

Analisis dan Penerapan Metode *Regresi Linear* dan *Decision Tree* untuk memprediksi Penjualan pada Rayyan Bakery Simpang Marbau

Analysis and Implementation of Linear Regression and Decision Tree Methods to Predict Sales at Rayyan Bakery, Simpang Marbau

¹Natari Dia Alike Hidayah*, ²Ibnu Rasyid Munthe, ³Angga Putra Juledi, ⁴Marnis Nasution

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu, Sumatera Utara, Indonesia

^{1,2,3,4}Jl. S.M. Raja No.126 A Aek Tapa Rantauprapat, Kab. Labuhanbatu, Sumatera Utara, Indonesia

*e-mail: natarigia@gmail.com

(received: 27 January 2026, revised: 6 February 2026, accepted: 7 February 2026)

Abstrak

Perkembangan teknologi informasi dan analitik data mendorong pelaku usaha untuk memanfaatkan data historis sebagai dasar pengambilan keputusan bisnis. Pada sektor usaha kecil dan menengah (UKM), khususnya di bidang kuliner, kemampuan memprediksi penjualan menjadi aspek penting dalam perencanaan produksi dan pengelolaan stok agar operasional usaha dapat berjalan secara efisien. Rayyan Bakery Simpang Marbau sebagai UKM bakery menghadapi permasalahan fluktuasi penjualan yang selama ini masih dikelola berdasarkan pengalaman tanpa analisis data yang sistematis. Permasalahan utama dalam penelitian ini adalah belum adanya penerapan metode prediksi penjualan berbasis data mining yang dapat membantu pemilik usaha dalam memperkirakan jumlah penjualan secara akurat. Oleh karena itu, diperlukan pendekatan prediksi yang mampu memanfaatkan data historis penjualan untuk mendukung pengambilan keputusan manajerial. Metode yang digunakan dalam penelitian ini adalah regresi linear dan *decision tree*. Data yang dianalisis berupa data penjualan Rayyan Bakery Simpang Marbau dalam periode tertentu. *Regresi linear* digunakan untuk memodelkan hubungan linier antara variabel penjualan, sedangkan *decision tree* digunakan untuk menangkap pola *non-linier* serta menghasilkan aturan keputusan yang mudah diinterpretasikan. Kinerja kedua metode kemudian dianalisis dan dibandingkan berdasarkan hasil prediksi yang dihasilkan. Hasil penelitian menunjukkan bahwa baik metode regresi linear maupun *decision tree* mampu digunakan untuk memprediksi penjualan, namun *decision tree* memberikan hasil yang lebih fleksibel dalam menangkap pola penjualan yang berfluktuasi. Temuan ini diharapkan dapat membantu Rayyan Bakery dalam perencanaan produksi dan pengelolaan stok, serta menjadi referensi penerapan metode prediksi penjualan pada UKM sejenis.

Kata kunci: analisis, *regresi linier*, *decision tree*, prediksi, penjualan, bakery

Abstract

The development of information technology and data analytics has encouraged business actors to leverage historical data as a basis for decision-making. In the small and medium enterprise (SME) sector, particularly in the culinary field, the ability to predict sales is a crucial aspect of production planning and stock management to ensure operational efficiency. Rayyan Bakery Simpang Marbau, as a bakery SME, faces challenges due to fluctuating sales that have traditionally been managed based on experience rather than systematic data analysis. The main problem addressed in this study is the absence of a data-driven sales prediction method that can assist the business owner in estimating sales accurately. Therefore, a predictive approach that utilizes historical sales data is required to support managerial decision-making. This study employs linear regression and decision tree methods. The analyzed data consist of historical sales records of Rayyan Bakery Simpang Marbau over a specific period. Linear regression is used to model the linear relationship between sales variables, while the decision tree captures non-linear patterns and produces easily interpretable decision rules. The performance of both methods is analyzed and compared based on the accuracy of

<http://sistemasi.ftik.unisi.ac.id>

the predictions they generate. The results indicate that both linear regression and decision tree methods can be effectively used to predict sales; however, the decision tree provides greater flexibility in capturing fluctuating sales patterns. These findings are expected to assist Rayyan Bakery in production planning and stock management, as well as serve as a reference for applying sales prediction methods in similar SMEs.

