan menghentikan proses pelatihan secara otomatis.

Dengan demikian, model tidak dilatih secara berlebihan dan risiko overfitting dapat dikendalikan.



BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengumpulan Data

Data penelitian diperoleh dari sistem basis data internal PT Intan Pariwara melalui proses ekstraksi menggunakan SQL. Data mentah mencakup tanggal transaksi, semester, cabang penjualan, kategori produk, nama produk, dan jumlah penjualan harian. Total data mentah berjumlah ± ribuan baris transaksi yang mencakup periode Januari 2024 – April 2025.

Tahap pengumpulan data dilakukan melalui langkah berikut:

1. Ekstraksi Data dari Database

- Query SQL mengambil data berdasarkan periode dan kategori produk tertentu.
- Format tanggal dinormalisasi menjadi datetime agar sesuai dengan kebutuhan analisis time series.

2. Penyaringan (Filtering)

- Menghapus transaksi non-penjualan (retur, koreksi).
- Menjaga hanya produk dengan frekuensi penjualan signifikan.

3. Audit Kualitas Data

- Mengecek **missing values** pada kolom jumlah penjualan dan kategori.
- Memastikan tidak ada duplikasi record berdasarkan kunci (tanggal, produk, cabang).

Hasil akhir proses pengumpulan data berupa dataset multivariat dengan struktur stabil dan siap untuk preprocessing.

4.2 Tahap Preprocessing Data

Tahapan praproses dilakukan untuk memastikan data dalam kondisi optimal untuk model LSTM. Berikut adalah tahapan detail yang dilakukan.

4.2.1 Handling Missing Values

1. Missing values ditemukan pada kolom cabang dan kategori (<1%).
2. Penanganan:
 - a. **Jumlah penjualan kosong** → diisi 0 (penjualan tidak terjadi).
 - b. **Kategori/cabang kosong** → diisi dengan mode (kategori paling sering).

4.2.2 Deteksi dan Penanganan Outlier

Outlier teridentifikasi melalui:

1. Boxplot menggunakan distribusi log-transform.
2. Nilai ekstrem ditemukan pada periode semester dan awal tahun ajaran.

Teknik penanganan:

1. **Clipping / Winsorization** pada persentil 1–99:

```
yclip ~ clip(y,p1,p99)
```

Ini menjaga kestabilan data dan menghindari lonjakan ekstrem yang memicu overfitting.

4.2.3 Feature Engineering

Untuk mendukung kemampuan model dalam menangkap pola temporal dan variasi musiman pada data penjualan, dilakukan proses *feature engineering*.

guna memperkaya informasi yang tersedia pada setiap observasi. Beberapa fitur tambahan dibangun untuk merepresentasikan aspek waktu, tren jangka pendek, serta karakteristik kategori dan cabang penjualan. Rincian fitur yang digunakan beserta deskripsinya disajikan dalam **Tabel 4.1 Fitur dan Deskripsi**, yang memberikan gambaran mengenai peran masing-masing fitur dalam memperkuat kualitas representasi data sebelum proses pemodelan.

Tabel 4.1. Fitur dan Deskripsi

Fitur	Deskripsi
dayofweek	Hari ke-0 s/d ke-6
month	Bulan transaksi
semester	Semester akademik
rolling_mean_5	Rata-rata 5 hari terakhir
kategori/cabang	Di-encode numerik

Teknik encoding:

1. **Label Encoding** → semester, cabang, kategori, produk

4.2.4 Transformasi Logaritmik & Standardisasi

1. Target penjualan ditransformasi dengan:

$$y' = \log(1+y)$$

2. Semua fitur dinormalisasi dengan **StandardScaler**.

4.2.5 Pembentukan Struktur Time Series (Sliding Window)

Dataset diubah untuk memenuhi format input LSTM:

Window = 5 time-steps

Contoh:

| t-5 | t-4 | t-3 | t-2 | t-1 | Target(t) |

Sehingga bentuk input menjadi:

(samples,5,features)

4.3 Pembangunan dan Pelatihan Model Bi-LSTM

Model yang digunakan terdiri dari:

1. **Bi-LSTM layer (100 units, return_sequences=True)**

Belajar pola temporal dua arah.

2. **LSTM layer (50 units)**

Memperdalam pemahaman urutan data.

3. **Dropout 0.2**

Mencegah overfitting dengan menghilangkan 20% neuron saat training.

4. **Dense output (1 unit)**

Untuk memastikan proses pelatihan model berlangsung secara optimal dan terkontrol, penelitian ini menetapkan serangkaian parameter pelatihan yang disesuaikan dengan karakteristik data dan tujuan pemodelan. Parameter-parameter tersebut mencakup algoritma optimasi, fungsi loss, ukuran batch, batas maksimum epoch, serta mekanisme *early stopping* untuk mencegah overfitting. Rincian lengkap konfigurasi yang digunakan disajikan dalam **Tabel 4.2 Parameter Pelatihan**, yang menjadi acuan utama dalam pelaksanaan proses training model.

