

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN
MULTIPLE LINEAR REGRESSION DALAM MEMPREDIKSI
PENGELUARAN OBAT HARIAN STUDI KASUS PADA
APOTEK BAMBUAN**

SKRIPSI

DISUSUN OLEH

MUAMMAR FARHAN

NPM. 2109020097



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN
MULTIPLE LINEAR REGRESSION DALAM MEMPREDIKSI
PENGELUARAN OBAT HARIAN STUDI KASUS PADA
APOTEK BAMBUAN**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Komputer (S.Kom) dalam Program Studi Teknologi Informasi pada
Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas
Muhammadiyah Sumatera Utara**

MUAMMAR FARHAN

NPM. 2109020097

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

LEMBAR PENGESAHAN

Judul Skripsi : PERBANDINGAN ALGORITMA RANDOM FOREST
DAN MULTIPLE LINEAR REGRESSION DALAM
MEMPREDIKSI PENGELUARAN OBAT HARIAN
STUDI KASUS PADA APOTEK BAMBUAN
Nama Mahasiswa : MUAMMAR FARHAN
NPM : 2109020097
Program Studi : TEKNOLOGI INFORMASI

Menyetujui
Komisi Pembimbing



(Zuli Agustina Gultom, M.SI)
NIDN. 0130089003

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom., M.Kom.)
NIDN. 0117019301

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN MULTIPLE
LINEAR REGRESSION DALAM MEMPREDIKSI PENGELUARAN
OBAT HARIAN STUDI KASUS PADA APOTEK BAMBUAN**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 22 Agustus 2025

Yang membuat pernyataan



Muammar Farhan

NPM. 2109020097

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Muammar Farhan
NPM	: 2109020097
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

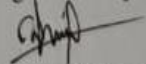
**PERBANDINGAN ALGORITMA RANDOM FOREST DAN MULTIPLE
LINEAR REGRESSION DALAM MEMPREDIKSI PENGELUARAN
OBAT HARIAN STUDI KASUS PADA APOTEK BAMBUAN**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 22 Agustus 2025

Yang membuat pernyataan



Muammar Farhan

NPM. 2109020097

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : MUAMMAR FARHAN
Tempat dan Tanggal Lahir : STABAT, 31 DESEMBER 2003
Alamat Rumah : Jl. Perdamaian Lingk VI Bambuan, Stabat
Telepon/Faks/HP : +62-812-7057-3358
E-mail : muammarfarhan09@gmail.com
Instansi Tempat Kerja : UNIVERSITAS MUHAMMDIYAH
SUMATERA UTARA
Alamat Kantor : Jl. Kapten Muchtar Basri No.3, Glugur
Darat II, Kec. Medan Tim., Kota Medan,
Sumatera Utara 20238

DATA PENDIDIKAN

SD	: SDN 054904 BAMBUAN	TAMAT: 2015
SMP	: SMPN 1 STABAT	TAMAT: 2018
SMA	: SMAN 1 STABAT	TAMAT: 2021

KATA PENGANTAR



Assalamu'alaikum Warahmatullahi Wabarakaatuh.

Alhamdulillah, puji dan Syukur penulis panjatkan kepada Allah SWT, yang telah memberikan Rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas akhir yaitu penulisan skripsi ini. Tak lupa shalawat serta beriringkan salam senangtiawsa penulis haturkan kepada Nabi Muhammad SAW. Yang telah membawa kita semua dari zaman jahiliyah menuju zaman yang terang benderang.

Skripsi ini merupakan salah satu syarat wajib yang harus dipenuhi untuk memperoleh gelar sarjana pada jenjang Strata Satu (S1) di Jurusan Teknologi, Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara. Pelaksanaan tugas ini merupakan salah satu kuliah wajib dari Jurusan Teknologi, Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara dan menjadi sarana penulis untuk menambah ilmu pengetahuan serta pengalaman dalam menerapkan keilmuan sesuai bidang yang sebelumnya sudah didapatkan pada proses perkuliahan.

Adapun pembahasan terkait skripsi ini adalah Perbandingan Algoritma Random Forest dan Multiple Linear Regression Dalam Memprediksi Pengeluaran Obat Harian Studi Kasus Pada Apotek Bambuan. Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom. Ketua Program Studi Teknologi Informasi.
4. Ibu Zuli Agustina Gultom, M.SI. Pembimbing Skripsi yang telah membantu dalam membimbing saya agar dapat menyelesaikan tugas akhir ini.
5. Mama dan Papa saya yang telah membesarkan saya dengan penuh kasih. Terima kasih atas doa, kesabaran, dan dukungan yang tiada henti. Teristimewa untuk Papa tercinta yang telah berpulang satu tahun yang lalu, semoga Papa tenang di sisi Tuhan. Skripsi ini saya persembahkan sebagai wujud cinta dan terima kasih yang tak terhingga.
6. Kepada pacar penulis yang sudah menemani penulis dari awal perkuliahan hingga akhir perjalanan kuliah, memberikan semangat dan motivasi agar terus maju, Saudari Dhea Fahma, semoga setelah ini diberikan rezeki menikah dikemudian hari.
7. Kepada teman-teman seperjuangan penulis yang berada di dalam group "Punyaku-Punyaku" yaitu Rifqi, Gatot, Yoga, Qodri, Anggi, Ikhsan dan Mutiara Akbar yang telah bersama sama selama perkuliahan baik senang maupun susah serta menjadi teman diskusi sekaligus motivasi selama menempuh pendidikan di UMSU.
8. Teman-teman kelas TI C1 Stambuk 2021 terutama buat saudara Rian, Setyo, Bahdi, Fina, Meysa, Daffa, Mirza dan teman-teman lainnya yang

tidak dapat saya sebutkan satu persatu yang sudah membantu dan memberikan dukungan kepada penulis sedari awal perkuliahan hingga akhir perjalanan kuliah

9. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

PERBANDINGAN ALGORITMA RANDOM FOREST DAN MULTIPLE LINEAR REGRESSION DALAM MEMPREDIKSI PENGELUARAN OBAT HARIAN STUDI KASUS PADA APOTEK BAMBUAN

ABSTRAK

Pengelolaan stok obat yang akurat sangat penting bagi apotek untuk mencegah kekurangan maupun kelebihan persediaan. Penelitian ini bertujuan membandingkan kinerja *Multiple Linear Regression* (MLR) dan *Random Forest Regression* (RFR) dalam memprediksi penjualan obat harian pada Apotek Bambuan. Data yang digunakan adalah catatan penjualan periode 2022–2024, yang telah melalui tahap *preprocessing* dan pembagian data latih serta uji. Kedua model dievaluasi menggunakan metrik *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*. Hasil penelitian menunjukkan bahwa Random Forest menghasilkan prediksi lebih akurat dengan nilai kesalahan lebih rendah dibandingkan MLR, meskipun begitu MLR tetap bermanfaat untuk interpretasi kontribusi variabel. Dengan demikian, Random Forest direkomendasikan untuk prediksi penjualan obat harian karena tingkat akurasinya yang lebih tinggi, sedangkan MLR unggul dari sisi transparansi model.

Kata Kunci: Machine Learning, Multiple Linear Regression, Random Forest, Prediksi Penjualan Obat, Apotek.

COMPARISON OF RANDOM FOREST AND MULTIPLE LINEAR REGRESSION ALGORITHM IN PREDICTING DAILY DRUG EXPENDITURES: A CASE STUDY AT APOTEK BAMBUAN

ABSTRACT

Accurate drug inventory management is essential for pharmacies to avoid shortages or excess stock. This study aims to compare the performance of *Multiple Linear Regression* (MLR) and *Random Forest Regression* (RFR) in predicting daily drug sales at Apotek Bambuan. The dataset consists of sales records from 2022–2024, which were preprocessed and divided into training and testing sets. Both models were evaluated using *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), and *R² Score*. The results show that Random Forest provides higher prediction accuracy with lower error values compared to MLR, although MLR remains useful for interpreting the contribution of predictor variables. Therefore, Random Forest is recommended for daily drug sales prediction due to its superior accuracy, while MLR offers advantages in model interpretability.

Keywords: *Machine Learning, Random Forest, Multiple Linear Regression, Drug sales prediction, Pharmacy.*

DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS	iiError! Bookmark not defined.
ERNYATAAN PERSETUJUAN PUBLIKASI	iiiError! Bookmark not defined.
RIWAYAT HIDUP	iiiv
KATA PENGANTAR	v
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	6
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	7
1.5. Manfaat Penelitian.....	8
BAB II LANDASAN TEORI	9
2.1. Apotek	9
2.2. Machine Learning.....	10
2.2.1. Konsep Machine Learning	10
2.2.2. Penerapan Machine Learning dalam Dunia Kesehatan dan Apotek	12
2.3. Multiple Linear Regression	14
2.4. Random Forest.....	15
2.5. Metrik Evaluasi Model	16
2.5.1. Root Mean Squared Error (RMSE).....	16
2.5.2. R ² Score (Koefisien Determinasi)	17
2.5.3. Mean Absolute Error (MAE)	18
2.5.4. Mean Absolute Percentage Error (MAPE).....	19
2.5.5. Mean Squared Error (MSE)	19
2.6. Studi Literatur.....	20
BAB III METODOLOGI PENELITIAN	24
3.1. Jenis Penelitian	24
3.2. Teknik Pengumpulan Data	25
3.3. Alat dan Perangkat Penelitian.....	26
3.4. Diagram Alur Penelitian atau Tahapan Penelitian.....	28
3.5. Teknik Analisis Data	30

3.5.1.	Preprocessing Data.....	30
3.5.2.	Pelatihan Model	31
3.5.3.	Evaluasi Kinerja Model.....	31
3.5.4.	Perbandingan Hasil dan Visualisasi	32
3.6.	Kerangka Waktu Penelitian	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Deskripsi Dataset	34
4.2	Hasil Preprocessing Data	34
4.3	Implementasi Model Random Forest dan MLR.....	39
4.3.1	Arsitektur Model	39
4.3.2.	Hasil Konfigurasi Awal Model.....	43
4.4.	Analisis Pengaruh Jumlah Data Historis.....	50
4.4.1	Data Historis 1 Tahun	50
4.4.2	Data Historis 2 Tahun	51
4.5.	Pembahasan Hasil Penelitian	52
BAB V PENUTUP		55
5.1	Kesimpulan	55
5.2	Saran	56
DAFTAR PUSTAKA		58
LAMPIRAN.....		60

DAFTAR TABEL

Tabel 2.1 Studi Literatur	20
Tabel 3.1 Contoh Dataset Apotek Bambuan.....	26
Tabel 3.2 Alat dan Perangkat Penelitian.....	26
Tabel 3.3 Kerangka Waktu Penelitian	33
Tabel 4.1 Pseudocode Preprocessing	35
Tabel 4.2 Pseudocode 2 Data Split	36
Tabel 4.3 Pseudocode 3 Model.....	40
Tabel 4.4 Pseudocode 4 Model Random Forest	41
Tabel 4.5 Perbandingan Data 1 Tahun Terhadap Kedua Model	50
Tabel 4.6 Perbandingan Data 2 Tahun Terhadap Kedua Model	51

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	28
Gambar 4.1 Prediksi vs Aktual Random Forest.....	44
Gambar 4.2 Perbandingan matriks evaluasi MAE.....	45
Gambar 4.3 Perbandingan matriks evaluasi MSE	46
Gambar 4.4 Perbandingan matriks evaluasi RMSE.....	47
Gambar 4.5 Perbandingan matriks evaluasi MAPE.....	48
Gambar 4.6 Perbandingan matriks evaluasi.....	49
Gambar 4.7 Perbandingan Keseluruhan Matriks Data 1 Tahun	53
Gambar 4.8 Perbandingan Keseluruhan Matriks Data 2 Tahun	53

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Apotek merupakan fasilitas pelayanan kesehatan yang menyelenggarakan kegiatan pelayanan kefarmasian, meliputi pengadaan, penyimpanan, peracikan, peresepan, dan distribusi obat kepada masyarakat. Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 73 Tahun 2016, apotek dipandang sebagai fasilitas tempat apoteker melaksanakan praktik kefarmasian sesuai ketentuan hukum yang berlaku. (Kementerian Kesehatan RI, 2016).

Apotek Bambuan merupakan salah satu fasilitas kefarmasian yang beroperasi di wilayah Kabupaten Langkat. Tiap harinya apotek bambuan menyalurkan puluhan obat kepada orang yang membutuhkan, Proses pencatatan transaksi penjualan, pengeluaran, maupun pemasukan obat masih dilakukan dengan sistem manual melalui pencatatan di buku.. Metode tersebut berpotensi menyebabkan kekeliruan pencatatan, memperlambat pemrosesan data, serta menyulitkan prediksi kebutuhan stok obat konvensional di waktu berikutnya. Kondisi tersebut dapat menghambat efektivitas pelayanan serta meningkatkan risiko kekurangan atau kelebihan persediaan obat.

Untuk itu, diperlukan pendekatan yang mampu mengubah data penjualan harian menjadi informasi prediktif yang bermanfaat. Salah satu solusi yang berkembang saat ini adalah pemanfaatan teknologi analisis berbasis data, yang memungkinkan pola penjualan masa lalu digunakan untuk memperkirakan kebutuhan di masa depan. Dengan demikian, pengelolaan stok tidak lagi hanya

mengandalkan pencatatan manual, tetapi sudah ditunjang oleh metode analisis yang lebih akurat dan sistematis.

Dalam era digital saat ini, penggunaan teknologi dan data dalam pengambilan keputusan bisnis menjadi hal yang sangat penting. Data historis penjualan dapat dimanfaatkan untuk memperkirakan kebutuhan stok obat di masa depan secara lebih akurat, sehingga apotek dapat memaksimalkan operasional, meningkatkan efisiensi, mengurangi risiko, serta menjaga kepuasan pelanggan. Salah satu pendekatan yang relevan untuk menjawab tantangan ini adalah penerapan metode *machine learning*, khususnya algoritma regresi yang mampu memprediksi jumlah penjualan obat harian berdasarkan pola data historis.

Penelitian ini secara khusus membandingkan dua model *machine learning*, yaitu *Multiple Linear Regression (MLR)* sebagai model linier, dan *Random Forest Regression (RFR)* sebagai model non-linier. MLR dipilih karena sifatnya sederhana, mudah diinterpretasikan, dan mampu menjelaskan kontribusi setiap variabel terhadap hasil prediksi. Sementara itu, Random Forest dipilih karena kemampuannya dalam menangkap pola non-linier dan mengurangi risiko *overfitting*. Dengan membandingkan kedua model ini, diharapkan dapat ditemukan model yang paling akurat dalam memprediksi jumlah penjualan obat harian di Apotek Bambuan.