Keywords: *analysis, linear regression, decision tree, prediction, sales, bakery*

1 Pendahuluan

Perkembangan teknologi informasi yang semakin pesat telah mendorong terjadinya transformasi digital di berbagai sektor, termasuk sektor bisnis dan ekonomi [1]. Pemanfaatan data sebagai aset strategis menjadi salah satu faktor kunci dalam meningkatkan efektivitas pengambilan keputusan [2]. Seiring dengan meningkatnya volume data yang dihasilkan dari aktivitas operasional, pendekatan berbasis analitik data dan kecerdasan buatan mulai banyak diterapkan untuk mendukung proses perencanaan, evaluasi, serta peramalan dalam dunia usaha [3]. Salah satu bentuk pemanfaatan data yang banyak digunakan adalah prediksi penjualan, yang berperan penting dalam menjaga keseimbangan antara permintaan pasar dan ketersediaan produk [4].

Pada sektor Usaha Kecil dan Menengah (UKM), kemampuan dalam melakukan prediksi penjualan menjadi semakin penting mengingat keterbatasan sumber daya yang dimiliki [5]. UKM dituntut untuk mampu mengelola produksi dan persediaan secara efisien agar dapat bersaing dan mempertahankan keberlangsungan usaha [6]. Ketidaktepatan dalam memperkirakan jumlah penjualan dapat menyebabkan berbagai permasalahan, seperti kelebihan stok yang berujung pada pemborosan biaya produksi, maupun kekurangan stok yang berpotensi menurunkan tingkat kepuasan pelanggan [7]. Oleh karena itu, penerapan teknologi informasi dan analisis data menjadi solusi yang relevan bagi UKM untuk meningkatkan efisiensi operasional dan kualitas pengambilan keputusan [8].

Salah satu pendekatan yang dapat digunakan untuk memprediksi penjualan adalah penerapan metode data mining dan machine learning [9]. Metode ini memungkinkan pemanfaatan data historis untuk menemukan pola, tren, dan hubungan antar variabel yang sebelumnya sulit diidentifikasi secara manual [10]. *Regresi linear* merupakan metode statistik yang umum digunakan untuk memodelkan hubungan linier antara variabel dependen dan variabel independen, sehingga dapat memberikan gambaran kuantitatif mengenai pengaruh faktor-faktor tertentu terhadap penjualan [11]. Di sisi lain, *decision tree* merupakan metode yang mampu menangkap pola hubungan non-linier dan menghasilkan model prediksi yang bersifat interpretatif melalui aturan keputusan berbentuk pohon, sehingga mudah dipahami oleh pengguna non-teknis.

Rayyan Bakery Simpang Marbau sebagai salah satu UKM di bidang kuliner menghadapi permasalahan fluktuasi penjualan dari waktu ke waktu. Selama ini, proses perencanaan produksi dan pengelolaan stok masih dilakukan berdasarkan pengalaman dan intuisi pemilik usaha, tanpa didukung oleh analisis data yang sistematis. Kondisi tersebut berpotensi menimbulkan ketidakefisienan operasional, seperti ketidaksesuaian jumlah produksi dengan permintaan aktual pasar. Padahal, Rayyan Bakery memiliki data historis penjualan yang dapat dimanfaatkan sebagai dasar dalam membangun model prediksi penjualan yang lebih akurat dan objektif.

Berdasarkan permasalahan tersebut, diperlukan suatu pendekatan prediksi penjualan berbasis data yang mampu membantu pemilik usaha dalam memperkirakan jumlah penjualan secara lebih tepat [12]. Penelitian ini mengusulkan penerapan dan perbandingan metode regresi linear dan decision tree untuk memprediksi penjualan pada Rayyan Bakery Simpang Marbau. Pemilihan kedua metode ini didasarkan pada karakteristiknya yang saling melengkapi, di mana regresi linear unggul dalam memodelkan hubungan linier, sementara decision tree lebih fleksibel dalam menangani pola data yang kompleks dan non-linier. Dengan membandingkan kinerja kedua metode, diharapkan dapat diperoleh model prediksi yang paling sesuai untuk mendukung pengambilan keputusan manajerial berbasis data pada UKM bakery.