Tabel 4.2. Parameter Pelatihan

Parameter	Nilai
Optimizer	Adam
Loss	MSE
Batch size	32
Epoch max	100
Early Stopping	patience=10 (monitor val_loss)

Early stopping menghentikan training ketika validasi tidak membaik, sehingga mengurangi overfitting dan mempercepat proses.

4.4 Analisis Perbedaan Data Penjualan Antar Cabang

Sebelum dilakukan pemodelan dan pengujian kinerja model peramalan, penting untuk memahami karakteristik dasar dari dataset penjualan yang digunakan dalam penelitian ini. Salah satu karakteristik utama yang ditemukan pada data penjualan PT Intan Pariwara adalah adanya perbedaan skala penjualan yang sangat signifikan antar cabang, meskipun produk yang dianalisis adalah sama dan berada pada periode waktu yang seragam.

Perbedaan skala ini menunjukkan bahwa dataset penjualan bersifat heterogen, di mana setiap cabang memiliki tingkat permintaan yang sangat bervariasi tergantung pada kondisi wilayah, jumlah sekolah, serta dinamika pasar lokal. Kondisi tersebut menyebabkan distribusi data penjualan tidak merata dan berpotensi menimbulkan nilai ekstrem pada sebagian observasi.

Untuk memberikan gambaran empiris mengenai perbedaan data yang sangat signifikan tersebut, Tabel 4.3 menyajikan perbandingan jumlah penjualan produk yang sama pada beberapa cabang PT Intan Pariwara.

Tabel 4.3. Perbandingan Jumlah Penjualan Produk Pada Beberapa Cabang

Nama Produk	Cabang	Jumlah Penjualan (Eksemplar)
Bahasa Indonesia: Kawan Seiring untuk SD/MI Kelas III	Sintang	1
Bahasa Indonesia: Kawan Seiring untuk SD/MI Kelas III	Karawang	229
Bahasa Indonesia: Kawan Seiring untuk SD/MI Kelas III	Wonosobo	913
Bahasa Indonesia: Kawan Seiring untuk	Bontang	2119

SD/MI Kelas III		
-----------------	--	--

Tabel 4.3 menunjukkan adanya perbedaan jumlah penjualan yang sangat signifikan pada dataset penjualan buku PT Intan Pariwara, meskipun produk yang diamati adalah sama. Sebagai contoh, produk “Bahasa Indonesia: Kawan Seiring untuk SD/MI Kelas III” tercatat hanya terjual sebanyak 1 eksemplar pada cabang Sintang, sementara pada cabang Karawang terjual 229 eksemplar, pada cabang Wonosobo terjual 913 eksemplar, dan pada cabang Bontang mencapai 2119 eksemplar dalam periode pengamatan yang sama.

Perbedaan skala penjualan yang sangat lebar antar cabang ini menunjukkan bahwa dataset penjualan memiliki karakteristik heterogenitas yang tinggi, baik dari sisi distribusi nilai maupun variasi antar entitas. Kondisi tersebut menyebabkan data penjualan tidak bersifat homogen dan cenderung mengandung nilai ekstrem (outlier) jika dianalisis secara agregat.

Heterogenitas data seperti ini menjadi tantangan utama dalam pemodelan peramalan penjualan, karena model prediktif harus mampu menangkap pola penjualan dengan skala kecil maupun besar secara bersamaan tanpa bias terhadap cabang tertentu. Oleh karena itu, diperlukan pendekatan pemodelan yang mampu menangani perbedaan skala data secara efektif, baik melalui arsitektur model yang sesuai maupun melalui teknik praproses dan regulasi seperti clipping dan early stopping yang diterapkan dalam penelitian ini.

4.5 Pengujian Model Berdasarkan Skenario

Untuk mengevaluasi pengaruh teknik regulasi dan penanganan outlier terhadap kinerja model, penelitian ini menerapkan empat skenario pengujian yang dirancang sesuai metodologi. Setiap skenario merepresentasikan kombinasi berbeda dari penggunaan *clipping* dan *early stopping*, sehingga memungkinkan analisis terstruktur mengenai kontribusi masing-masing teknik maupun efek keduanya ketika diterapkan secara bersamaan. Rincian setiap skenario disajikan dalam **Tabel 4.4 Skenario Pengujian**, yang menjadi dasar dalam membandingkan performa model pada berbagai konfigurasi eksperimen.

Tabel 4.4. Skenario Pengujian

Kode	Deskripsi
S1	Baseline (tanpa clipping & tanpa early stopping)
S2	Baseline + Clipping
S3	Baseline + Early Stopping
S4	Clipping + Early Stopping (Full Optimization)

4.5.1 Skenario 1 – Baseline

Tanpa regularisasi:

1. Model cepat overfitting.
2. Gap train-test besar.
3. Nilai error tinggi.

Hasil:

1. MSE: tinggi (sekitar >1)
2. Kurva prediksi kurang stabil

4.5.2 Skenario 2 – Penerapan Clipping

Clipping meredam nilai ekstrem yang memengaruhi training.

Hasil:

1. MSE dan RMSE menurun signifikan.
2. Model lebih stabil, terutama pada puncak penjualan.

4.5.3 Skenario 3 – Early Stopping

Early stopping menghentikan training pada epoch ~25–30.