Penelitian mengenai prediksi penjualan dan manajemen persediaan barang, khususnya di sektor farmasi seperti apotek, telah banyak dilakukan oleh berbagai peneliti sebelumnya. Studi-studi ini memberikan landasan penting dalam pengembangan model prediksi penjualan obat harian yang diangkat dalam penelitian ini. Beberapa di antaranya menggunakan berbagai pendekatan seperti

regresi linier, *K-Nearest Neighbors*, *Random Forest* dan berbagai pendekatan lainnya. Perbedaan dari studi-studi tersebut dapat terlihat dari jenis data yang digunakan, algoritma yang diterapkan, dan tingkat akurasi prediksi yang dicapai.

Supriyanto, Ilhamsyah, dan Enri (2020) melakukan penelitian berjudul *Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression dan Random Forest*. Meskipun objek penelitian bukan dalam bidang farmasi, perbandingan dua algoritma yang hampir sama yaitu *Linear Regression* dan *Random Forest* sangat relevan dengan penelitian ini. Di dalam penelitian tersebut juga dijelaskan tentang hasil dari penelitian tersebut dimana penulis menyatakan Penelitian tersebut menunjukkan bahwa algoritma *Linear Regression* memberikan performa yang lebih baik dibandingkan *Random Forest* dari segi perbandingan nilai matriks evaluasi RMSE.

Penelitian lain dilakukan oleh Efendi dan Zyen (2024) yang berjudul *Penerapan Algoritma Random Forest untuk Prediksi Penjualan dan Sistem Persediaan Produk*. Meskipun fokus penelitian ini bukan pada apotek, hasilnya tetap relevan karena mengkaji efektivitas *Random Forest* dalam sistem *Point of Sale* (POS) ritel. Hasil penelitian tersebut menunjukkan bahwa penerapan algoritma *Random Forest* mampu meningkatkan efisiensi pengelolaan bisnis dengan tingkat akurasi prediksi mencapai 85%. Sistem ini berhasil merangkum penjualan mingguan dengan hasil yang cukup akurat, sehingga dapat membantu mencegah terjadinya kelebihan maupun kekurangan stok yang berpotensi menimbulkan kerugian dan menurunkan tingkat kepuasan pelanggan..

Sementara itu, Pratiwi (2023) dalam penelitiannya yang berjudul *Prediksi Persediaan Obat pada Apotek Menggunakan Algoritma Decision Tree*

menggunakan algoritma *Decision Tree* sebagai pembanding. Penelitian ini difokuskan pada prediksi stok obat generik di sebuah apotek dan memperoleh hasil akhir yaitu upaya memprediksi ketersediaan obat generik di salah satu apotek. Hasil pengujian menunjukkan bahwa dengan pembagian data pelatihan dan pengujian sebesar 80:20, model *Decision Tree* mampu mencapai tingkat akurasi hingga 98,71%. Nilai metrik lainnya seperti Precision, Recall, dan F1-Score juga menunjukkan konsistensi, masing-masing sebesar 0,9872, 0,9872, dan 0,9867. Ketika dilakukan pengujian dengan rasio data 70:30, akurasi model sedikit menurun menjadi 98,28%, dengan Precision sebesar 0,9832, Recall 0,9828, dan F1-Score 0,9804, namun perbedaannya relatif kecil dibandingkan skenario sebelumnya.. Walaupun tidak menggunakan *Random Forest*, hasil penelitian ini tetap memberikan kontribusi penting sebagai pembanding terhadap algoritma regresi pohon lainnya dalam konteks prediksi stok obat.

Penelitian yang dilakukan oleh Pratiwi et al. (2023) berjudul “*Prediksi Persediaan Obat pada Apotek Mitha Farma Menggunakan Algoritma Random Forest*” merupakan salah satu studi penting yang menginspirasi penelitian ini. Penelitian tersebut menerapkan *Random Forest* untuk mengklasifikasikan jenis obat yang cepat habis di Apotek Mitha Farma, Medan. Hasilnya menunjukkan bahwa algoritma ini mampu memetakan tren pembelian dengan akurat, di mana obat Neurobion putih menjadi yang paling sering habis terjual hingga 60%. Pendekatan *Random Forest* memberikan solusi dalam menentukan strategi pengisian stok yang lebih responsif dan efisien dibanding metode manual.

Studi lain yang juga relevan adalah yang dilakukan oleh Rusdy, Purnawansyah, dan Herman (2022), berjudul “*Penerapan Metode Regresi Linier*

pada Prediksi Permintaan dan Penawaran Obat: Studi Kasus Aplikasi Point of Sales". Penelitian ini menggunakan data historis transaksi penjualan obat dan menerapkan metode Multiple Linear Regression (MLR). Hasil evaluasi model menunjukkan bahwa regresi linier memberikan hasil prediksi yang cukup akurat dengan nilai *Mean Absolute Percentage Error* (MAPE) sebesar 4,91%, *Mean Squared Error* (MSE) sebesar 1,07, dan *Root Mean Squared Error* (RMSE) sebesar 1,03. Walaupun tidak menggunakan Random Forest, hasil penelitian ini tetap memberikan kontribusi penting sebagai pembanding terhadap algoritma regresi pohon lainnya dalam konteks prediksi stok obat.

Urgensi penelitian ini terletak pada pentingnya pemanfaatan machine learning (ML) untuk mendukung pengelolaan stok obat harian di apotek. *Machine learning* memiliki kemampuan untuk mempelajari pola dari data historis penjualan dan menghasilkan prediksi yang lebih akurat serta dapat menyesuaikan diri dengan fluktuasi permintaan. Dengan pendekatan ini, proses pengambilan keputusan tidak lagi hanya mengandalkan intuisi atau pencatatan manual, melainkan berdasarkan analisis data yang sistematis dan terukur.

Dalam konteks Apotek Bambuan, akurasi prediksi penjualan obat sangat penting untuk menghindari dua risiko utama, yaitu kekurangan stok yang dapat mengganggu pelayanan pasien, serta kelebihan stok yang berpotensi menimbulkan kerugian akibat obat kadaluwarsa. Oleh karena itu, penelitian ini menempatkan penerapan *machine learning* sebagai langkah strategis untuk meningkatkan efisiensi, efektivitas, dan ketepatan dalam perencanaan persediaan obat.

Selain manfaat praktis, penelitian ini juga memiliki nilai akademis, yaitu dengan membandingkan dua algoritma *machine learning* yang berbeda pendekatan:

Multiple Linear Regression (MLR) sebagai model linier yang sederhana dan mudah dipahami, serta Random Forest Regression (RFR) sebagai model non-linier yang lebih kompleks dan mampu menangkap pola data yang tidak linier. Evaluasi kedua model dilakukan menggunakan metrik *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*, sehingga hasil penelitian dapat memberikan gambaran menyeluruh mengenai keunggulan masing-masing algoritma.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi nyata bagi Apotek Bambuan dan apotek lokal lainnya dalam meningkatkan ketepatan perencanaan stok obat, sekaligus menjadi referensi akademik mengenai perbandingan kinerja model linier dan non-linier dalam konteks prediksi penjualan obat harian.

1.2. Rumusan Masalah

Permasalahan utama dalam penelitian ini berdasarkan latar belakang yang dirincikan sebelumnya adalah bagaimana membandingkan kinerja model *Multiple Linear Regression* (MLR) yang merupakan model linier, *Random Forest* sebagai model non linier dalam memprediksi total jumlah penjualan obat harian menggunakan data historis penjualan periode data tahun 2022-2024. Fokusnya terletak pada efektivitas masing-masing model.

1.3. Batasan Masalah

Adapun batasan-batasan yang diberlakukan dalam penelitian ini agar terfokus dan menghasilkan analisis yang jelas. Batasan tersebut adalah sebagai berikut:

1. Penelitian hanya berfokus pada Apotek Bambuan yang berlokasi di Jalan Perdamaian, Kecamatan Stabat, Kabupaten Langkat.

2. Penelitian ini hanya membandingkan performa model *Multiple Linear Regression* (MLR) dan *Random Forest*. Evaluasi performa dilakukan menggunakan metrik *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*,
3. Data yang digunakan dalam penelitian ini adalah data penjualan obat-obatan yang telah tersedia di Apotek Bambuan dalam jangka waktu tertentu yang telah ditentukan dalam studi kasus ini periode tahun 2022 sampai dengan 2024, sehingga hasil terbatas pada rentang waktu tersebut.

1.4. Tujuan Penelitian

1. Mengevaluasi akurasi model *Multiple Linear Regression* (MLR) dan *Random Forest* dalam memprediksi total jumlah penjualan obat menggunakan metrik *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*,
2. Menganalisis pengaruh jumlah data historis terhadap akurasi prediksi kedua model.
3. Memberikan rekomendasi model *Machine Learning* yang paling akurat dan efisien untuk prediksi total jumlah penjualan obat harian bagi toko di sektor farmasi khususnya apotek di Indonesia, berdasarkan hasil akhir matriks evaluasi *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*.

1.5. Manfaat Penelitian

1. Bagi Peneliti

Penelitian ini memberikan pengalaman langsung dalam penerapan metode machine learning *Multiple Linear Regression* (MLR) dan *Random Forest* untuk prediksi data deret waktu, sekaligus memperluas pemahaman teknis terhadap implementasi praktis menggunakan *Python*, serta *libray* seperti *Scikit Learn*, *Matplotlib* dan lain sebagainya, Selain itu, peneliti memperoleh wawasan tentang proses pengolahan data historis penjualan.

2. Bagi Apotek

Hasil penelitian ini dapat dijadikan sebagai bahan pertimbangan dan bahan pengambilan keputusan mengenai perkiraan jumlah stok obat yang akan dibutuhkan di masa yang akan datang.

3. Bagi Pembaca

Penelitian ini diharapkan dapat digunakan sebagai salah satu referensi mengenai penggunaan *Multiple Linear Regression* dan *Random Forest* dalam studi kasus prediksi penjualan, serta diharapkan dapat menjadi salah satu studi perbandingan dimasa yang akan datang.

BAB II

LANDASAN TEORI

2.1. Apotek

Apotek adalah salah satu fasilitas pelayanan kesehatan yang bertugas melaksanakan kegiatan kefarmasian, meliputi proses pengadaan, penyimpanan, peracikan, pemberian resep, hingga penyaluran obat kepada masyarakat. Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 73 Tahun 2016, apotek diposisikan sebagai tempat bagi apoteker dalam melaksanakan praktik kefarmasian sesuai dengan ketentuan hukum yang berlaku (Kementerian Kesehatan RI, 2016).

Dalam praktiknya, apotek memiliki dua fungsi utama yaitu sebagai tempat penjualan obat fungsi komersial dan sebagai penyedia layanan informasi obat kepada masyarakat fungsi profesional. Oleh karena itu, pengelolaan apotek harus memperhatikan dua hal sekaligus yaitu pelayanan yang cepat dan akurat kepada pasien serta efisiensi operasional dalam pengelolaan barang, khususnya obat-obatan.

Permasalahan umum yang sering terjadi di apotek adalah manajemen persediaan yang belum optimal. Kekurangan stok dapat menyebabkan keterlambatan pengobatan, sementara kelebihan stok dapat meningkatkan risiko kadaluwarsa obat dan pemborosan biaya. Oleh karena itu, sistem manajemen apotek yang baik harus mampu memprediksi kebutuhan stok obat secara akurat.

Seiring dengan perkembangan teknologi informasi, berbagai inovasi telah diterapkan untuk mendukung pengelolaan apotek, seperti penggunaan sistem *Point of Sales* (POS), aplikasi resep digital, dan sistem manajemen inventori otomatis.

Salah satu pendekatan terkini yang mulai diterapkan dalam sektor ini adalah penggunaan teknologi prediktif berbasis data, seperti machine learning, untuk memperkirakan kebutuhan obat berdasarkan pola penjualan sebelumnya.

2.2. Machine Learning

2.2.1. Konsep Machine Learning

Menurut Wijoyo et al. (2024) *Machine Learning* (ML) merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer untuk belajar dan mengambil keputusan secara mandiri tanpa harus diprogram berulang kali oleh manusia. Dengan kata lain, komputer dapat meningkatkan kemampuannya seiring waktu melalui pengalaman yang diperoleh dari data. Secara prinsip, Machine Learning beroperasi dengan cara menganalisis data, mengenali pola, serta membuat keputusan atau prediksi dengan campur tangan manusia yang seminimal mungkin. Teknologi ini dirancang untuk membantu sistem dalam melakukan pembelajaran otomatis dari data-data historis yang tersedia. Machine Learning umumnya diklasifikasikan ke dalam tiga pendekatan utama:

1. Supervised Learning

Menurut Wijoyo et al. (2024) Supervised learning merupakan salah satu metode pembelajaran mesin yang paling umum digunakan. Dalam konteks bahasa Indonesia, istilah ini sering diartikan sebagai “pembelajaran dengan pengawasan,” karena model dibor menggunakan data yang telah memiliki label atau target yang diketahui. Algoritma supervised learning bekerja dengan mempelajari hubungan antara data input dan output berdasarkan contoh yang diberikan selama proses pelatihan. Pendekatan ini banyak digunakan untuk tugas-tugas seperti klasifikasi

dan regresi. Keunggulan utama dari metode ini adalah prosesnya yang relatif sederhana dan mudah dipahami. Selain itu, algoritma ini juga efektif dalam melakukan klasifikasi terhadap data. Namun, kelemahannya terletak pada kebutuhan akan jumlah data berlabel yang besar untuk memperoleh hasil prediksi yang akurat, serta waktu komputasi yang cukup lama dalam proses pelatihan. Karena model ini mengandalkan data berlabel, maka penerapannya lebih kompleks dibandingkan dengan unsupervised learning yang tidak memerlukan label pada setiap input data.

2. Unsupervised Learning

Menurut Wijoyo et al. (2024) unsupervised learning merupakan jenis algoritma pembelajaran mesin yang tidak memerlukan data berlabel. Pada metode ini, sistem tidak menggunakan label keluaran atau target selama proses pelatihan, melainkan menganalisis data secara mandiri untuk menemukan pola dan hubungan yang tersembunyi. Pendekatan ini sering diterapkan untuk tujuan seperti pengelompokan (clustering) dan pencarian hubungan antar data (association rule). Karena tidak menggunakan label, algoritma pembelajaran tanpa pengawasan cenderung lebih fleksibel dalam mengeksplorasi struktur data yang belum diketahui sebelumnya. Keunggulan utama metode ini adalah kemampuannya dalam menemukan pola-pola baru tanpa harus bergantung pada data yang telah dipecah. Namun kekurangannya terletak pada tingkat kesulitan dalam memancarkan hasil, karena tidak adanya label membuat perbandingan antara input dan output menjadi lebih kompleks. Setelah memahami konsep pembelajaran terbimbing dan tidak terbimbing, diketahui bahwa tidak semua algoritma dapat digolongkan secara

mutlak ke dalam dua kategori tersebut. Beberapa algoritma berada di antara keduanya, yang dikenal sebagai pembelajaran semi-supervisi.