Dengan demikian, penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja metode regresi linear dan decision tree dalam memprediksi penjualan, serta menentukan metode yang paling sesuai untuk mendukung pengambilan keputusan manajerial di Rayyan Bakery. Hasil penelitian diharapkan tidak hanya memberikan kontribusi akademis dalam penerapan metode prediksi

penjualan pada UKM, tetapi juga memberikan manfaat praktis bagi pelaku usaha dalam meningkatkan efisiensi operasional dan perencanaan bisnis berbasis data.

2 Tinjauan Literatur

Penerapan algoritma Regresi Linear dalam meramalkan penjualan mobil Toyota di industri otomotif Indonesia. Dengan menggunakan data penjualan dari tahun 2018 hingga 2023, model Regresi Linear berhasil memprediksi penjualan mobil Toyota untuk beberapa bulan ke depan. Hasil prediksi menunjukkan angka yang cukup dekat dengan penjualan aktual untuk bulan Januari, namun pada bulan Februari, Maret, dan April 2024 menunjukkan adanya perbedaan signifikan antara prediksi dan realisasi, yang mungkin disebabkan oleh faktor-faktor musiman atau kejadian khusus yang tidak tertangkap oleh model [13].

Algoritma regresi linear untuk memprediksi penjualan produk kategori musik dan kerajinan di toko Agung0na9, dilakukan beberapa proses tahapan KDD (*Knowledge Discovery In Data Base*) dan didapatkan hasil untuk memprediksi 3 bulan selanjutnya yaitu bulan pertama musik 68 pcs, bulan dua 69 dan bulan tiga 70 pcs untuk kategori kerajinan bulan pertama 1078, bulan kedua 1029 dan bulan ketiga 1066. dengan nilai evaluasi untuk kategori musik 13% dan kerajinan 22,65% yang bisa dikatakan cukup atau bisa digunakan hasil dari persamaan regresi linear, karena hasil evaluasi Relative Error cukup mendekati [14].

Algoritma Decision Tree adalah model prediksi yang paling efektif untuk memperkirakan jumlah pendaftar PPDB di SMKN3 Metro. Dengan nilai *Mean Squared Error* (MSE) terendah, *Root Mean Squared Error* (RMSE) terendah, *Mean Absolute Error* (MAE) terendah, serta koefisien determinasi (R^2) tertinggi, model ini memberikan akurasi yang sangat baik dibandingkan dengan model lainnya, termasuk *Random Forest*, *k-Nearest Neighbors* (kNN), *Support Vector Machine* (SVM), dan Neural Network. Decision Tree juga mampu mendekati data real untuk kompetensi utama seperti Rekayasa Perangkat Lunak dan Teknik Komputer & Jaringan, menunjukkan kemampuannya dalam memprediksi jumlah pendaftar dengan baik. Implikasi dari penelitian ini adalah bahwa