Efek:

1. Menghindari overfitting.
2. Generalisasi lebih baik.
3. Nilai validasi membaik dibanding baseline.

4.5.4 Skenario 4 – Clipping + Early Stopping (Full Optimization)

Hasil lengkap dari proses evaluasi tersebut disajikan dalam **Tabel 4.5. Hasil Evaluasi Model Bidirectional LSTM**, yang merangkum performa model pada setiap konfigurasi epoch. Tabel ini menjadi dasar analisis untuk menentukan jumlah epoch yang paling efektif dalam menghasilkan prediksi dengan tingkat akurasi yang optimal.

Tabel 4.5. Hasil Evaluasi Model Bidirectional LSTM

Evaluation				
Epoch	MSE	MAE	RMSE	R ²
10	1.07	0.75	1.03	0.48
15	0.88	0.64	0.94	0.58
20	0.68	0.54	0.82	0.66
50	0.35	0.4	0.59	0.78

100	0.19	0.32	0.44	0.87
-----	------	------	------	------

$R^2 = 0.87$: model menjelaskan 87% variasi data aktual

RMSE = 0.44: error rendah dan stabil

Model mampu mengikuti pola musiman dan fluktuasi harian

Pengujian jumlah epoch dilakukan secara bertahap untuk mengamati perilaku pembelajaran model Bidirectional LSTM terhadap data pelatihan dan data uji. Pada epoch rendah (10–15), model belum mampu menangkap pola temporal secara optimal, yang ditunjukkan oleh nilai error yang masih tinggi dan nilai R^2 yang relatif rendah. Kondisi ini mengindikasikan terjadinya underfitting, di mana kapasitas pembelajaran model belum mencukupi untuk merepresentasikan kompleksitas data penjualan.

Seiring dengan peningkatan jumlah epoch (20–50), kinerja model menunjukkan perbaikan yang signifikan. Nilai MSE, MAE, dan RMSE menurun secara konsisten, sementara nilai R^2 meningkat, menandakan bahwa model mulai mampu mempelajari pola musiman dan tren penjualan secara lebih stabil.

Pada epoch yang lebih tinggi (100), model mencapai performa terbaik dengan nilai R^2 sebesar 0,87 dan RMSE sebesar 0,44. Namun demikian, pengujian tidak dihentikan hanya pada titik tertinggi performa, melainkan dianalisis secara menyeluruh untuk memastikan bahwa peningkatan kinerja bersifat stabil dan bukan hasil fluktuasi sesaat. Pendekatan ini penting untuk menghindari kesimpulan yang bias terhadap satu konfigurasi epoch tertentu serta memastikan kemampuan generalisasi model.

Berdasarkan hasil pengujian, titik terendah performa model terjadi pada epoch 10, yang ditandai dengan nilai error tertinggi. Sebaliknya, titik tertinggi performa dicapai pada epoch 100. Analisis terhadap rentang epoch tersebut menunjukkan bahwa peningkatan kinerja model berlangsung secara gradual dan konsisten. Oleh karena itu, pemilihan epoch optimal didasarkan pada tren pembelajaran secara keseluruhan, bukan semata-mata pada satu titik maksimum performa.

Untuk memastikan bahwa performa model yang diperoleh tidak bergantung pada satu kali proses pelatihan saja, dilakukan pengujian berulang sebanyak 15 kali pada konfigurasi terbaik, yaitu model Bidirectional LSTM dengan clipping dan early stopping pada epoch 100. Setiap pengujian menggunakan inisialisasi bobot yang berbeda, kemudian dievaluasi menggunakan metrik MSE, MAE, RMSE, dan R^2 .

Hasil pengujian menunjukkan bahwa nilai evaluasi pada setiap run berada pada rentang yang relatif sempit, dengan fluktuasi yang tidak signifikan antar pengujian. Nilai rata-rata MSE sebesar 1.01, MAE sebesar 0.65, RMSE sebesar 0.57, dan R^2 sebesar 0.77 menunjukkan bahwa performa model bersifat stabil dan konsisten. Hal ini mengindikasikan bahwa hasil prediksi yang diperoleh bukan merupakan kebetulan dari satu kali pelatihan, melainkan mencerminkan kemampuan generalisasi model yang baik. Hasil pengujian berulang model Bidirectional LSTM dengan clipping dan early stopping pada epoch 100 disajikan pada Tabel 4.6.

Tabel 4.6 Hasil Pengujian Berulang Model Bi-LSTM dengan Clipping dan Early Stopping pada Epoch 100

Run Ke	MSE	MAE	R^2	RMSE
1	0.96	0.61	0.9	0.42
2	0.99	0.63	0.89	0.43
3	1.01	0.6	0.88	0.44
4	0.99	0.66	0.86	0.4
5	1.05	0.68	0.9	0.46
6	1	0.64	0.88	0.44
7	0.98	0.62	0.86	0.43
8	1.02	0.65	0.88	0.45
9	1.06	0.73	0.9	0.42
10	1.01	0.65	0.88	0.44
11	0.95	0.6	0.68	0.88
12	1.02	0.66	0.56	0.78
13	1.03	0.61	0.49	0.67
14	1.06	0.69	0.52	0.89
15	1.07	0.76	0.48	1.03
Average	1.01	0.65	0.77	0.57

Tabel 4.7 menyajikan perbandingan kinerja model Bidirectional LSTM pada empat skenario pengujian, yaitu model tanpa regularisasi, model dengan early stopping, model dengan clipping, serta kombinasi clipping dan early stopping. Seluruh model diuji pada epoch yang sama, yaitu epoch 100, untuk memastikan perbandingan dilakukan secara adil dan konsisten.