3. Reinforcement Learning

Menurut Wijoyo et al. (2024) *Reinforcement learning* atau *semi-supervised learning* merupakan pendekatan mesin pembelajaran yang tidak sepenuhnya termasuk dalam kategori diawasi maupun tidak diawasi. Pada metode ini, sistem belajar melalui proses interaksi dengan lingkungannya untuk mencapai hasil terbaik berdasarkan umpan balik yang diterima. Pendekatan pembelajaran penguatan umumnya digunakan pada kasus di mana data terbagi menjadi dua bagian sebagian data memiliki label dilatih secara diawasi, sementara sebagian lainnya tidak dipelajari secara tanpa pengawasan. Model akan berusaha memaksimalkan hasil yang diperoleh melalui proses percobaan dan kesalahan (*trial and error*) dengan bantuan fungsi yang diberikan (*reward function*). Keunggulan metode ini adalah efisiensinya dalam mempelajari strategi optimal dengan jumlah data berlabel yang relatif sedikit. Selain itu, pembelajaran penguatan dapat menyesuaikan perilaku sistem berdasarkan pengalaman, sehingga cocok diterapkan pada bidang yang membutuhkan pengambilan keputusan berkelanjutan seperti robotika, sistem rekomendasi, dan otomatisasi proses bisnis.

2.2.2. Penerapan Machine Learning dalam Dunia Kesehatan dan Apotek

Penerapan *Machine learning* telah banyak diterapkan dalam berbagai bidang, termasuk dalam dunia kesehatan. Di rumah sakit, ML digunakan untuk prediksi penyakit (*diagnosis prediktif*), klasifikasi gambar medis (seperti rontgen atau MRI), manajemen jadwal pasien, hingga sistem rekomendasi terapi. Sedangkan dalam

industri farmasi, ML digunakan untuk penemuan obat baru, optimasi formulasi obat, hingga prediksi interaksi antarobat.

Di sektor apotek, *machine learning* semakin banyak dimanfaatkan untuk memprediksi kebutuhan stok obat, mendeteksi tren permintaan, serta mengoptimalkan rantai pasok. Beberapa penelitian lokal di Indonesia menunjukkan hasil yang signifikan. Penelitian oleh Rahma et al. (2024) menggunakan algoritma K-Nearest Neighbor untuk memprediksi penjualan obat di PT XYZ Palembang dan berhasil mencapai akurasi hingga 90%.

Penerapan machine learning di apotek memberikan keuntungan dalam hal efisiensi dan akurasi. Dengan memprediksi kebutuhan obat secara otomatis, apotek dapat menghindari kekosongan atau kelebihan stok, mengurangi biaya operasional, dan meningkatkan pelayanan kepada pelanggan. Hal ini sangat penting mengingat ketersediaan obat yang konsisten menjadi faktor kunci dalam menjamin keberlanjutan pelayanan kesehatan.

Dalam penelitian ini, penerapan machine learning dilakukan dengan membandingkan dua algoritma *supervised learning*, yaitu *Multiple Linear Regression* (MLR) dan *Random Forest Regression* (RFR). Masing-masing algoritma akan dievaluasi berdasarkan performanya dalam memprediksi jumlah penjualan obat harian di Apotek Bambuan, dengan menggunakan metrik evaluasi *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*. Multiple Linear Regression (MLR).

2.3. Multiple Linear Regression

Menurut Farhanuddin dkk. (2024), Regresi Linier Berganda (MLR) merupakan salah satu algoritma dalam pembelajaran mesin terawasi (*supervised learning*) yang berfungsi untuk memprediksi nilai numerik berdasarkan satu atau lebih variabel independen. Algoritma ini membangun model berbentuk linier yang menghubungkan variabel bebas (variabel bebas) dengan variabel terikat (variabel terikat). Model MLR dipilih karena bersifat sederhana, mudah digunakan, dan memiliki kemampuan yang baik dalam menggambarkan hubungan linier antar variabel. Keunggulan utamanya terletak pada kemampuannya memberikan interpretasi yang jelas mengenai pengaruh masing-masing variabel independen terhadap hasil prediksi..

MLR menjadi salah satu pendekatan yang umum dalam prediksi numerik karena sifatnya yang mudah diinterpretasikan dan mampu memberikan wawasan langsung mengenai kontribusi masing-masing variabel bebas terhadap hasil. Dalam konteks penelitian ini, MLR digunakan untuk memprediksi jumlah penjualan obat harian berdasarkan faktor-faktor seperti hari, akhir pekan, hari libur, dan rolling average penjualan hari sebelumnya. Pemodelan seperti ini penting dalam pengambilan keputusan untuk manajemen stok dan perencanaan distribusi.

Rumus MLR dapat ditulis sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Dimana:

Y = variabel dependen (misalnya, jumlah penjualan obat)

$X_1, X_2, \dots, X_n$ = variabel independen (misalnya, hari, stok sebelumnya, jenis obat)

β_0 = konstanta

$\beta_1, \beta_2, \dots, \beta_n$ = koefisien regresi untuk setiap variabel $X_1, X_2, \dots, X_n$

ε = error (residual) antara nilai aktual dan prediksi

Nilai koefisien β biasanya dihitung menggunakan metode *Ordinary Least Squares (OLS)*, yang bertujuan untuk meminimalkan nilai Sum of Squared Errors (SSE).

Tujuan dari regresi linear berganda adalah untuk menemukan nilai-nilai koefisien β yang meminimalkan jumlah kuadrat error (least squares). Proses estimasi biasanya menggunakan metode *Ordinary Least Squares (OLS)*, yaitu pendekatan untuk mencari solusi terbaik yang mengurangi jumlah kesalahan kuadrat total.

2.4. Random Forest

Random Forest merupakan salah satu algoritma dalam bidang machine learning yang termasuk ke dalam kategori pembelajaran ensemble. Pendekatan ini menggabungkan sejumlah model prediksi (biasanya berupa pohon keputusan) untuk memperoleh hasil yang lebih stabil dan akurat. Konsep *Random Forest* pertama kali diperkenalkan oleh Breiman pada tahun 2001. Sejak saat itu, metode ini terus dikembangkan dan banyak digunakan pada berbagai permasalahan analisis data, baik untuk tujuan klasifikasi maupun regresi..

Secara umum, *Random Forest* terdiri dari dua varian utama, yaitu *Random Forest Classifier* dan *Random Forest Regressor*. Keduanya menggunakan prinsip dasar pohon keputusan (*decision tree*) namun dibedakan berdasarkan jenis output yang dihasilkan. *Random Forest Classifier* digunakan untuk memprediksi variabel

kategorikal, sementara *Random Forest Regressor* digunakan untuk memprediksi nilai kontinu, seperti jumlah penjualan, suhu, atau harga.

Metode ini bekerja dengan cara membangun banyak *decision tree* secara paralel melalui teknik *bootstrap sampling*, lalu menggabungkan prediksi semua pohon dengan metode voting (klasifikasi) atau rata-rata (regresi). Proses ini membuat *Random Forest* sangat efektif dalam mengurangi *overfitting* dan meningkatkan kestabilan model.

2.5. Metrik Evaluasi Model

Evaluasi model merupakan proses penting dalam pembelajaran mesin (*machine learning*) dan pembelajaran mendalam (*deep learning*) untuk mengukur sejauh mana model mampu menghasilkan prediksi yang akurat terhadap data yang tidak pernah dilihat sebelumnya, yaitu data uji. Tujuan utama dari evaluasi ini adalah untuk menilai kemampuan generalisasi model, mengidentifikasi kemungkinan *overfitting* atau *underfitting*, serta membandingkan performa dari beberapa model atau konfigurasi berbeda (Simamora et al., 2025).

2.5.1. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) adalah akar kuadrat dari MSE. Metrik ini berguna karena nilai akhirnya memiliki satuan yang sama dengan nilai target, misalnya dalam konteks jumlah obat, sehingga lebih mudah untuk diinterpretasikan secara praktis. Rumus RMSE ditulis sebagai:

$$\mathbf{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}$$

RMSE sering digunakan sebagai metrik utama dalam evaluasi model regresi karena mencerminkan secara langsung seberapa jauh hasil prediksi dari model

menyimpang dari nilai sebenarnya. Dalam penelitian oleh Rosyid et al. (2024), RMSE digunakan untuk membandingkan performa beberapa model regresi dalam memprediksi nilai aktivitas senyawa obat. Model dengan nilai RMSE paling rendah dianggap paling baik dalam menangkap pola data yang akurat.

2.5.2. R² Score (Koefisien Determinasi)

R² Score atau Koefisien Determinasi merupakan metrik evaluasi yang mengukur seberapa besar variansi dari nilai target y yang dapat dijelaskan oleh model. Metrik ini memberikan informasi tentang kualitas prediksi model dalam menjelaskan dinamika data. Rumus R² Score ditulis sebagai:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Dimana:

$$SS_{res} = \sum (y_i - \hat{y}_i)^2 = \text{jumlah kuadrat residual (kesalahan model)}$$

$$SS_{tot} = \sum (y_i - \bar{y})^2 = \text{total variansi dari nilai aktual terhadap rata-rata}$$

$\bar{y}$ = rata-rata dari seluruh nilai aktual.

Nilai R² berkisar antara 0 dan 1. Semakin mendekati 1, semakin baik model dalam menjelaskan pola data. Jika nilai R² negatif, artinya model lebih buruk daripada model baseline yang hanya menggunakan rata-rata. Dalam konteks prediksi permintaan produk, R² digunakan sebagai indikator penting yang menunjukkan seberapa besar kepercayaan terhadap model prediksi yang dibangun (Rosyid et al., 2024). menekankan pentingnya metrik ini dalam prediksi obat, karena selain mampu menunjukkan kualitas kuantitatif dari model, ia juga membantu dalam mengevaluasi kemampuan model dalam menangkap struktur data kompleks secara menyeluruh.

2.5.3. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah metrik evaluasi yang menghitung rata-rata dari nilai absolut selisih antara nilai aktual dan nilai prediksi. MAE memberikan gambaran langsung tentang seberapa besar kesalahan rata-rata yang dihasilkan model tanpa memperhitungkan arah kesalahan (positif atau negatif). Rumus MAE adalah sebagai berikut:

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|$$

Dimana :

$MAE = \text{Mean Absolute Error}$

$m = \text{jumlah data}$

$X_i = \text{nilai prediksi (predicted value) dari data ke-}i$

$Y_i = \text{nilai aktual dari data ke-}i$

MAE memberikan informasi dalam satuan yang sama dengan variabel target, sehingga mudah dipahami secara praktis. Sebagai contoh, jika targetnya adalah jumlah penjualan obat harian, maka MAE menunjukkan rata-rata jumlah unit obat yang diprediksi salah per hari.

Dalam penelitian oleh Rosyid et al. (2024), MAE digunakan bersama RMSE dan R^2 untuk mengevaluasi model regresi dalam prediksi aktivitas senyawa obat. Hasil penelitian menunjukkan bahwa MAE sangat berguna dalam memberikan gambaran umum kesalahan model secara konsisten dan dapat dijadikan acuan utama dalam pengambilan keputusan berbasis data prediktif.

2.5.4. Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan ukuran kesalahan yang dihitung berdasarkan rata-rata selisih absolut antara nilai hasil prediksi dan nilai aktual, yang kemudian dinyatakan dalam bentuk persentase terhadap nilai aktual tersebut. Penggunaan Mean Absolute Percentage Error (MAPE) pada evaluasi dari hasil peramalan dapat melihat tingkat akurasi terhadap angka peramalan dan angka realisasi.

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{X_i - Y_i}{Y_i} \right| \times 100$$

Dimana :

$MAPE = \text{Mean Absolute Percentage Error}$

$m = \text{jumlah data}$

$X_i = \text{nilai prediksi oleh model dari data ke-}i$

$Y_i = \text{nilai aktual dari data ke-}i$

2.5.5. Mean Squared Error (MSE)

Mean Squared Error (MSE) merupakan salah satu ukuran kinerja regresi yang paling sering digunakan, berfungsi untuk menghitung nilai rata-rata dari selisih kuadrat antara data aktual dan hasil prediksi. Secara matematis, perhitungan MSE dapat dinyatakan dengan rumus berikut:

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2$$

Dimana :

$MSE = \text{Mean Squared Error}$

$n = \text{jumlah total data (observasi)}$

X_i = nilai aktual pada data ke-i

Y_i = nilai prediksi oleh model pada data ke-i

2.6. Studi Literatur

Tabel 2.1 Studi Literatur

No.	Penulis (Tahun)	Judul Penelitian	Hasil Penelitian
1	K. Puteri dan A. Silvanie (2020)	<i>Machine Learning Untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linier Berganda</i>	Melakukan perbandingan antara metode regresi linier dan random forest dalam proses prediksi harga minyak kelapa sawit. Hasil penelitian menunjukkan bahwa regresi linier memberikan kinerja yang lebih unggul dengan nilai RMSE sebesar 31.174, sementara model random forest memperoleh RMSE sebesar 32.850. Penelitian ini menggunakan pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian.
2	Y. Supriyanto, M. Ilhamsyah, dan U. Enri	<i>Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression Dan Random Forest</i>	Hasil penelitian menunjukkan bahwa model menghasilkan nilai R-squared sebesar 84,2%, dengan tiga variabel paling dominan yaitu tanggal, jenis komoditas, dan faktor pasar.
3	Arifuddin, D., Kusrini, & Kusnawi. (2025)	<i>Perbandingan Performansi Algoritma Multiple Linear Regression dan Multi</i>	Penelitian ini membandingkan performa <i>Multiple Linear Regression</i> (MLR) dan <i>Multi -Layer Perceptron Neural Network</i> (MLPNN) dalam

		<i>Layer Perceptron Neural Network dalam Memprediksi Penjualan Obat</i>	memprediksi penjualan obat analgesik di Apotek XYZ. Namun, kedua model masih memiliki keterbatasan dalam menghasilkan prediksi yang akurat, dengan nilai MSE di atas 19.000 dan rentang kesalahan prediksi mencapai 22,3%, yang mengindikasikan adanya underfitting.
4	Pramesti, D., & Baihaqi, W. M. (2023).	<i>Perbandingan Prediksi Jumlah Transaksi Ojek Online Menggunakan Regresi Linier dan Random Forest</i>	Berdasarkan nilai R-squared, baik model Random Forest maupun regresi linier memperoleh hasil yang sangat tinggi, hampir mencapai 1.0 pada data pelatihan maupun pengujian. Hal ini menunjukkan bahwa kedua model tersebut memiliki kemampuan yang baik dalam menjelaskan variasi dari variabel target.
5	Ayuni, G. N., & Fitriana, D. (2022).	<i>Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PTXYZ</i>	Hasil peramalan penjualan properti dengan metode regresi linier diperoleh memiliki kinerja yang cukup baik. Penilaian tersebut didasarkan pada hasil pengujian menggunakan metrik MSE, RMSE, dan MAPE terhadap seluruh jenis properti di PT. XYZ yang dinilai telah memenuhi standar.