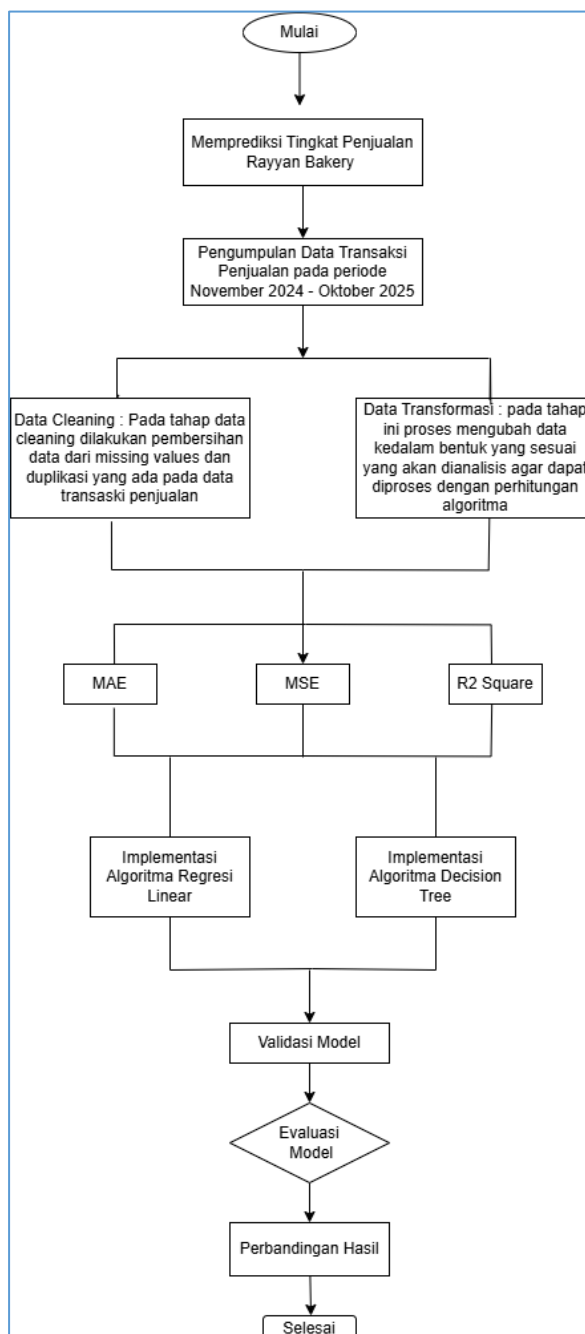
SMKN3 Metro dapat memanfaatkan model *Decision Tree* untuk merencanakan kapasitas penerimaan siswa dan meningkatkan efisiensi pengelolaan PPDB. Dengan mengintegrasikan model ini ke dalam sistem manajemen sekolah, proses perencanaan dan alokasi sumber daya dapat dilakukan dengan lebih akurat dan strategis. Selain itu, penelitian ini menyoroti perlunya evaluasi berkala dan pengembangan berkelanjutan dari model prediksi untuk memastikan keakuratan hasil sesuai dengan dinamika dan perubahan tren pendaftar [15].

Keterbaruan penelitian ini terletak pada integrasi antara metode regresi linear dan decision tree dalam konteks UKM bakery. Pendekatan ini tidak hanya memberikan hasil prediksi yang lebih akurat, tetapi juga menyediakan dasar ilmiah untuk pengambilan keputusan strategis berbasis data yang relevan dengan karakteristik usaha kecil menengah di Indonesia.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data historis penjualan sebagai sumber utama analisis. Pendekatan kuantitatif dipilih karena mampu memberikan gambaran numerik yang jelas mengenai hubungan antar variabel serta memungkinkan pengukuran kinerja model prediksi secara objektif melalui metrik evaluasi tertentu [16]. Data historis penjualan diolah dan dianalisis menggunakan metode data mining untuk menemukan pola dan tren yang dapat dimanfaatkan dalam proses peramalan penjualan.

Selanjutnya, tahapan penelitian yang dilakukan disajikan dalam bentuk alur penelitian yang mencakup proses identifikasi masalah, pengumpulan data, pembersihan data, transformasi data, pembangunan model, evaluasi model, serta perbandingan hasil dari kedua metode yang digunakan. Berikut gambar 1 merupakan alur dari tahapan penelitian yang dilakukan:



Gambar 1 Alur penelitian

3.1 Mulai

Penelitian dimulai dengan mengidentifikasi kebutuhan untuk memprediksi penjualan Rayyan Bakery menggunakan data historis penjualan.

3.2 Identifikasi Masalah

Rayyan Bakery Simpang Marbau prediksi penjualan yang lebih akurat untuk mencegah kelebihan atau kekurangan stok. Dengan memanfaatkan data historis dan metode *Regresi Linear* serta *Decision Tree*, pola penjualan dapat dianalisis sehingga keputusan produksi menjadi lebih efisien dan menguntungkan.