Hasil pengujian menunjukkan bahwa model tanpa teknik regulasi menghasilkan nilai error yang relatif tinggi, yang mengindikasikan adanya pengaruh outlier dan kecenderungan overfitting. Penerapan early stopping memberikan perbaikan kinerja, namun belum mampu menurunkan error secara signifikan. Sebaliknya, penerapan clipping menghasilkan penurunan nilai MSE dan RMSE yang sangat signifikan, menunjukkan efektivitas teknik tersebut dalam menstabilkan distribusi target.

Kombinasi clipping dan early stopping menghasilkan performa terbaik dengan nilai MSE sebesar 1,01, MAE sebesar 0,65, RMSE sebesar 0,57, dan R^2 sebesar 0,77. Hasil ini menunjukkan bahwa integrasi teknik penanganan outlier dan mekanisme pengendalian overfitting mampu meningkatkan akurasi dan generalisasi model secara optimal. Perbandingan kinerja model Bidirectional LSTM pada keempat skenario pengujian disajikan pada Tabel 4.7.

Tabel 4.7 Perbandingan Model Bi-LSTM vs Bi-LSTM + Early Stopping vs Clipping vs Full

Model	MSE	MAE	R^2	RMSE
Bi-LSTM	4414.54	26.37	0.35	66.44
Bi-LSTM + Early Stopping	4030.24	24.78	0.4	63.48
Bi-LSTM + Clipping	1.17	0.78	0.43	1.08
Bi-LSTM + Clipping + Early Stopping	1.01	0.65	0.77	0.57

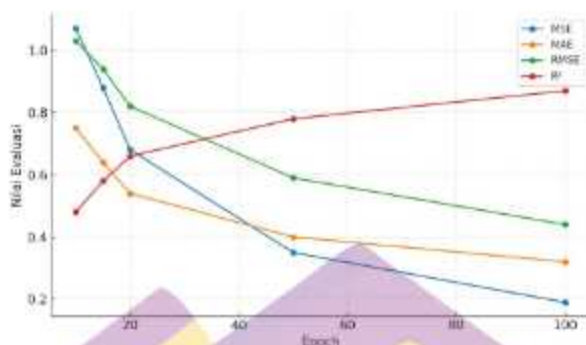
4.6 Visualisasi Kinerja Model

1. Tren Evaluasi

Untuk memberikan pemahaman yang lebih komprehensif mengenai perubahan kinerja model pada berbagai jumlah epoch, hasil evaluasi yang sebelumnya disajikan dalam bentuk tabel divisualisasikan dalam diagram tren. Visualisasi ini bertujuan untuk memperlihatkan pola penurunan error serta peningkatan akurasi model secara lebih jelas, sehingga dinamika proses pembelajaran Bidirectional LSTM dapat diamati dengan lebih intuitif.

Diagram tren memungkinkan pembaca melihat secara langsung bagaimana metrik error MSE, MAE, dan RMSE mengalami penurunan seiring bertambahnya epoch, yang menunjukkan bahwa model semakin mampu mempelajari pola data historis secara lebih efektif. Di sisi lain, kenaikan nilai R^2 menggambarkan meningkatnya kemampuan model dalam menjelaskan variabilitas data aktual. Melalui grafik ini, kecenderungan model menuju kondisi optimal dapat diidentifikasi dengan lebih mudah dibandingkan hanya melalui interpretasi numerik dari tabel.

Secara lengkap, pola perubahan setiap metrik evaluasi pada berbagai konfigurasi epoch digambarkan dalam **Gambar 4.1 Tren Evaluasi MSE, MAE, RMSE, dan R^2** , yang menjadi dasar bagi analisis lebih lanjut terkait pemilihan jumlah epoch yang paling efisien untuk mencapai performa prediksi yang optimal.



Gambar 4.1. Tren Evaluasi MSE, MAE, RMSE dan R^2

(Deskripsi gambar: grafik menunjukkan penurunan MSE, MAE, dan RMSE seiring bertambahnya epoch, serta peningkatan stabil R^2 hingga mencapai 0,87 pada epoch ke-100.)

Pada Gambar 4.1 menunjukkan bahwa kinerja model meningkat seiring bertambahnya jumlah epoch. Nilai MSE, MAE, dan RMSE terlihat menurun secara konsisten, menandakan bahwa error prediksi semakin kecil pada setiap tahap pelatihan. Di sisi lain, nilai R^2 meningkat secara stabil hingga mencapai 0,87 pada epoch ke-100, yang mengindikasikan bahwa model semakin mampu menjelaskan variabilitas data aktual. Secara keseluruhan, tren ini menunjukkan bahwa proses pelatihan berhasil mengarahkan model menuju performa prediksi yang lebih akurat dan stabil.