2.7.1 Kesenjangan Penelitian

Berdasarkan studi pustaka yang telah dikaji sebelumnya, terlihat bahwa penggunaan metode regresi seperti *Random Forest* dan *Multiple Linear Regression* (MLR) telah banyak diterapkan dalam konteks prediksi penjualan, baik di sektor retail umum maupun farmasi. Hasilnya menunjukkan bahwa *Random Forest* cenderung memiliki keunggulan dalam hal akurasi, terutama ketika data yang digunakan bersifat kompleks dan memiliki pola non-linear.

Namun, terdapat beberapa kesenjangan penelitian yang masih bisa dijadikan ruang kontribusi dalam penelitian ini, antara lain:

1. Minimnya penelitian yang secara langsung membandingkan *Random Forest* dan MLR dalam konteks prediksi penjualan harian obat secara lokal di apotek individu bukan hanya dataset ritel umum. Kebanyakan studi masih menggunakan data agregat atau berskala perusahaan besar, bukan unit kecil seperti Apotek Bambuan.
2. Evaluasi kinerja model pada studi sebelumnya sebagian besar hanya mengandalkan satu atau dua metrik, sedangkan penelitian ini secara eksplisit akan membandingkan model berdasarkan MAE, MAPE, MSE, RMSE, dan R^2 secara bersamaan untuk memberikan evaluasi yang lebih komprehensif.
3. Belum banyak penelitian yang fokus pada prediksi berbasis data harian di sektor apotek, padahal pola penjualan obat sangat dipengaruhi oleh hari, musim, dan faktor lokal lainnya yang berbeda dengan pola penjualan barang konsumsi umum.

Dengan mempertimbangkan kesenjangan-kesenjangan tersebut, penelitian ini bertujuan untuk memberikan kontribusi baru dalam bentuk komparasi model *prediktif* yang relevan secara praktis dan operasional bagi pengelola apotek, khususnya dalam pengambilan keputusan harian berbasis data.

BAB III

METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif komparatif dengan pendekatan eksperimen komputasional. Metode kuantitatif dipilih karena penelitian ini bertujuan untuk mengukur dan membandingkan performa dua model regresi, yaitu *Random Forest Regression* dan *Multiple Linear Regression*, dalam memprediksi jumlah penjualan obat harian berdasarkan data historis dari Apotek Bambuan. Data yang digunakan bersifat numerik dan dianalisis secara objektif menggunakan perhitungan statistik.

Pendekatan komparatif digunakan karena fokus utama penelitian ini adalah membandingkan akurasi dan efisiensi dari dua algoritma prediktif yang berbeda. Kedua model dianalisis dan diuji menggunakan metrik evaluasi yang sama, yaitu *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Root Mean Squared Error* (RMSE), *Mean Squared Error* (MSE) dan *Coefficient of Determination* (R^2). Dengan demikian, hasil evaluasi dapat digunakan untuk menentukan model mana yang lebih unggul dalam memodelkan pola penjualan obat harian yang cenderung fluktuatif.

Sementara itu, pendekatan eksperimen komputasional digunakan karena seluruh proses pelatihan, validasi, dan pengujian model dilakukan dengan bantuan perangkat lunak berbasis *Python*, serta pustaka *machine learning* seperti *Scikit-learn* dan *Pandas*. Model dilatih menggunakan data historis penjualan yang telah diproses terlebih dahulu, dan hasil prediksinya dibandingkan secara langsung.

Dengan lingkungan eksperimen yang terstruktur dan sistematis, penelitian ini diharapkan mampu memberikan gambaran empiris tentang efektivitas kedua metode dalam konteks dunia farmasi, khususnya dalam perencanaan stok dan pengambilan keputusan di Apotek Bambuan.

3.2. Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari sistem pencatatan transaksi penjualan harian di Apotek Bambuan. Data ini mencakup informasi jumlah obat yang terjual setiap hari selama periode waktu tertentu. Selain itu, data dapat mencakup atribut tambahan seperti tanggal transaksi, nama atau kode obat, dan kategori obat jika tersedia.

Teknik pengumpulan data dilakukan dengan cara mengakses data historis penjualan yang telah direkap oleh pihak apotek, baik melalui file digital seperti Microsoft Excel (CSV/XLSX) maupun melalui database internal jika tersedia. Peneliti kemudian melakukan ekstraksi dan penyimpanan data tersebut ke dalam format yang sesuai untuk proses analisis lebih lanjut menggunakan perangkat lunak *Python*..

Untuk menjaga validitas data, dilakukan proses verifikasi awal guna memastikan tidak terdapat data duplikat, nilai kosong atau *missing values*, atau inkonsistensi format. Selanjutnya, data akan digunakan sebagai basis untuk membangun model prediksi, di mana pembagian dataset dilakukan menjadi data latih dan data uji.

Penggunaan data sekunder ini dinilai efisien karena memungkinkan peneliti melakukan eksplorasi dan eksperimen secara langsung tanpa harus melakukan proses observasi atau pengumpulan data primer yang memakan waktu.

Tabel 3.1 Contoh Dataset Apotek Bambuan

Tanggal	Hari	Jumlah_Obat_Terjual	Toko_Tutup	Total_Omset
2022-12-29	Kamis	28	0	344000
2022-12-30	Jumat	56	0	524500
2022-12-31	Sabtu	62	0	708000
2023-01-01	Minggu	65	0	831500
2023-01-02	Senin	0	1	0

3.3. Alat dan Perangkat Penelitian

Penelitian ini menggunakan berbagai alat bantu, baik perangkat keras maupun perangkat lunak, yang mendukung proses eksperimen komputasional. Alat dan perangkat tersebut dijabarkan dalam Tabel 3.5 berikut:

Tabel 3.2 Alat dan Perangkat Penelitian

No	Nama	Jenis	Spesifikasi / Keterangan
1	Asus TUF F15	Perangkat	RAM 16 GB, SSD 1024 GB, GPU NVIDIA RTX 3050, OS Windows 11
2	Google Colab	Alat	Platform komputasi awan berbasis Python, mendukung GPU/TPU
3	Python	Alat	Versi 3.10 atau lebih baru
4	Scikit-learn	Alat	Library machine learning (versi 1.x)
5	Matplotlib & Seaborn	Alat	Library visualisasi data
6	Numpy	Alat	Untuk penyusunan dataset time series

Penelitian ini didukung oleh seperangkat alat dan tools berbasis komputasi yang digunakan untuk menunjang proses pelatihan dan evaluasi model machine learning. Berdasarkan Tabel 3.1, laptop yang peneliti gunakan adalah ASUS TUF Gaming F15 FX506HC dengan spesifikasi RAM 16 GB, SSD 512 GB, dan GPU NVIDIA GeForce RTX 3050, yang berjalan pada sistem operasi Windows 11. Spesifikasi ini cukup untuk mendukung proses eksperimen komputasi lokal, seperti pelatihan model, pengolahan data, dan visualisasi hasil prediksi.

Proses pelatihan model dilakukan sepenuhnya secara lokal, tanpa menggunakan platform komputasi awan, karena spesifikasi perangkat keras yang digunakan telah memadai untuk menjalankan pemodelan regresi skala menengah. Pengolahan data dilakukan dengan bantuan bahasa pemrograman *Python* versi 3.10, yang dipilih karena fleksibel, efisien, dan memiliki banyak pustaka pendukung untuk analisis data dan *machine learning*.

Pustaka utama yang digunakan dalam penelitian ini meliputi:

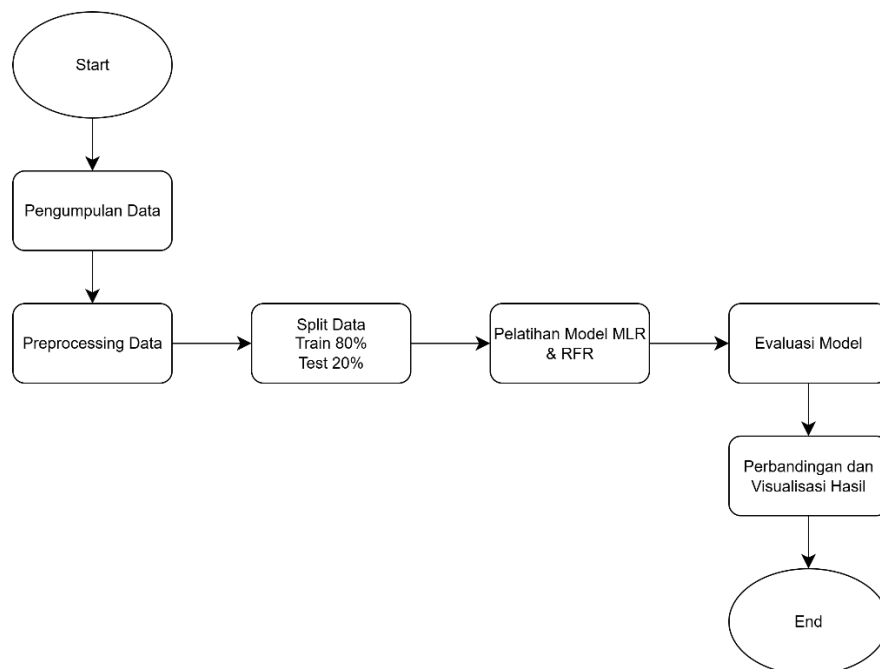
- a. *Pandas* dan *NumPy* untuk proses eksplorasi dan manipulasi data numerik,
- b. *Scikit-learn* untuk membangun dan melatih model *Multiple Linear Regression* (MLR) dan *Random Forest Regression* (RFR),
- c. serta *Matplotlib* dan *Seaborn* untuk proses visualisasi grafik hasil prediksi, sebaran error, dan perbandingan performa antar model.

Data yang digunakan dalam penelitian diperoleh dalam bentuk file Excel (.xlsx) atau CSV (.csv) yang berisi catatan penjualan harian dari Apotek Bambuan. Data ini menjadi dasar dalam membangun model prediktif terhadap jumlah obat yang terjual.

Dengan kombinasi perangkat keras dan pustaka perangkat lunak tersebut, seluruh tahapan penelitian dapat dijalankan secara efisien, sistematis, dan dapat direplikasi oleh peneliti lain dengan sumber daya serupa.

3.4. Diagram Alur Penelitian atau Tahapan Penelitian

Diagram alur digunakan untuk menggambarkan tahapan-tahapan utama dalam penelitian ini secara sistematis. Visualisasi ini mempermudah pemahaman terhadap proses penelitian, mulai dari pengumpulan data hingga evaluasi dan perbandingan model.



Gambar 3.1 Alur Penelitian

Untuk memperjelas alur pada Gambar 3.1, berikut ini adalah uraian setiap tahapan berdasarkan urutan pada diagram:

1. Mulai (*Start*)

Penelitian dimulai dengan merumuskan tujuan dan ruang lingkup, serta menetapkan dua model utama yang akan dibandingkan, yaitu *Multiple Linear Regression* (MLR) dan Random Forest.

2. Pengumpulan Data

Data Pada tahap ini, peneliti mengumpulkan data historis penjualan obat harian dari sistem pencatatan transaksi di Apotek Bambuan. Data ini biasanya mencakup tanggal penjualan dan jumlah unit obat yang terjual. Format data bisa berupa Excel (.xlsx) atau CSV (.csv). Data mencakup periode 2022 hingga 2025.

3. Preprocessing Data

Data yang diperoleh kemudian dibersihkan dan dipersiapkan. Proses ini meliputi penghapusan data duplikat, penanganan nilai kosong (missing value), konversi tanggal menjadi variabel numerik (misalnya hari ke-), dan normalisasi jika diperlukan. Tahap ini krusial agar data sesuai untuk dianalisis oleh model regresi. Normalisasi data menggunakan Min-Max Scaler.

4. Split Dataset

Setelah data diproses, data dibagi menjadi dua bagian:

- a. Data latih (training set) sebesar 80%, digunakan untuk melatih model
- b. Data uji (test set) sebesar 20%, digunakan untuk menguji akurasi model terhadap data yang belum pernah dilihat sebelumnya.

5. Pelatihan Model

Proses training dilakukan pada data latih untuk kedua model.

6. Evaluasi Model

Performa masing-masing model diuji pada data uji menggunakan tiga metrik evaluasi regresi: MAE, MAPE, MSE, RMSE, dan R^2 Score.

7. Perbandingan Hasil dan Visualisasi

Hasil evaluasi dari model MLR dan Random Forest dibandingkan secara numerik (berdasarkan nilai metrik) dan visual (plot harga aktual vs prediksi), untuk menilai akurasi dan kemampuan generalisasi masing-masing model.

8. Selesai (End)

Penelitian disimpulkan dengan menentukan model terbaik serta menyusun hasil dan temuan ke dalam laporan skripsi.

3.5. Teknik Analisis Data

Teknik analisis data dalam penelitian ini dilakukan dengan pendekatan eksperimen komputasional berbasis *Python*, yang mencakup tahapan *preprocessing*, pelatihan model, evaluasi matriks, dan perbandingan hasil antar model.

3.5.1. Preprocessing Data

Sebelum dilakukan pelatihan model, data penjualan yang didapatkan langsung dari Apotek Bambuan perlu melalui tahap *preprocessing* untuk memastikan struktur data sesuai dengan kebutuhan model sehingga dapat *ditraining*. Proses ini bertujuan untuk meningkatkan akurasi prediksi, menghindari *noise*, dan mengatasi perbedaan skala data yang dapat memengaruhi hasil pembelajaran model. Tahapan yang umum dilakukan pada tahap ini adalah menentukan input dan target atau variabel independen dan dependen, melakukan normalisasi data sehingga nilai yang memiliki range jauh dapat diatasi dan dipelajari dengan mudah, dan pembagian dataset atau *split data* menjadi 80% data *training* dan 20% data *testing*.

3.5.2. Pelatihan Model

Pada Tahapan Pada tahap ini, terdapat dua algoritma yang digunakan dalam proses perbandingan, yaitu *Multiple Linear Regression* (MLR) dan *Random Forest Regression* (RFR), dilatih menggunakan data yang telah diproses sebelumnya. Model MLR dikembangkan dengan menerapkan fungsi regresi linier berganda yang menghitung koefisien berdasarkan hubungan linier antara variabel bebas dan variabel terikat. Sementara itu, model *Random Forest* dilatih dengan cara membentuk beberapa pohon keputusan secara paralel, di mana setiap pohon yang dibor menggunakan subset data dan fitur yang dipilih secara acak.

Proses pelatihan dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka seperti scikit-learn. Parameter default digunakan terlebih dahulu untuk mendapatkan baseline performa awal dari masing-masing model. Selama proses ini, model mempelajari pola hubungan antar variabel untuk meminimalkan kesalahan prediksi pada data training.

3.5.3. Evaluasi Kinerja Model

Setelah proses pelatihan selesai, kedua model kemudian dievaluasi menggunakan data pengujian yang telah disiapkan secara terpisah.. Evaluasi ini bertujuan untuk menilai sejauh mana model dapat memprediksi jumlah penjualan obat harian berdasarkan data yang belum pernah digunakan sebelumnya.

Evaluasi dilakukan menggunakan beberapa metrik, yaitu:

- a. *Mean Absolute Error* (MAE): Mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi.
- b. *Mean Absolute Percentage Error* (MAPE): Mengukur persentase tingkat kesalahan prediksi sebuah model dalam bentuk persentase.