3.3 Pengumpulan Data

Tahap pertama adalah pengumpulan data transaksi penjualan pada periode November 2024 hingga Oktober 2025. Data transaksi ini akan menjadi dasar bagi pembuatan model prediksi penjualan

3.4 Pembersihan Data

Pada tahap ini, dilakukan pembersihan data dengan cara mengatasi nilai yang hilang (missing values) dan menghapus duplikasi data. Pembersihan data sangat penting untuk memastikan bahwa data yang digunakan untuk pemodelan adalah akurat dan lengkap.

3.5 Transformasi Data

Setelah data dibersihkan, tahap ini mengubah data ke dalam bentuk yang sesuai untuk analisis lebih lanjut. Data yang telah ditransformasi akan siap untuk diproses oleh algoritma yang digunakan.

3.6 Metrik Evaluasi Model

Tiga metrik yang digunakan untuk mengevaluasi kinerja model adalah MAE (*Mean Absolute Error*), MSE (*Mean Squared Error*), dan R2 *Square*. Metrik-metrik ini digunakan untuk mengukur akurasi dan kemampuan prediksi model.

3.7 Implementasi Algoritma

- Regresi Linear: Pada tahap ini, algoritma regresi linear diterapkan untuk mengetahui hubungan antara variabel independen dengan prediksi penjualan.
- Decision Tree: Algoritma decision tree diterapkan untuk memprediksi penjualan dengan membuat keputusan berdasarkan berbagai fitur data yang ada.

3.8 Validasi Model

Setelah model dibangun, dilakukan validasi untuk memastikan bahwa model dapat bekerja dengan baik dan menghasilkan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya.

3.9 Evaluasi Model

Model yang telah divalidasi akan dievaluasi dengan menggunakan metrik-metrik yang telah disebutkan (MAE, MSE, R2) untuk menilai seberapa efektif model dalam memprediksi penjualan.

3.10 Perbandingan Hasil

Setelah evaluasi, hasil dari kedua model (*Regresi Linear dan Decision Tree*) dibandingkan untuk melihat model mana yang memberikan kinerja terbaik dalam memprediksi penjualan di Rayyan Bakery.

3.11 Selesai

Tahap terakhir adalah menarik kesimpulan dari hasil perbandingan model dan memberikan rekomendasi berdasarkan model yang memiliki kinerja terbaik.

4 Hasil dan Pembahasan

4.1 Transformasi Data

Transformasi data dilakukan untuk menyiapkan dataset agar sesuai dengan kebutuhan algoritma. Transformasi yang digunakan meliputi: (1) pembentukan variabel target dan variabel prediktor; (2) penyesuaian tipe data; dan (3) normalisasi Min-Max jika diperlukan untuk menguji dampak perbedaan skala fitur pada model regresi.

4.1.1 Transformasi Skala Pendapatan

Pendapatan ditransformasikan ke dalam satuan jutaan rupiah untuk menyederhanakan perhitungan dan menghindari nilai numerik yang terlalu besar dengan persamaan 1:

$$X_2 = \frac{X_2}{1.000.000} \quad (1)$$

Setelah dilakukan transformasi maka hasilnya dapat dilihat pada tabel 1 berikut:

Tabel 1 Atribut data penelitian	
Pendapatan (Rp)	Transformasi
Rp5.200.000	5,2
Rp5.950.000	5,95
Rp5.950.000	5,95
Rp4.900.000	4,9
Rp5.800.000	5,8
Rp6.100.000	6,1