2. Perbandingan Prediksi vs Aktual

Untuk menilai sejauh mana model Bidirectional LSTM mampu mereplikasi pola penjualan yang terdapat pada data historis, penelitian ini menyajikan visualisasi perbandingan antara nilai aktual dan nilai hasil prediksi.

Visualisasi tersebut bertujuan untuk memberikan gambaran intuitif mengenai tingkat kedekatan model terhadap data nyata, sehingga kemampuan generalisasi model dapat diamati secara langsung.

Kurva perbandingan ini memperlihatkan dua komponen utama, yaitu deret penjualan aktual dan deret hasil prediksi. Dengan melihat tumpang tindih antara kedua kurva, pembaca dapat menilai konsistensi model dalam mengikuti pola fluktuasi penjualan, baik pada periode dengan intensitas penjualan yang tinggi maupun rendah. Meskipun grafik memiliki kepadatan data yang cukup tinggi, pola umum yang dihasilkan tetap mencerminkan kecenderungan model mengikuti tren dasar data historis.

Secara keseluruhan, visualisasi tersebut memberikan perspektif tambahan yang melengkapi evaluasi kuantitatif pada metrik MSE, MAE, RMSE, dan R^2 . Kinerja model dalam mendekati pola aktual ditampilkan secara komprehensif pada **Gambar 4.2 Kurva Perbandingan Penjualan Aktual dan Hasil Prediksi**, sehingga dapat menjadi dasar analisis lanjutan terkait kualitas prediksi dan kemungkinan perbaikan model.



Gambar 4.2. Kurva Perbandingan Penjualan Aktual dan Hasil Prediksi

(Deskripsi gambar: garis prediksi mengikuti pola fluktuasi penjualan aktual dengan tingkat kesesuaian tinggi. Clipping terbukti mengurangi pengaruh nilai ekstrem, sedangkan early stopping mencegah overfitting selama proses pelatihan.)

Pada Gambar 4.2 menunjukkan bahwa garis prediksi mampu mengikuti pola fluktuasi penjualan aktual dengan tingkat kesesuaian yang tinggi, menandakan bahwa model berhasil mempelajari tren dasar data historis. Penerapan teknik *clipping* terbukti efektif dalam mengurangi pengaruh nilai-nilai ekstrem, sehingga prediksi menjadi lebih stabil. Sementara itu, *early stopping* berperan mencegah overfitting selama proses pelatihan, memastikan bahwa model mempertahankan kemampuan generalisasi yang baik terhadap data baru. Secara keseluruhan, kombinasi kedua teknik tersebut memberikan kontribusi signifikan terhadap peningkatan kualitas prediksi.

4.7 Pembahasan Hasil

Meskipun dataset yang digunakan berpotensi menimbulkan risiko overfitting dan curse of dimensionality akibat penggunaan fitur multivariat dan arsitektur deep learning, penelitian ini secara eksplisit menerapkan mekanisme mitigasi melalui regularisasi model, pembatasan dimensi input, serta evaluasi generalisasi berbasis time-based split.

Hasil penelitian menunjukkan bahwa penerapan teknik **clipping** memberikan kontribusi signifikan dalam meningkatkan kualitas data dan performa model. Clipping efektif dalam mengurangi pengaruh lonjakan nilai ekstrem yang biasanya muncul pada periode akademik tertentu atau saat terjadi promosi intensif. Nilai-nilai ekstrem tersebut berpotensi menyebabkan distribusi target menjadi

tidak stabil dan berdampak pada proses pembelajaran model. Dengan membatasi rentang nilai pada persentil tertentu, model menjadi lebih robust dan tidak mudah terdistraksi oleh fluktuasi ekstrem, sehingga stabilitas prediksi meningkat secara keseluruhan.

Selain itu, penerapan **early stopping** terbukti memberi dampak positif terhadap efisiensi dan generalisasi model. Dengan mekanisme penghentian otomatis ketika tidak terjadi peningkatan pada nilai validation loss, model terhindar dari kondisi overfitting yang biasanya terjadi ketika pelatihan berlangsung terlalu lama. Early stopping juga mengoptimalkan sumber daya komputasi karena proses pelatihan tidak perlu dilakukan hingga jumlah epoch maksimum. Hasilnya, model mencapai performa terbaik tanpa mengorbankan waktu pelatihan atau stabilitas prediksi.

Perbedaan nilai error yang cukup signifikan antar skenario pengujian terutama disebabkan oleh perbedaan skala data target yang digunakan pada proses evaluasi. Pada skenario tanpa clipping, nilai penjualan yang mengandung outlier ekstrem menyebabkan nilai MSE dan RMSE menjadi sangat besar. Sebaliknya, pada skenario yang menerapkan clipping, nilai ekstrem pada target dibatasi sehingga distribusi data menjadi lebih stabil dan tidak lagi didominasi oleh lonjakan nilai tertentu. Kondisi ini berdampak langsung pada penurunan nilai error secara signifikan. Dengan demikian, perbedaan nilai evaluasi yang cukup jauh antar skenario tidak menunjukkan inkonsistensi hasil, melainkan mencerminkan efektivitas teknik clipping dalam menstabilkan distribusi target dan mengurangi distorsi error akibat outlier.