- c. *Mean Squared Error (MSE)*: Mengukur seberapa akurat suatu model dalam memprediksi nilai numerik.
- d. *Root Mean Squared Error (RMSE)*: Menghitung akar dari rata-rata kuadrat selisih antara nilai aktual dan prediksi.
- e. *R-squared (R² Score)*: Menunjukkan proporsi variasi dari variabel dependen yang dapat dijelaskan oleh variabel independen.

Ketiga metrik ini digunakan untuk memberikan gambaran yang komprehensif tentang akurasi prediksi dan kemampuan generalisasi model.

3.5.4. Perbandingan Hasil dan Visualisasi

Tahap akhir setelah proses evaluasi selesai dilakukan terhadap masing-masing model, langkah berikutnya adalah membandingkan kinerja antara model *Multiple Linear Regression (MLR)* dan *Random Forest Regression (RFR)* berdasarkan nilai metrik evaluasi seperti MAE, RMSE, dan R². Tujuan dari perbandingan ini adalah untuk mengidentifikasi model yang paling optimal dalam memprediksi jumlah penjualan obat harian.

Perbandingan dilakukan dengan menampilkan nilai evaluasi masing-masing model dalam bentuk tabel dan grafik untuk mempermudah interpretasi. Visualisasi ini dibuat menggunakan pustaka Python seperti matplotlib dan seaborn, yang dapat menggambarkan sejauh mana akurasi prediksi dari setiap model serta memperlihatkan perbedaan performa secara visual. Dengan adanya visualisasi ini, pembaca dapat dengan mudah memahami efektivitas masing-masing algoritma dalam memodelkan data penjualan obat harian.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Deskripsi Dataset

Penelitian ini menggunakan data penjualan obat harian yang diperoleh dari Apotek Bambuan, berlokasi di Kecamatan Stabat, Kabupaten Langkat. Data ini bersifat sekunder, karena peneliti tidak melakukan pengumpulan data primer (observasi atau survei), melainkan memanfaatkan data historis yang sudah tersedia. Data ini diambil langsung dari catatan transaksi apotek selama periode tiga tahun terakhir. Data awal diperoleh dalam bentuk file Excel (.xlsx) / CSV (.csv) yang merupakan hasil rekap penjualan harian apotek dengan isi data terdapat lima komponen utama seperti tanggal transaksi, hari, jumlah obat terjual, toko tutup atau sedang buka dan total omset per harinya, informasi ini nantinya digunakan sebagai basis untuk membangun model prediksi perkiraan jumlah obat terjual perharinya.

Proses pengambilan data dilakukan melalui koordinasi langsung dengan pengelola Apotek Bambuan. Data transaksi yang awalnya masih berbentuk pencatatan manual dalam buku direkap ke dalam file digital. Selanjutnya peneliti melakukan verifikasi awal terhadap data digital tersebut untuk memastikan kelengkapan dan konsistensinya. Kemudian untuk data yang diambil merupakan periode data dari tahun 2022 hingga tahun 2024, rentang waktu ini dipilih untuk memastikan ketersediaan data historis yang cukup panjang sehingga model dapat mengenali data dengan baik.

4.2 Hasil Preprocessing Data

Sebelum data digunakan untuk melatih model, dilakukan tahap preprocessing agar data bersih, konsisten, dan sesuai dengan kebutuhan algoritma machine

learning. Proses preprocessing ini sangat penting karena kualitas data akan sangat memengaruhi akurasi model yang dihasilkan. Namun pada umumnya dikarenakan data yang digunakan adalah data yang dipindahkan secara manual dari buku ke excel maka pada tahap preprocessing ini tidak memakan waktu terlalu banyak.

Tabel 4.1 Pseudocode Preprocessing

Pseudocode 1 : Preprocessing

Input : File Data Skripsi

Output : File Hasil Preprocessing

1. `df = pd.read_excel("Data_Skripsi")`

//Deklarasi variabel df untuk membaca file excel dengan nama file Data_Skripsi yang kemudian keseluruhan isi file dapat dipanggil dengan menggunakan variabel df.

2. `df['Tanggal'] = pd.to_datetime(df['Tanggal'], errors='coerce')`

//Mengganti type data yang terdapat pada kolom tanggal menjadi datetime atau waktu serta untuk memastikan bahwa type data pada kolom Tanggal sesuai.

3. `df = pd.get_dummies(df, columns=['Hari'])`

//Menggunakan fungsi dari pandas untuk melakukan fungsi One Hot Encoding pada kolom Hari sehingga setelah proses ini akan terbentuk kolom baru yaitu Hari_Senin, Hari_Selasa dan seterusnya sampai Hari_Minggu dengan isi pada kolom berupa nilai 1/0 dengan contoh apabila hari itu hari selasa maka kolom hari selasa bernilai 1 dan kolom lainnya memiliki nilai 0

4. `df['Hari_ke_tahun'] = df['Tanggal'].dt.dayofyear`

`df['Bulan'] = df['Tanggal'].dt.month`

```
df['Tahun'] = df['Tanggal'].dt.year
```

//Menggunakan pandas untuk mengambil nilai dari kolom tanggal dan membuat kolom baru yaitu hari keberapa pada tahun ini, bulan keberapa dan tahun berapa sehingga model bisa mempelajari pola musiman yaitu harian bulanan ataupun jika data yang dimiliki mempunyai jangka waktu hingga puluhan tahun maka sangat mungkin membaca pola tahunan.

```
5. df['Rolling7'] = df['Total_Obat'].shift(1).rolling(window=7, min_periods=1).mean()
```

//Membuat kolom baru bernama Rolling 7 yaitu rata rata penjualan obat dalam 7 hari sebelumnya.

Setelah proses preprocessing dilakukan data yang sudah akan digunakan untuk tahap selanjutnya adalah membagi data menjadi 2 yaitu 80% data training dan 20% data testing.

Tabel 4.2 Pseudocode 2 Data Split

Pseudocode 2 : Data Split
Input : Data hasil preprocessing, feature_cols, target_col Output : X_train_scaled, y_train, X_test_scaled, y_test, X_train, X_test 1. target_col = 'Total_Obat' //Deklarasi variabel target_col menentukan nilai atau kolom mana yang digunakan sebagai prediksi, disini digunakan kolom Total_Obat sebagai output akhir. 2. feature_cols = [c for c in df.columns if c not in ['Tanggal', target_col]]

```

//deklarasi variabel feature_cols memiliki fungsi untuk memilih semua
kolom sebagai faktor pendukung atau variabel input kecuali kolom
Tanggal dan kolom target yaitu Total Obat.

3. test_start = pd.Timestamp('2024-09-15')

//Menentukan tanggal mulai data test yaitu mulai dari tanggal 15 bulan 9
tahun 2024 sebagai data awal test data dibawahnya akan dijadikan data
training.

4. test_end = pd.Timestamp('2024-11-15')

//Menentukan data akhir data test yaitu 15 November 2024 sebagai data
test akhir.

5. mask_test = (df['Tanggal'] >= test_start) & (df['Tanggal'] <= test_end)

//deklarasi variabel mask test digunakan sebagai penanda untuk memilih
baris data diantara variabel test start dan variabel test end yang akan
digunakan sebagai data test.

6. if mask_test.sum() == 0:

    split_idx = int(len(df)*0.8)

    train_df = df.iloc[:split_idx].copy()

    test_df = df.iloc[split_idx:].copy()

    //Jika variabel mask test deklarasi sebelumnya ternyata tidak ada data
    didalamnya maka data akan dibagi secara otomatis 80% untuk data latih
    dan 20% untuk data uji berdasarkan panjang data.

    else:

        train_df = df.loc[~mask_test].copy()

        test_df = df.loc[mask_test].copy()

```

//Jika ternyata di dalam variabel mask terdapat data yang ditetapkan periodenya ada maka akan dijadikan data test di dalam periode tersebut dan diluar periode tersebut akan dijadikan data uji.

7. `X_train, y_train = train_df[feature_cols], train_df[target_col]`

//variabel x_train yaitu variabel yang memiliki isi faktor pendukung untuk melatih model, isinya semua kolom fitur kecuali tanggal dan total_obat, dan y_train adalah kolom yang mau diprediksi yaitu total_obat.

`X_test, y_test = test_df[feature_cols], test_df[target_col]`

//X_test dan Y_test memiliki fungsi yang sama seperti x_train dan y_train hanya saja digunakan sebagai data test bukan sebagai data training.

8. `scaler = MinMaxScaler()`

`X_train_scaled = scaler.fit_transform(X_train)`

`X_test_scaled = scaler.transform(X_test)`

//Pada tahap ini melakukan penyamaan skala data agar range yang dihasilkan tidak terlalu jauh sehingga nilai pada tiap kolom hanya dikisaran 0-1 supaya adil.

Setelah proses preprocessing dan split data yaitu membagi data menjadi 2, 80% data training dan 20% data testing serta melakukan normalisasi yaitu menggunakan min max scaler untuk membuat nilai dari tiap kolom adil sehingga tiap kolom memiliki nilai hanya diantara range 0-1 langkah selanjutnya yaitu pelatihan kedua model menggunakan Random Forest dan Multiple Linear Regression.

4.3 Implementasi Model Random Forest dan MLR

4.3.1 Arsitektur Model

Model Multiple Linear Regression (MLR) dibangun menggunakan pustaka *scikit-learn*. MLR berfungsi untuk memodelkan hubungan linear antara variabel dependen, yaitu jumlah penjualan obat (*Total_Obat*), dengan sejumlah variabel independen yang telah dihasilkan dari proses preprocessing (misalnya fitur waktu, dummy hari, lag, dan rolling average). Implementasi program menggunakan *LinearRegression()* dari *scikit-learn* tanpa parameter tambahan karena baseline ini sudah cukup merepresentasikan pendekatan linear. Proses fitting dilakukan pada data training yang telah dinormalisasi, sedangkan hasil evaluasi menggunakan metrik *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan R^2 Score.

Random Forest Regression dibangun menggunakan pustaka *scikit-learn* dengan kelas *RandomForestRegressor*. Berbeda dengan MLR yang hanya membentuk satu persamaan linear, Random Forest merupakan ensemble learning berbasis pohon keputusan (*decision trees*). Model ini bekerja dengan membangun banyak pohon keputusan secara acak, kemudian hasil prediksi tiap pohon digabungkan atau dirata-rata untuk menghasilkan prediksi akhir yang lebih stabil dan akurat. Skema kerja Random Forest dapat dijelaskan sebagai berikut:

1. Dataset training di-bootstrapping menjadi subset berbeda untuk tiap pohon.
2. Setiap pohon dilatih menggunakan subset data tersebut dengan pemilihan fitur secara acak.
3. Saat melakukan prediksi, setiap pohon memberikan hasilnya masing-masing.

4. Hasil akhir regresi adalah rata-rata dari seluruh prediksi pohon.

Evaluasi dilakukan dengan metrik yang sama seperti MLR, yaitu *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*.

Tabel 4.3 Pseudocode 3 Model

Pseudocode 3 Membuat Model Multiple Linear Regression
Input : X_train_scaled, y_train, X_test_scaled, y_test Output : pred_mlr, mae_mlr, rmse_mlr, r2_mlr 1. mlr = LinearRegression() //Inisiasi variabel mlr untuk memanggil pustaka linear regression 2. mlr.fit(X_train_scaled, y_train) //Latih model dengan menggunakan fitur training yang sudah dinormalisasi yaitu semua kolom yang ada pada dataset kecuali tanggal dan jumlah_obat dan target training sebagai hasil akhir. 3. pred_mlr = mlr.predict(X_test_scaled) //Prediksi nilai target untuk data uji dengan data yang sudah dinormalisasi dengan scaler yang sama yaitu yang memiliki range 0-1. 4. mae_mlr = mean_absolute_error(y_test, pred_mlr) rmse_mlr = mean_squared_error(y_test, pred_mlr) r2_mlr = r2_score(y_test, pred_mlr) //Menghitung matriks evaluasi dengan mae yaitu rata-rata selisih absolut, kemudian RMSE yaitu akar mse dan r2 yaitu proporsi variasi yang dapat dijelaskan model semakin mendekati 1 maka semakin baik. 5. print("\n=== J1. Evaluasi MLR ===")

```

print(f"MAE : {mae_mlr:.3f}")

print(f"RMSE: {rmse_mlr:.3f}")

print(f"R2 : {r2_mlr:.3f}")

//Menampilkan hasil evaluasi dengan fungsi print dan menampilkan
ketiga matriks evaluasi dengan format 3 desimal.

```

Pseudocode diatas menjelaskan tentang training model dari Multiple Linear Regression dengan memanggil pustaka dari scikit learn dengan menggunakan 3 matriks evaluasi untuk mengetahui hasil dari model untuk mengenali pola data dari dataset yang tersedia.

Tabel 4.4 Pseudocode 4 Model Random Forest

Pseudocode 4 Membuat Model Random Forest Regression
Input : X_train_scaled, y_train, X_test_scaled, y_test Output : pred_mlr, mae_mlr, rmse_mlr, r2_mlr <pre> from sklearn.ensemble import RandomForestRegressor from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score //Memanggil pustaka model random forest dan matriks evaluasi. </pre> <pre> rf = RandomForestRegressor(n_estimators=400, max_depth=None, random_state=42, n_jobs= -1) </pre>


```
//Menginisiasi variabel df dengan memanggil random forest dengan
menetapkan jumlah pohon sebanyak 400, kedalaman pohon bebas dan
memberikan keterangan memakai semua core cpu yang tersedia agar
training lebih cepat misalkan ingin dibatasi dapat menerapkan misalkan
n_jobs = 4 maka hanya menggunakan 4 core.
```

```
3. rf.fit(X_train, y_train)
```

```
//Melatih model dengan menggunakan fitur training tanpa normalisasi
karena random forest tidak butuh scaling dikarenakan berbasis pohon
keputusan.
```

```
4. pred_rf = rf.predict(X_test)
```

```
//Prediksi nilai target untuk data uji.
```

```
5. mae_rf = mean_absolute_error(y_test, pred_rf)
```

```
rmse_rf = mean_squared_error(y_test, pred_rf, squared=False)
```

```
r2_rf = r2_score(y_test, pred_rf)
```

```
// Sama seperti sebelumnya pada pseudocode 3 yaitu menghitung matriks
evaluasi dengan menggunakan 3 matriks evaluasi yaitu MAE, RMSE dan
R2 Score.
```

```
6. print("\n=== Evaluasi Random Forest ===")
```

```
print(f'MAE : {mae_rf:.3f}')
```

```
print(f'RMSE: {rmse_rf:.3f}')
```

```
print(f'R2 : {r2_rf:.3f}')
```

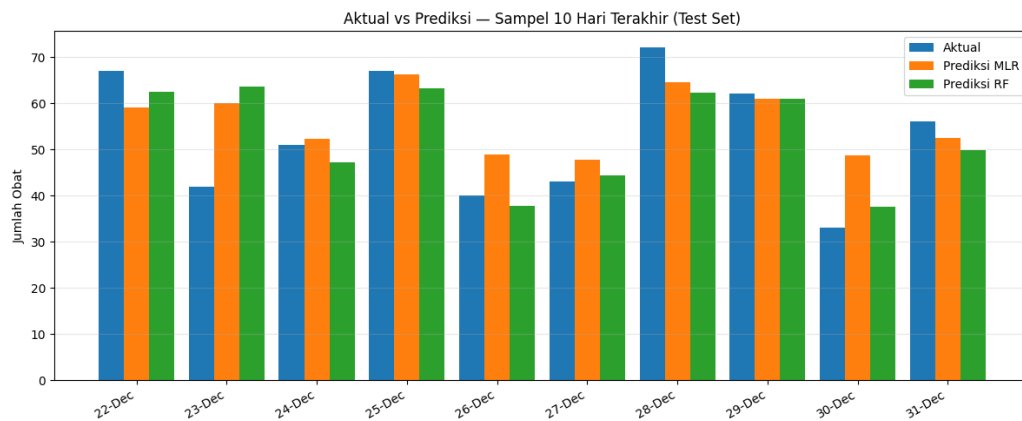
```
//Menampilkan hasil evaluasi dari matriks evaluasi dengan menggunakan
fungsi print.
```

4.3.2. Hasil Konfigurasi Awal Model

Setelah arsitektur model *Multiple Linear Regression* (MLR) dan *Random Forest Regression* (RFR) dibangun dengan konfigurasi awal (baseline), kedua model kemudian dilatih menggunakan dataset penjualan obat dari Apotek Bambuan. Tujuan dari pelatihan awal ini adalah untuk mengevaluasi kemampuan dasar masing-masing model dalam memprediksi jumlah obat terjual berdasarkan pola historis sebelum dilakukan optimasi lebih lanjut.