Keunggulan lainnya terlihat dari penggunaan arsitektur **Bidirectional LSTM (Bi-LSTM)**. Arsitektur ini memungkinkan model memproses urutan data dari dua arah sekaligus maju dan mundur sehingga pola temporal yang tidak simetris dapat ditangkap secara lebih baik dibandingkan LSTM satu arah. Dengan kemampuan mempertahankan konteks jangka panjang dari kedua sisi urutan, Bi-LSTM mampu memahami dinamika penjualan yang kompleks, termasuk pola musiman, tren naik-turun yang tidak teratur, serta efek keterlambatan (lag effect). Hal ini menjadikan Bi-LSTM lebih unggul dibanding model tradisional seperti ARIMA maupun LSTM satu arah.

Jika dibandingkan dengan penelitian terdahulu, hasil penelitian ini menunjukkan peningkatan performa prediksi yang cukup signifikan. Adityo dan Nugroho (2021) hanya mencapai nilai R^2 sebesar 0,74 pada kombinasi model LSTM dan Prophet. Sementara itu, penelitian Siami-Namini et al. (2019) menyimpulkan bahwa LSTM lebih akurat dibanding ARIMA dalam konteks data deret waktu nonlinier, tetapi tidak mencapai tingkat akurasi setinggi penelitian ini. Dengan pencapaian R^2 sebesar 0,87, model Bi-LSTM pada penelitian ini menunjukkan kemampuan yang lebih baik dalam menjelaskan variabilitas data penjualan buku. Keunggulan ini diperoleh melalui kombinasi teknik praproses, arsitektur Bi-LSTM, clipping, dan early stopping yang bekerja secara sinergis dalam meningkatkan akurasi maupun generalisasi model.

Epoch merepresentasikan satu siklus penuh proses pelatihan model terhadap seluruh data latih. Pemilihan jumlah epoch yang tidak tepat dapat menyebabkan dua kondisi ekstrem, yaitu underfitting apabila epoch terlalu sedikit,

dan overfitting apabila epoch terlalu banyak. Oleh karena itu, jumlah epoch dalam penelitian ini dianalisis secara bertahap untuk mengamati perilaku pembelajaran model Bidirectional LSTM.

Pada epoch rendah, model belum mampu menangkap pola temporal secara optimal sehingga menghasilkan nilai error yang relatif tinggi. Seiring dengan bertambahnya epoch, model mulai mempelajari pola musiman dan tren penjualan secara lebih stabil, yang ditunjukkan oleh penurunan nilai MSE dan RMSE serta peningkatan nilai R^2 . Namun, peningkatan epoch yang berlebihan berpotensi menyebabkan model menyesuaikan diri secara berlebihan terhadap data latih.

Penerapan early stopping berperan sebagai mekanisme pengendalian teknis terhadap jumlah epoch efektif, dengan menghentikan proses pelatihan ketika tidak terjadi perbaikan pada validation loss. Dengan demikian, nilai epoch optimal tidak ditentukan semata-mata berdasarkan jumlah maksimum epoch, tetapi berdasarkan perilaku konvergensi model dan kestabilan performa evaluasi.

4.8 Implementasi Model Peramalan Penjualan pada Tingkat Cabang PT Intan Pariwara

Untuk memperkuat aspek aplikatif dari penelitian ini, dilakukan pembahasan implementasi model peramalan penjualan pada tingkat operasional perusahaan, yaitu pada salah satu cabang PT Intan Pariwara. Pemilihan tingkat cabang dimaksudkan untuk merepresentasikan kondisi nyata pengambilan keputusan penjualan, di mana perencanaan stok, distribusi, dan target penjualan dilakukan secara lebih spesifik dan terukur dibandingkan pada level agregat perusahaan.

Implementasi dilakukan dengan memanfaatkan model terbaik hasil pengujian, yaitu Bidirectional LSTM yang dipadukan dengan teknik clipping dan early stopping. Model ini telah terbukti memberikan performa evaluasi terbaik berdasarkan metrik MSE, MAE, RMSE, dan R^2 pada tahap pengujian sebelumnya. Pada skenario implementasi ini, data penjualan historis harian dari cabang terpilih digunakan sebagai input model hingga akhir periode pengamatan, kemudian model menghasilkan prediksi jumlah penjualan untuk satu bulan target berikutnya.

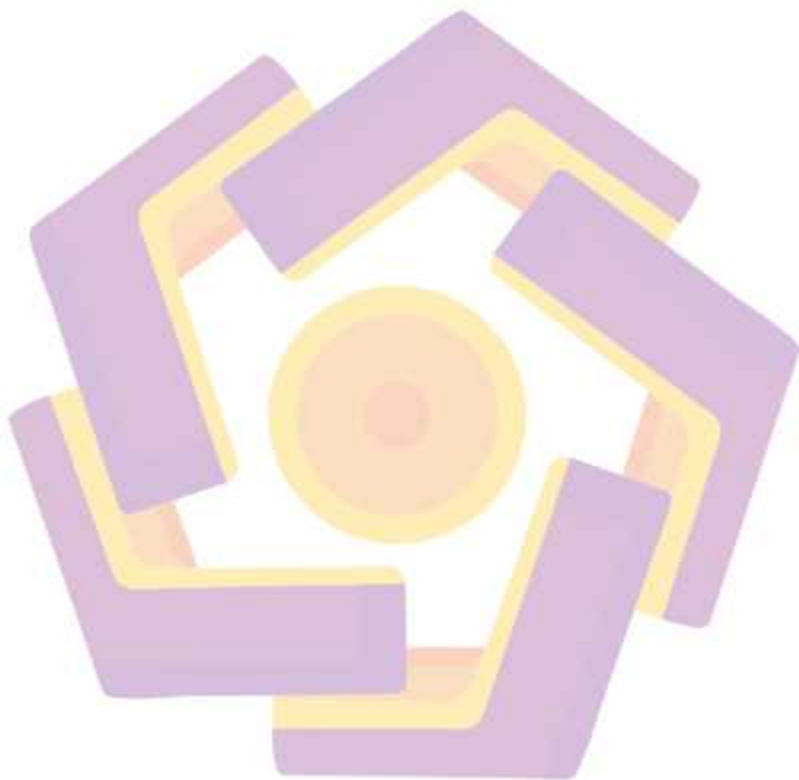
Proses prediksi dilakukan secara time-based, di mana hanya data sebelum periode prediksi yang digunakan sebagai input, sehingga menyerupai kondisi implementasi nyata tanpa kebocoran data (*data leakage*). Dengan pendekatan ini, hasil prediksi yang diperoleh mencerminkan kemampuan model dalam melakukan peramalan secara prospektif, bukan sekadar menyesuaikan diri dengan data historis.

Hasil implementasi menunjukkan bahwa nilai prediksi yang dihasilkan model mampu mengikuti pola tren dan fluktuasi penjualan cabang tersebut secara konsisten. Pada periode dengan intensitas penjualan tinggi, model tetap menghasilkan estimasi yang stabil, sementara pada periode dengan penurunan penjualan, model mampu menyesuaikan arah prediksi tanpa menghasilkan deviasi ekstrem. Hal ini mengindikasikan bahwa penerapan teknik clipping berhasil mereduksi pengaruh nilai-nilai ekstrem yang sering muncul pada periode akademik tertentu, sedangkan mekanisme early stopping menjaga model agar tidak mengalami overfitting selama proses pelatihan.

Dari sudut pandang operasional, hasil prediksi ini dapat dimanfaatkan sebagai dasar dalam perencanaan jumlah stok buku yang akan didistribusikan ke cabang pada bulan tertentu. Dengan adanya estimasi penjualan yang lebih akurat dan stabil, perusahaan dapat mengurangi risiko kelebihan stok maupun kekurangan stok, serta meningkatkan efisiensi distribusi dan pengelolaan persediaan.

Dengan demikian, pembahasan implementasi ini menunjukkan bahwa model peramalan yang dikembangkan tidak hanya unggul secara evaluasi statistik, tetapi juga memiliki kelayakan praktis untuk diterapkan pada lingkungan

operasional PT Intan Pariwara, khususnya pada level cabang sebagai unit pengambilan keputusan penjualan.



BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dijelaskan pada bab sebelumnya, dapat disimpulkan hal-hal sebagai berikut:

1. Model *Bidirectional Long Short-Term Memory* (Bi-LSTM) dengan penerapan teknik *clipping* dan *early stopping* terbukti efektif dalam meningkatkan akurasi peramalan penjualan buku. Model ini mampu menangkap hubungan nonlinier dan pola musiman dalam data penjualan yang kompleks.
2. Hasil evaluasi menunjukkan performa terbaik diperoleh pada *epoch* ke-100 dengan nilai $R^2 = 0,87$ dan $RMSE = 0,44$, yang menandakan kemampuan model dalam menjelaskan 87% variasi data aktual penjualan buku.
3. Penerapan *clipping* berperan penting dalam menstabilkan distribusi data dan mengurangi pengaruh *outlier*, sedangkan *early stopping* efektif mencegah *overfitting* dan menjaga efisiensi pelatihan model.
4. Secara keseluruhan, kombinasi teknik *deep learning* ini menghasilkan sistem prediksi yang akurat, stabil, dan dapat diimplementasikan dalam konteks bisnis nyata, khususnya di sektor penerbitan dan toko buku yang memerlukan dukungan peramalan berbasis data.