Selama proses pelatihan, MLR langsung menghasilkan persamaan linear yang menghubungkan variabel prediktor dengan target. Berbeda dengan model jaringan saraf yang memerlukan iterasi epoch, MLR tidak mencatat *training loss* secara bertahap. Oleh karena itu, evaluasi dilakukan langsung terhadap data uji dengan menggunakan metrik *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*. Jadi bukan seperti model LSTM ataupun Neural Network yang melakukan update di dalam 1 model berulang kali menggunakan data yang sama sampai konvergen namun nilai matriks yang dihasilkan merupakan nilai sekali hitung yang langsung mendapatkan nilai solusi optimal tanpa memerlukan mengetahui epoch atau berapa kali percobaan yang diperlukan sehingga tidak ada training loop.

Pada model Random Forest, pelatihan dilakukan dengan membangun ratusan pohon keputusan yang masing-masing dilatih pada subset data berbeda, proses ensemble menghasilkan prediksi yang lebih stabil dan tahan terhadap *overfitting*. misalkan menggunakan 400 tree, setelah proses training selesai hasil evaluasi akhir adalah hasil final 1 kali training Evaluasi dilakukan pada data uji menggunakan metrik yang sama dengan MLR.

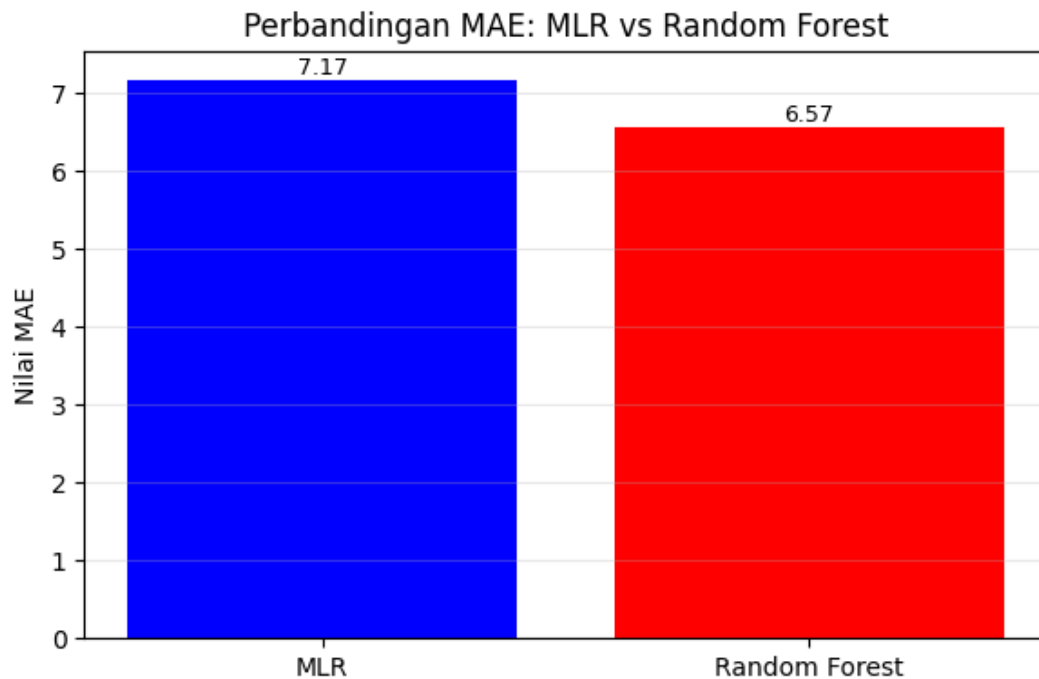


Gambar 4.1 Prediksi vs Aktual Random Forest

Dari gambar 4.1 diatas memperlihatkan grafik hasil dari pelatihan model menggunakan *Random Forest* dan *Multiple Linear Regression* untuk memprediksi jumlah pengeluaran obat yang dikeluarkan oleh Apotek Bambuan setiap harinya, dapat terlihat bahwa grafik mengikuti pola jumlah penjualan obat dengan baik walaupun terdapat selisih yang tidak terlalu jauh antara nilai aktual dan prediksi namun nilai yang dihasilkan kedua model sudah hampir mendekati nilai aktual dapat terlihat pada contoh dari gambar 4.1 dimana selisih yang didapatkan hanya sekitar 8 obat dari aktual yang mana jika melihat jumlah obat yang dikeluarkan toko perharinya nilai selisih yang dihasilkan sudah cukup bagus dikarenakan nilai jumlah pengeluaran obat yang dikeluarkan perharinya sangat fluktuatif dimana nilai aktual dari data historis penjualan itu sendiri pada tiap hari nya bisa memiliki *range* yang cukup jauh.

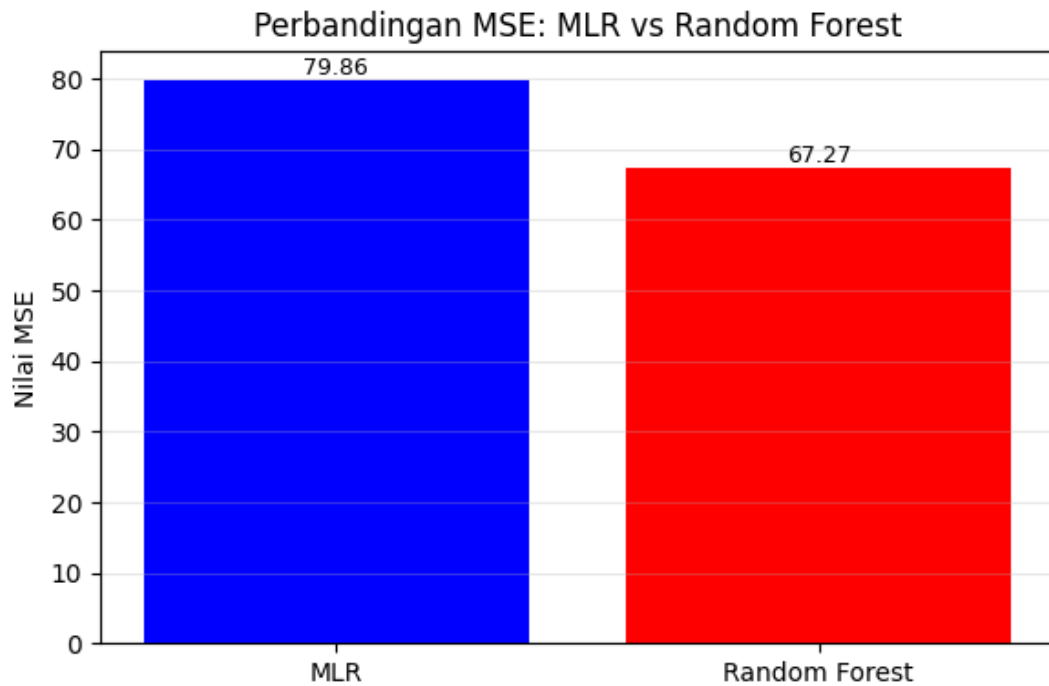
Selanjutnya hasil akhir dari kelima matriks evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R² Score*. Kelima matriks ini akan digunakan untuk mengukur seberapa besar kesalahan prediksi model terhadap data aktual. Evaluasi dilakukan terhadap hasil prediksi pada data uji (test set) untuk

masing masing model yaitu Random Forest dan Multiple Linear Regression terhadap data penjualan obat di Apotek Bambuan.



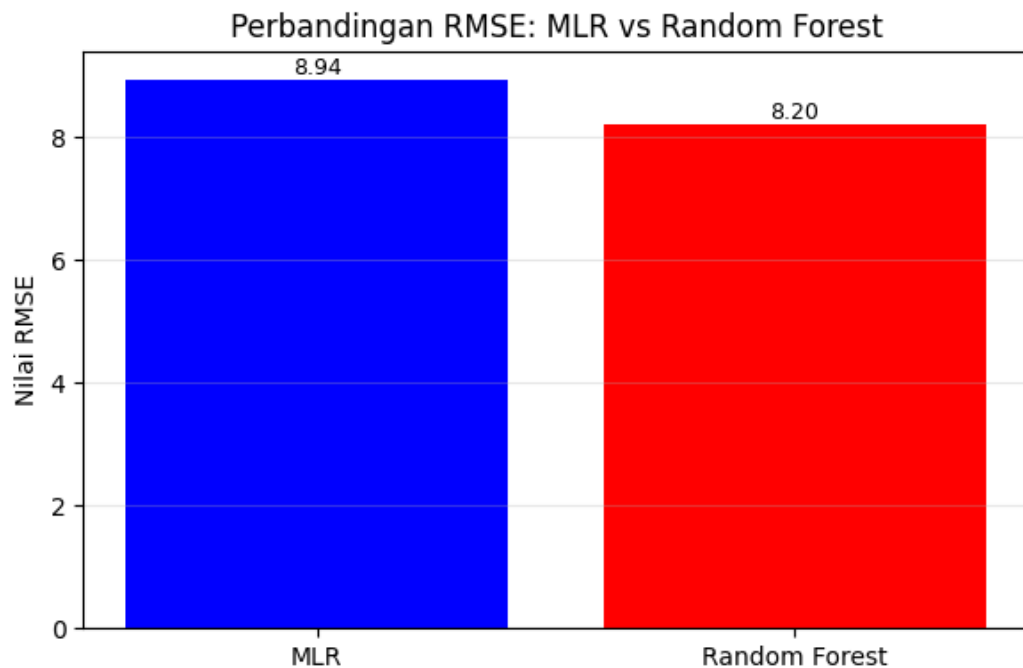
Gambar 4.2 Perbandingan matriks evaluasi MAE

Dari gambar 4.2 diatas dapat terlihat hasil akhir dari matriks evaluasi *Mean Absolute Error (MAE)* yang dihasilkan oleh kedua model yaitu *Multiple Linear Regression* dan *Random Forest* pada data penjualan obat di Apotek Bambuan, dimana hasil yang didapatkan Random Forest memiliki nilai yang lebih kecil dibandingkan *Multiple Linear Regression*, dan untuk matriks evaluasi *Mean Absoulte Error (MAE)* semakin kecil nilai yang dihasilkan maka semakin baik oleh karena itu berdasarkan hasil matriks ini dapat ditentukan bahwa *Random Forest* memiliki hasil prediksi yang lebih baik dibandingkan *Multiple Linear Regression*. Selain itu dari gambar 4.2 juga bisa didapatkan Kesimpulan rata rata kesalahan prediksi model berada pada kisaran 6-7 unit obar per harinya dimana untuk nilai tersebut menurut penulis sudah baik dikarenakan data yang dihasilkan dalam penjualan obat per harinya fluktuatif.



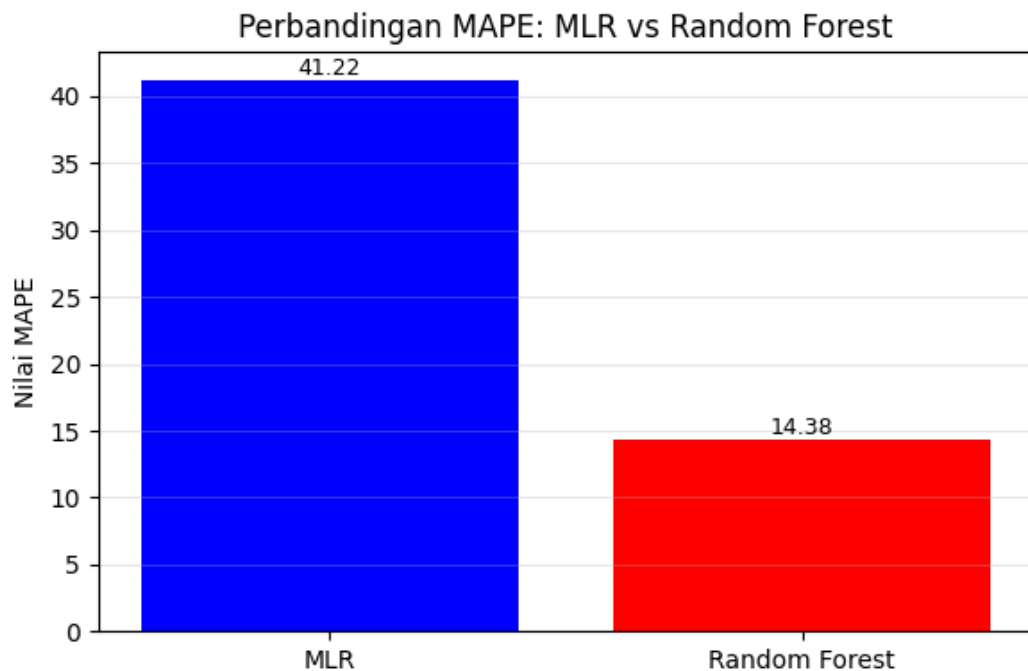
Gambar 4.3 Perbandingan matriks evaluasi MSE

Dari gambar 4.3 diatas dapat terlihat hasil akhir dari matriks evaluasi *Mean Squared Error (MSE)* yang dihasilkan oleh kedua model yaitu *Multiple Linear Regression* dan *Random Forest* pada data penjualan obat di Apotek Bambuan, dimana hasil yang didapatkan Random Forest memiliki nilai yang lebih lebih kecil dibandingkan *Multiple Linear Regression*, dan untuk matriks evaluasi *Mean Squared Error (MSE)* semakin kecil nilai yang dihasilkan maka semakin baik oleh karena itu berdasarkan hasil matriks ini dapat ditentukan bahwa *Random Forest* memiliki hasil prediksi yang lebih baik dibandingkan *Multiple Linear Regression*.



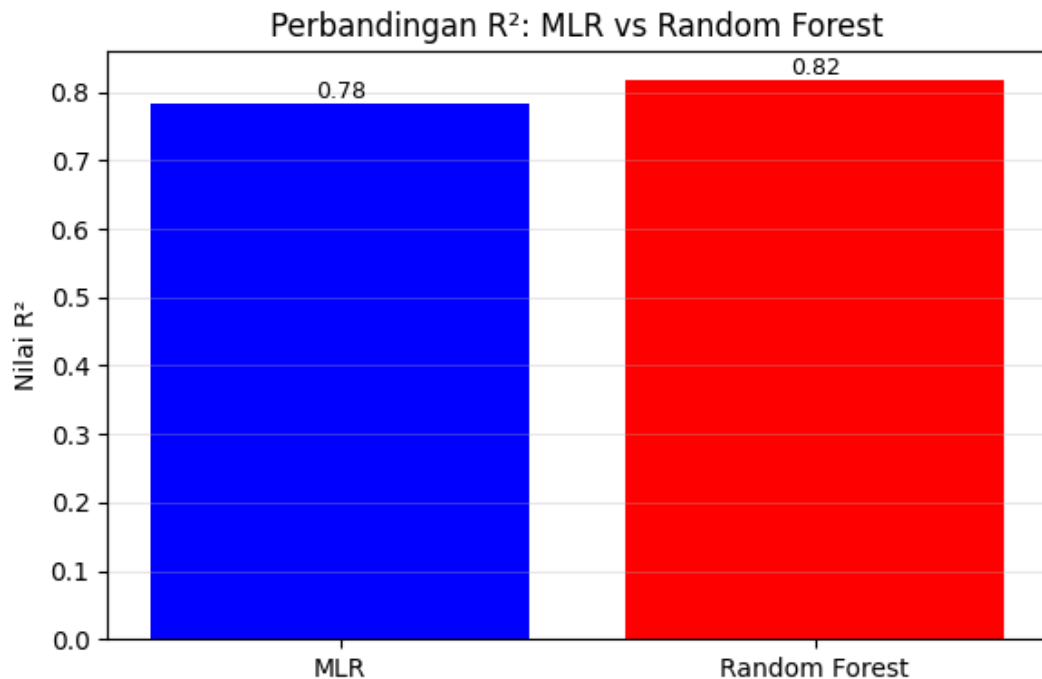
Gambar 4.4 Perbandingan matriks evaluasi RMSE

Dari gambar 4.4 diatas dapat terlihat hasil akhir dari matriks evaluasi *Root Mean Squared Error (RMSE)* yang dihasilkan oleh kedua model yaitu *Multiple Linear Regression* dan *Random Forest* pada data penjualan obat di Apotek Bambuan, dimana hasil yang didapatkan *Random Forest* memiliki nilai yang lebih kecil dibandingkan *Multiple Linear Regression*, dan untuk matriks evaluasi *Root Mean Squared Error (RMSE)* semakin kecil nilai yang dihasilkan maka semakin baik oleh karena itu berdasarkan hasil matriks ini dapat ditentukan bahwa *Random Forest* memiliki hasil prediksi yang lebih baik dibandingkan *Multiple Linear Regression*. Kemudian dari nilai *Root Mean Squared Error (RMSE)* pada gambar 4.4 diatas juga dapat diambil kesimpulan bahwa rata rata kesalahan prediksi sekitar 8 unit obat per harinya dimana pada hasil akhir ini *Random Forest* sedikit lebih unggul dibandingkan *Multiple Linear Regression*.



Gambar 4.5 Perbandingan matriks evaluasi MAPE

Dari gambar 4.5 diatas dapat terlihat hasil akhir dari matriks evaluasi *Mean Absolute Percentage Error (MAPE)* yang dihasilkan oleh kedua model yaitu *Multiple Linear Regression* dan *Random Forest* pada data penjualan obat di Apotek Bambuan, dimana hasil yang didapatkan *Random Forest* memiliki nilai yang lebih kecil dibandingkan *Multiple Linear Regression*, dan untuk matriks evaluasi *Mean Absolute Percentage Error (MAPE)* semakin kecil nilai yang dihasilkan maka semakin baik oleh karena itu berdasarkan hasil matriks ini dapat ditentukan bahwa *Random Forest* memiliki hasil prediksi yang lebih baik dibandingkan *Multiple Linear Regression*. Kemudian dari nilai *Mean Absolute Percentage Error (MAPE)* pada gambar 4.5 diatas juga dapat diambil kesimpulan bahwa *Multiple Linear Regression* rata rata meleset 41% dari nilai aktual dan *Random Forest* meleset dengan nilai 14% dari nilai aktual.



Gambar 4.6 Perbandingan matriks evaluasi

Dari gambar 4.5 diatas dapat terlihat hasil akhir dari matriks evaluasi R^2 Score yang dihasilkan oleh kedua model yaitu Multiple Linear Regression dan *Random Forest* pada data penjualan obat di Apotek Bambuan, dimana hasil yang didapatkan *Random Forest* memiliki nilai yang lebih besar dibandingkan *Multiple Linear Regression*, dan untuk matriks evaluasi R^2 Score semakin besar nilai yang dihasilkan maka semakin baik oleh karena itu berdasarkan hasil matriks ini dapat ditentukan bahwa *Random Forest* memiliki hasil prediksi yang lebih baik dibandingkan *Multiple Linear Regression*. Berdasarkan hasil dari gambar 4.5 diatas.

Secara keseluruhan, eksperimen baseline ini menunjukkan bahwa *Random Forest* memiliki keunggulan dibandingkan *Multiple Linear Regression* dalam konteks prediksi penjualan obat harian di Apotek Bambuan. Namun, MLR tetap bermanfaat karena mampu memberikan interpretasi langsung berupa kontribusi tiap

variabel terhadap target. Hal ini menjadikannya relevan untuk analisis penjelasan (*explainability*), sementara Random Forest lebih unggul untuk akurasi prediksi.

4.4. Analisis Pengaruh Jumlah Data Historis

Analisis pengaruh jumlah data historis 1 tahun, 2 tahun, dan 3 tahun dilakukan untuk membandingkan pengaruh dari jumlah data terhadap matriks evaluasi yang didapatkan oleh masing masing model serta seberapa cepat model memahami pola yang ada.

4.4.1 Data Historis 1 Tahun

Perbandingan pertama menggunakan data historis 1 tahun dimana membandingkan hasil kedua model berdasarkan data 1 tahun dimana hasil akhir dari kedua model akan ditampilkan.

Tabel 4.5 Perbandingan Data 1 Tahun Terhadap Kedua Model

Model	MAE	MAPE	MSE	RMSE	R ²
MLR	7.385	16.050%	105.396	10.266	0.099
RFR	6.482	14.437%	75.331	8.679	0.356

Berdasarkan tabel diatas didapatkan kesimpulan bahwa model Random Forest lebih cepat memahami model berdasarkan hasil akhir matriks evaluasi yaitu *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (Mape)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)* dan *R² Score* dimana hasil akhir dari kelima matriks evaluasi tersebut yaitu untuk Random Forest Regression mendapatkan nilai *Mean Absolute Error (MAE)* dengan nilai 6.482, *Mean Absolute Percentage Error (Mape)* 14.437%, *Mean Squared Error (MSE)* dengan nilai 75.331, *Root Mean Squared Error (RMSE)* dengan nilai 8.679 dan *R² Score* dengan nilai 0.356, dan untuk model *Multiple Linear Regression* pada matriks evaluasi

Mean Absolute Error (MAE) dengan nilai 7.385, *Mean Absolute Percentage Error (Mape)* 16.050%, *Mean Squared Error (MSE)* dengan nilai 105.396, *Root Mean Squared Error (RMSE)* dengan nilai 10.266 dan *R² Score* dengan nilai 0.099 walau hasil akhir yang didapatkan belum cukup memuaskan untuk kedua model dikarenakan hanya berdasarkan data 1 tahun namun berdasarkan hasil kedua model tersebut dapat disimpulkan bahwa Random Forest memiliki keunggulan untuk memprediksi penjualan obat berdasarkan data historis 1 tahun.

4.4.2 Data Historis 2 Tahun

Perbandingan Kedua menggunakan data historis berdasarkan matriks evaluasi untuk kedua model pada dataset periode 2 dimana hasil akhir dari kedua model akan ditampilkan pada tabel 4.4 dibawah.

Tabel 4.6 Perbandingan Data 2 Tahun Terhadap Kedua Model

Model	MAE	MAPE	MSE	RMSE	R ²
MLR	7.172	41.218%	79.855	8.936	0.784
RFR	6.569	14.382%	67.268	8.202	0.818

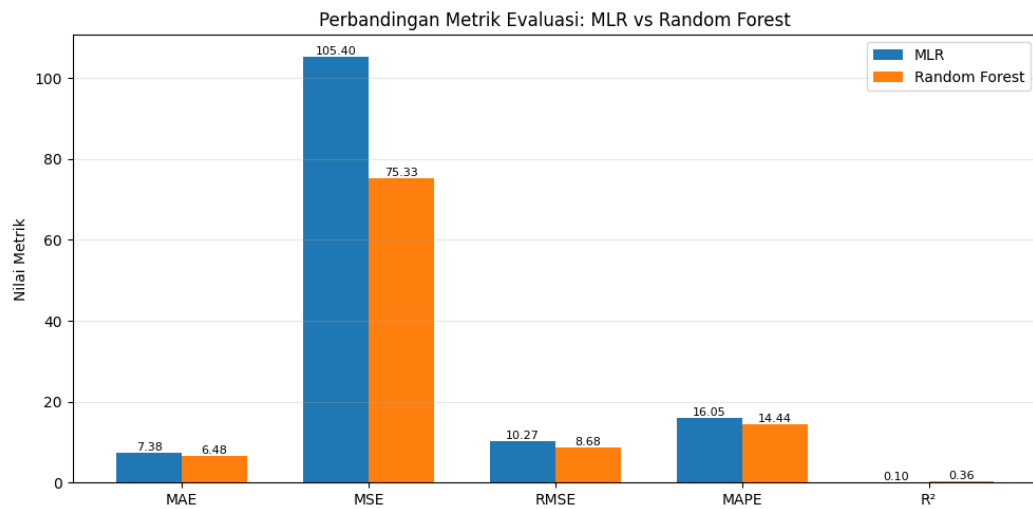
Berdasarkan tabel diatas didapatkan kesimpulan bahwa model Random Forest lebih cepat memahami model berdasarkan hasil akhir matriks evaluasi *Mean Absolute Error (MAE)* dengan nilai 6.569, *Mean Absolute Percentage Error (Mape)* 14.382%, *Mean Squared Error (MSE)* dengan nilai 67.268, *Root Mean Squared Error (RMSE)* dengan nilai 8.202 dan *R² Score* dengan nilai 0.818 dan Multiple Linear Regression hanya mendapatkan hasil selalu kurang baik dibandingkan Random Forest yaitu *Mean Absolute Error (MAE)* dengan nilai 7.172, *Mean Absolute Percentage Error (Mape)* 14.382%, *Mean Squared Error (MSE)* dengan nilai 79.855, *Root Mean Squared Error (RMSE)* dengan nilai 8.936

dan R^2 Score dengan nilai 0.784 walau hasil akhir yang didapatkan oleh *Multiple Linear Regression* belum sebaik *Random Forest* dalam hal akurasi prediksi, model ini tetap memiliki keunggulan dari sisi interpretasi. Hal ini terlihat pada koefisien regresi (β) dari setiap variabel independen, yang dapat digunakan untuk mengidentifikasi variabel mana yang memberikan pengaruh paling besar terhadap jumlah penjualan obat. Dengan demikian, MLR lebih transparan dan mudah dijelaskan dibandingkan model non-linier.

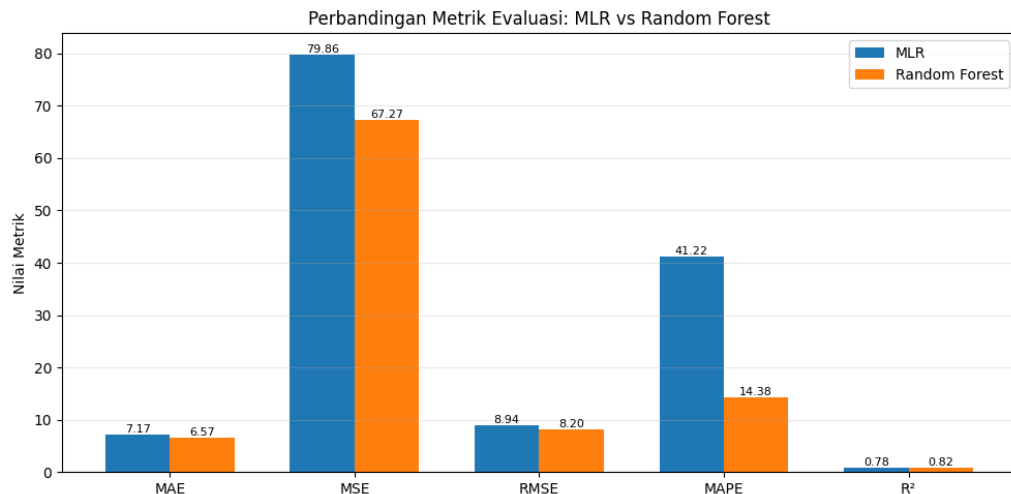
4.5. Pembahasan Hasil Penelitian

Berdasarkan visualisasi dan evaluasi performa yang telah dilakukan, baik model *Random Forest* maupun *Multiple Linear Regression* menunjukkan kemampuan yang baik dalam menangkap pola temporal dari data penjualan Apotek Bambuan. Model *Random Forest* dapat dengan cepat beradaptasi dan memahami pola data lebih cepat dibandingkan *Multiple Linear Regression* contoh pada data pola historis penjualan 1 tahun dimana *Random Forest* dapat memahami lebih cepat dibanding *Multiple Linear Regression* jika kita membandingkan berdasarkan hasil matriks evaluasi *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (Mape)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)* dan R^2 Score walaupun hasil yang didapatkan masih belum cukup untuk data historis penjualan 1 tahun, hal ini tercemin dari nilai matriks evaluasi yang didapatkan untuk R^2 Score pada *Random Forest* yaitu sekitar 0.3 dan *Multiple Linear Regression* 0.099. Faktor yang mempengaruhi prediksi berdasarkan hasil sebelumnya yaitu jumlah data yang digunakan untuk melatih kedua model, kedua model ketika menggunakan data 1 tahun kurang dapat memahami pola yang

terdapat pada data penjualan sedangkan ketika data yang digunakan bertambah 1 tahun lagi, kedua model dapat memahami lebih baik.