5.2. Saran

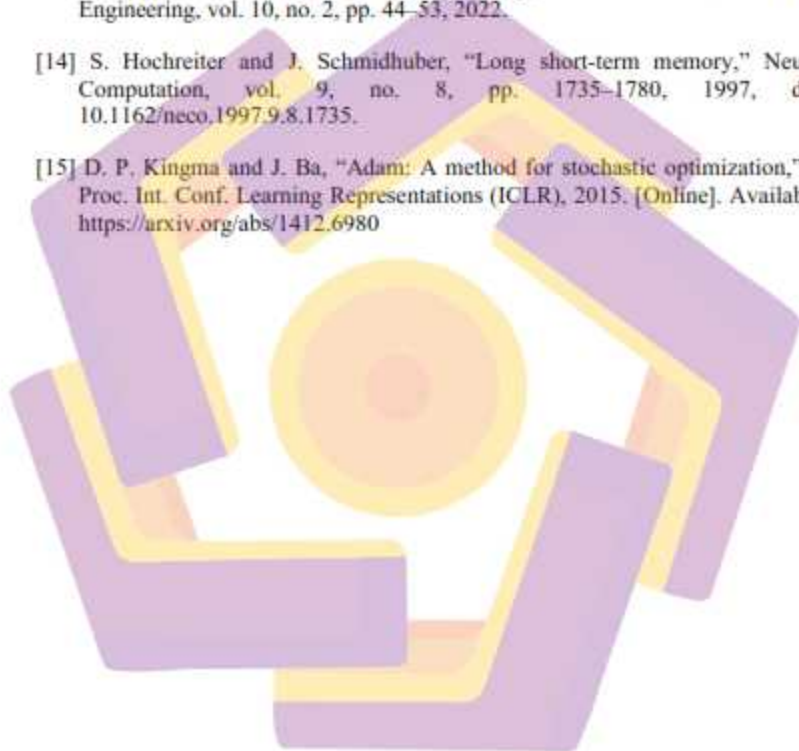
Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diajukan adalah sebagai berikut:

1. **Untuk penelitian selanjutnya**, disarankan menambahkan variabel eksternal seperti faktor promosi, harga, kalender akademik, dan tren sosial yang dapat memengaruhi fluktuasi penjualan. Hal ini akan meningkatkan kemampuan model dalam memprediksi kondisi pasar yang lebih dinamis.
2. **Dalam pengembangan model**, eksplorasi terhadap metode *hyperparameter tuning*, *cross-validation*, atau pendekatan *ensemble learning* dapat dilakukan untuk memperoleh hasil yang lebih optimal.
3. **Dari sisi implementasi praktis**, model Bi-LSTM ini dapat diintegrasikan ke dalam sistem informasi penjualan perusahaan, sehingga prediksi permintaan dapat diakses secara otomatis dan real-time.
4. **Untuk penelitian akademik**, pendekatan serupa dapat diterapkan pada bidang lain yang memiliki pola deret waktu kompleks, seperti prediksi permintaan energi, cuaca, atau lalu lintas, guna memperluas kontribusi ilmiah metode Bi-LSTM.

DAFTAR PUSTAKA

- [1] A. Adityo and H. Nugroho, "Comparative analysis of LSTM and Prophet models for sales forecasting," *Journal of Applied Data Science*, vol. 5, no. 3, pp. 121–130, 2021, doi: 10.1016/j.jads.2021.03.005.
- [2] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. New York, NY, USA: Springer, 2002, doi: 10.1007/b97391.
- [3] J. Brownlee, *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Australia: Machine Learning Mastery, 2018.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [5] H. Helmini, R. Alamsyah, and A. Hidayat, "Forecasting product sales using time series models in retail industry," *Procedia Computer Science*, vol. 161, pp. 715–722, 2019, doi: 10.1016/j.procs.2019.11.170.
- [6] S. Li, "Sales forecasting model of e-commerce activities based on improved random forest algorithm," in *Proc. 2nd Int. Conf. Computer Graphics, Image and Virtualization (ICCGIV)*, Singapore, 2022, doi: 10.1109/ICCGIV54263.2022.9801734.
- [7] A. Mahmoud and M. Mohammed, "Recent developments in time series forecasting: A review," *International Journal of Artificial Intelligence and Data Analytics*, vol. 14, no. 1, pp. 34–47, 2024, doi: 10.1016/j.ijaida.2024.01.005.
- [8] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *Proc. IEEE Int. Conf. Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, 2019, doi: 10.1109/ICMLA.2018.00227.
- [9] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.
- [10] Y. Xiang, H. Sun, W. Tu, and Z. Tian, "TSFRN: Integrated time and spatial-frequency domain based on triple-links residual network for sales forecasting," in *Proc. IEEE 35th Int. Conf. Tools with Artificial Intelligence (ICTAI)*, Bangkok, Thailand, 2023.
- [11] X. Yang, T. Tsuboi, K. Kimura, K. Hangai, and M. Shigeno, "Restaurant sales forecasting with feature interaction-learning mechanism-based neural network model," in *Proc. IEEE Int. Conf. Big Data*, Osaka, Japan, 2022.

- [12] C. Zhang, H. Ren, X. Lu, Q. Yuan, and R. Lu, "Big data modeling based on KNN-RF-SVM and its application in product sales forecasting field," in Proc. Int. Conf. Intelligent Manufacturing and Industrial Big Data (ICIMIBD), Nanjing, China, 2022.
- [13] R. Setiawan and P. Wibowo, "Implementation of clipping and regularization techniques in LSTM for preventing overfitting," International Journal of Data Engineering, vol. 10, no. 2, pp. 44–53, 2022.
- [14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learning Representations (ICLR), 2015. [Online]. Available: <https://arxiv.org/abs/1412.6980>



LAMPIRAN