Gambar 4.7 Perbandingan Keseluruhan Matriks Data 1 Tahun



Gambar 4.8 Perbandingan Keseluruhan Matriks Data 2 Tahun

Dapat dilihat pada gambar 4.7 dan 4.8 perbandingan hasil matriks evaluasi dari kedua data yaitu penggunaan data 1 tahun dan data 2 tahun dimana hasil akhir matriks evaluasi yang didapatkan ternyata memiliki hubungan yang besar terhadap jumlah data yang digunakan contohnya nilai akhir nilai model *Random Forest* pada data 2 tahun yaitu *Mean Absolute Error (MAE)* dengan nilai 6.569, *Mean Absolute Percentage Error (Mape)* 14.382%, *Mean Squared Error (MSE)* dengan nilai

67.268, *Root Mean Squared Error (RMSE)* dengan nilai 8.202 dan *R² Score* dengan nilai 0.818, sedangkan yang didapatkan jika hanya menggunakan data 1 tahun *Mean Absolute Error (MAE)* dengan nilai 6.482, *Mean Absolute Percentage Error (Mape)* 14.437%, *Mean Squared Error (MSE)* dengan nilai 75.331, *Root Mean Squared Error (RMSE)* dengan nilai 8.679 dan *R² Score* dengan nilai 0.356, begitu pula pada model *Multiple Linear Regression* yaitu pada data 2 tahun nilai akhir matriks evaluasi yang dihasilkan memiliki nilai *Mean Absolute Error (MAE)* dengan nilai 7.172, *Mean Absolute Percentage Error (Mape)* 14.382%, *Mean Squared Error (MSE)* dengan nilai 79.855, *Root Mean Squared Error (RMSE)* dengan nilai 8.936 dan *R² Score* dengan nilai 0.784 sedangkan pada data 1 tahun yang didapatkan hanya memiliki nilai *Mean Absolute Error (MAE)* dengan nilai 7.385, *Mean Absolute Percentage Error (MAPE)* 16.050%, *Mean Squared Error (MSE)* dengan nilai 105.396, *Root Mean Squared Error (RMSE)* dengan nilai 10.266 dan *R² Score* dengan nilai 0.099.

Berdasarkan data-data diatas dapat diambil kesimpulan bahwa *Random Forest* menghasilkan prediksi yang lebih baik dibandingkan *Multiple Linear Regression* baik dengan menggunakan data penjualan 1 tahun ataupun 2 tahun berdasarkan perbandingan hasil matriks evaluasi *Mean Absolute Error (MAE)* dengan nilai 6.57, *Mean Absolute Percentage Error (Mape)* dengan nilai 14,38%, *Mean Squared Error (MSE)* dengan nilai 67.27, *Root Mean Squared Error (RMSE)* dengan nilai 8.20 dan *R² Score* dengan nilai 0.82 pada studi kasus jumlah pengeluaran obat harian di Apotek Bambuan sehingga *Random Forest* unggul dalam semua aspek hasil akhir matriks evaluasi.

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini bertujuan untuk membandingkan tingkat akurasi Random Forest dan Multiple Linear Regression dalam memprediksi pengeluaran obat harian pada Apotek Bambuan. Berdasarkan hasil eksperimen, evaluasi metrik (MSE, RMSE, dan R^2 Score), serta analisis visualisasi hasil prediksi, diperoleh beberapa temuan utama sebagai berikut:

1. Model Random Forest menunjukkan tingkat akurasi yang lebih tinggi dibandingkan Multiple Linear Regression ketika melakukan test pada data 1 tahun dimana mencatat nilai RMSE sebesar 75.331, MAE sebesar 6.482 dan R^2 Score sebesar 0.356 yang mana lebih baik dibandingkan Multiple Linear Regression yang hanya mencatatkan nilai RMSE sebesar 105.396, MAE sebesar 7.385 dan R^2 Score sebesar 0.099 yang mana mencatatkan nilai kurang dari yang dihasilkan oleh Random Forest, selisih nilai matriks evaluasi ini menunjukkan bahwa Random Forest dapat mengenali pola data historis penjualan lebih cepat. Lalu pada data 2 historis penjualan 2 tahun dimana kedua model mencatat nilai selisih yang tidak terlalu jauh yaitu RMSE sebesar 74.229, MAE sebesar 6.871 dan R^2 Score sebesar 0.820 yang mana hanya selisih sedikit dibandingkan Multiple Linear Regression yang mencatatkan nilai RMSE sebesar 72.920, MAE sebesar 6.923 dan R^2 Score sebesar 0.784.
2. Keunggulan Random Forest yang mampu membaca pola data lebih cepat dibandingkan Multiple Linear Regression dikarenakan random forest dapat

memahami pola data pada data yang sedikit sehingga apabila data yang digunakan sedikit maka direkomendasikan menggunakan Random Forest.

3. Visualisasi hasil prediksi menunjukkan bahwa kurva prediksi yang dihasilkan oleh kedua apabila menggunakan data yang mencukupi maka kurva yang terbentuk antara nilai actual dan nilai predicted terbilang bagus dan untuk Random Forest ketika ada nilai yang terjadi perubahan yang tajam atau berbeda dari biasanya dapat beradaptasi dengan cepat yang dapat dilihat pada gambar visualisasi prediksi yang dihasilkan keduanya.
4. Secara keseluruhan, penelitian ini membuktikan bahwa jumlah data yang digunakan akan mempengaruhi hasil dari prediksi kedua model yang mana akan sangat berpengaruh terhadap model Multiple Linear Regression sedangkan Random Forest dapat beradaptasi dengan sangat baik walaupun hasil yang diberikan belum cukup baik, dan ketika kedua model diberikan jumlah data yang mencukupi maka kedua model mampu menangkap pola – pola yang terdapat pada data penjualan.

5.2 Saran

Adapun saran yang kiranya dapat dilakukan pada penelitian selanjutnya adalah sebagai berikut :

1. Menggunakan lebih banyak data historis dengan target nilai y yang lebih spesifik misalkan per obat contohnya paracetamol namun untuk melakukan penelitian ini diperlukan dilakukan di Apotek besar atau bahkan di perusahaan industri farmasi yang mana mengeluarkan obat per 1 macam misalkan paracetamol sehingga data yang didapatkan mencukupi.

2. Model yang dikembangkan serta struktur dataset yang dibuat pada penelitian ini dapat dijadikan acuan awal oleh individu yang ingin melakukan analisis prediktif terhadap perkiraan penjualan obat.
3. Diharapkan penelitian ini dapat menjadi dasar pengembangan sistem prediksi penjualan obat yang lebih akurat dan aplikatif, serta mendorong integrasi model machine learning dalam sektor farmasi.

DAFTAR PUSTAKA

- Arifuddin, D., Kusrini, & Kusnawi. (2025). Perbandingan performansi algoritma Multiple Linear Regression dan Multi Layer Perceptron Neural Network dalam memprediksi penjualan obat. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 722–737.
- Ayuni, G. N., & Fitrianah, D. (2022). Penerapan metode regresi linier untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika*, 14(2). Institut Teknologi Harapan Bangsa, Bandung.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Efendi, M. S., & Zyen, A. K. (2024). Penerapan algoritma Random Forest untuk prediksi penjualan dan sistem persediaan produk. *Resolusi*, 5(1), 12–20.
<https://djournals.com/resolusi/article/view/2149/1156>
- Farhanuddin, F., Sihombing, S. E. K., & Yahfizham, Y. (2024). Komparasi Multiple Linear Regression dan Random Forest Regression dalam memprediksi anggaran biaya manajemen proyek sistem informasi. *Journal of Computers and Digital Business*, 3(2), 86–97.
- Fahlepi, M. R., & Widjaja, A. (2019). Penerapan Metode Multiple Linear Regression untuk Prediksi Harga Sewa Kamar Kost. *Jurnal STRATEGI*.
- Hutahaeen, M., & Handoko, K. (2022). Penerapan data mining untuk memprediksi penjualan obat di Klinik Harapan Kita Batam. *Jurnal Comasie*, 6(5).
- Josaphat, B. P., & Pangestika, Z. (2025). Predicting stock price using convolutional neural network and long short-term memory (case study: Stock of BBCA). *Journal of the Indonesian Mathematical Society*.
<https://jimsa.org/index.php/jimsa/article/view/1512>

- Kementerian Kesehatan Republik Indonesia. (2016). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 73 Tahun 2016 tentang Standar Pelayanan Kefarmasian di Apotek*. Jakarta: Kemenkes RI.
- K. Puteri, & Silvanie, A. (2020). Machine Learning untuk model prediksi harga sembako dengan metode regresi linier berganda. *Jurnal Nasional Informatika*, 1(2), 82–94.
- M. R. Rosyid, L. Mawaddah, & M. Akrom. (2024). Investigasi model machine learning regresi pada senyawa obat sebagai inhibitor korosi. *Jurnal Algoritma*, 21(1), 332–342.
- Muzakir, A., Komari, A., & Ilham, M. (2024). *Penerapan konsep machine learning & deep learning*. Asosiasi Dosen Sistem Informasi Indonesia. <https://adsii.or.id/wp-content/uploads/2024/10/Penerapan-Konsep-Machine-Learning-Deep-Learning-Ari-Muzakir-et-al.pdf>
- Pramesti, D., & Baihaqi, W. M. (2023). Perbandingan prediksi jumlah transaksi ojek online menggunakan regresi linier dan Random Forest. *Generation Journal*, 7(3–Spesial).
- Rahma, S. M., Oktadini, N. R., Tania, K. D., & Indah, D. R. (2024). Penerapan data mining untuk prediksi penjualan obat menggunakan metode K-nearest neighbor. *JOISIE (Journal of Information Systems and Informatics Engineering)*, 8(2), 259–271.
- Simamora, F. P., Purba, R., & Pasha, M. F. (2025). Optimisasi hyperparameter BiLSTM menggunakan Bayesian Optimization untuk prediksi harga saham. *Jambura Journal of Machine Learning*, 7(1). <https://ejurnal.ung.ac.id/index.php/jjom/article/view/27166>
- Supriyanto, Y., Ilhamsyah, M., & Enri, U. (2022). Prediksi harga minyak kelapa sawit menggunakan Linear Regression dan Random Forest. *Jurnal Ilmiah Wahana Pendidikan*, 8(7), 1–8. <https://doi.org/10.5281/zenodo.6559603>
- Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'Ban, S. R., Amalia, M., & Febriansyah, R. (2024). Pembelajaran *Machine Learning*. *OKTAL: Jurnal Ilmu Komputer dan Science*, 3(2), 375–380.

LAMPIRAN



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/IBAN-PT/Akred/PT/III/2019
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<http://fkip.umsu.ac.id> fkip@umsu.ac.id [umsu.fkip](https://www.facebook.com/umsu.fkip) [umsu.fkip](https://www.instagram.com/umsu.fkip) [umsu.fkip](https://www.youtube.com/channel/UC...) [umsu.fkip](https://www.linkedin.com/company/umsu.fkip)

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 577/IL3-AU/UMSU-09/F/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Teknologi Informasi
Pada tanggal : 28 April 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Muammar Farhan
NPM : 2109020097
Semester : VIII (Delapan)
Program studi : Teknologi Informasi
Judul Proposal / Skripsi : Prediksi Penjualan Obat di Apotek Bambuan Menggunakan Model ARIMA (Autoregresif Integrated Moving Average) Berbasis Website Untuk Optimalisasi Manajemen Stok

Dosen Pembimbing : Zuli Agustina Gultom, M.Si

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi dinyatakan " BATAL "** bila tidak selesai sebelum Masa Kadaluarsa tanggal : 28 April 2026
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
 Pada Tanggal : 30 Syawwal 1446 H
 28 April 2025M



Dekan

Dr. Al-Khwarizmi, M.Kom.
 NIDN. 0127099201

Cc. File





UMSU

Ingat | Cerdas | Terpercaya

UMSU Sumatera Utara
Jl. Sekeloa Timur No. 1 Medan 20136
Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 29/SK/BAN-PT/Akred/PT/06/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
Website: <http://www.umsu.ac.id> Email: info@umsu.ac.id umsu@umsu.ac.id umsu@umsu.ac.id umsu@umsu.ac.id

PERUBAHAN TOPIK/JUDUL PENELITIAN

Nomor agenda :
Tanggal persetujuan : 17 Juli 2025
Topik yang disetujui Program Studi : Prediksi Penjualan Obat di Apotek Bambuan Menggunakan Model ARIMA (Autoregresif Integrated Moving Average) Berbasis Website Untuk Optimalisasi Manajemen Stok
Nama Dosen pembimbing : Zuli Agustina Gultom, M.Si
Judul yang disetujui Dosen Pembimbing : Perbandingan Algoritma Random Forest dan Multiple Linear Regression Dalam Memprediksi Pengeluaran Obat Harian Studi Kasus Pada Apotek Bambuan

Medan 17 Juli 2025

Disahkan oleh

Ketua Program Studi

Terdapat: ~~Staf~~ Informasi

(Fatma Sari Hutagalung, M.Kom)

Persetujuan

Dosen Pembimbing

(Zuli Agustina Gultom, M.Si)



UMSU
Unggul | Cerdas | Terpercaya

Waktu yang lebih singkat untuk lebih banyak prestasi
dan kesejahteraan

MAJLIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK.BAN-PT/Akred/PT/III/2019

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

BMN 1745/UMSU/RI/18

BMN 1745/UMSU/RI/18

umsumedan

umsumedan

umsumedan

umsumedan

Berita Acara Pembimbingan Skripsi

Nama Mahasiswa : Muammar Farhan
NPM : 2109020097
Program Studi : Teknologi Informasi
Nama Dosen Pembimbing : Zuli Agustina Gultom, M.Si
Konsentrasi : Machine Learning
Judul Penelitian : Perbandingan Algoritma Random Forest dan Multiple Linear Regression Dalam Memprediksi Pengeluaran Obat Harian Studi Kasus Pada Apotek Bambuan

Item	Hasil Evaluasi	Tanggal	Paraf Dosen
1	Pergantian judul sepuluh semesta dan penambahan bab 2.	16/09/2025	
2	Perisi gambar Mengas: Pseudocode	18/09/2025	
3	Perisi tabel, Penulisan tabel data	19/09/2025	
4	Penambahan matrix MAPE dan MSE untuk matrix evaluasi	21/09/2025	
5	Penambahan data, dan perbaikan Bab 4.	23/09/2025	
6	Perbaikan Penulisan Bab 4 dan 5.	29/09/2025	
7	Acc redang skripsi	30/09/2025	

Medan, 19 Agustus 2025

Diketahui oleh:
Kepala Program Studi
Teknologi Informasi

(Fatma Sari Hutagalung S.Kom., M.Kom.)

Disetujui oleh:
Dosen Pembimbing

(Zuli Agustina Gultom, M.Si)