<http://sistemasi.ftik.unisi.ac.id>

Pendapatan (Rp)	Transformasi
Rp6.750.000	6,75
Rp7.155.000	7,155
Rp5.950.000	5,95
Rp5.155.000	5,155
Rp5.100.000	5,1
Rp7.750.000	7,75
Rp5.150.000	5,15
Rp7.900.000	7,9
Rp5.250.000	5,25
Rp5.110.000	5,11
Rp5.110.000	5,11
Rp5.155.000	5,155
Rp8.150.000	8,15
Rp8.525.000	8,525
Rp7.000.000	7
Rp5.455.000	5,455
Rp5.310.000	5,31
Rp5.285.000	5,285
Rp6.950.000	6,95
Rp6.735.000	6,735
Rp6.870.000	6,87
Rp5.370.000	5,37
Rp6.810.000	6,81
Rp8.710.000	8,71
Rp5.550.000	5,55
Rp5.185.000	5,185
Rp4.990.000	4,99
Rp5.870.000	5,87
Rp6.820.000	6,82
Rp6.750.000	6,75
Rp6.595.000	6,595
Rp5.390.000	5,39
Rp6.885.000	6,885
Rp5.900.000	5,9
Rp8.790.000	8,79
Rp7.100.000	7,1
Rp6.770.000	6,77
Rp6.600.000	6,6
Rp5.000.000	5
Rp5.240.000	5,24
Rp6.850.000	6,85
Rp5.450.000	5,45
Rp6.900.000	6,9
Rp6.870.000	6,87
Rp6.890.000	6,89
Rp6.500.000	6,5

4.1.2 Pembagian Data Pelatihan dan Pengujian

Karena dataset berbentuk deret waktu mingguan, peneliti membagi data secara kronologis untuk mencegah kebocoran informasi masa depan. Sebanyak 80% data awal digunakan sebagai data pelatihan (training), sedangkan 20% data akhir digunakan sebagai data pengujian (testing). Dengan total 52 minggu, pembagian ini menghasilkan 41 minggu sebagai data pelatihan dan 11 minggu sebagai data pengujian yang mana hasilnya dapat dilihat pada Tabel 2 berikut.

Tabel 2 Data training penjualan rayyan bakery

Periode (Tanggal)	Stok (unit)	Pendapatan (Rp)	Penjualan (Unit)
07-Nov-24	450	5.200.000	390
14-Nov-24	450	5.950.000	410
21-Nov-24	440	5.950.000	410
28-Nov-24	420	4.900.000	370
05-Dec-24	450	5.800.000	408
12-Dec-24	450	6.100.000	416
19-Dec-24	500	6.750.000	435
26-Dec-24	500	7.155.000	460
02-Jan-25	500	5.950.000	410
09-Jan-25	450	5.155.000	395
16-Jan-25	450	5.100.000	395
23-Jan-25	500	7.750.000	460
30-Jan-25	450	5.150.000	380
06-Feb-25	500	7.900.000	458
13-Feb-25	450	5.250.000	400
20-Feb-25	450	5.110.000	395
27-Feb-25	420	5.110.000	390
06-Mar-25	425	5.155.000	393
13-Mar-25	600	8.150.000	470
20-Mar-25	600	8.525.000	482
27-Mar-25	600	7.000.000	450
03-Apr-25	450	5.455.000	423
10-Apr-25	450	5.310.000	410
17-Apr-25	450	5.285.000	408
24-Apr-25	450	6.950.000	448
01-May-25	450	6.735.000	430
08-May-25	450	6.870.000	447
15-May-25	450	5.370.000	398
22-May-25	460	6.810.000	435
29-May-25	500	8.710.000	473
05-Jun-25	450	5.550.000	410
12-Jun-25	450	5.185.000	387
19-Jun-25	450	4.990.000	370
26-Jun-25	450	5.870.000	381
03-Jul-25	450	6.820.000	407
10-Jul-25	450	6.750.000	398
17-Jul-25	450	6.595.000	405
24-Jul-25	450	5.390.000	395

Periode (Tanggal)	Stok (unit)	Pendapatan (Rp)	Penjualan (Unit)
31-Jul-25	450	6.885.000	419
07-Aug-25	450	5.900.000	389

Setelah menentukan data pelatihan dan data pengujian maka selanjutnya menentukan data testing. Data testing yang digunakan adalah sebanyak 12 data yang dapat dilihat pada Tabel 3 berikut:

Tabel 3 Data testing penjualan rayyan bakery

Periode (Tanggal)	Stok (unit)	Pendapatan (Rp)	Penjualan (Unit)
14-Aug-25	500	8.790.000	470
21-Aug-25	480	7.100.000	430
28-Aug-25	450	6.770.000	418
04-Sep-25	430	6.600.000	425
11-Sep-25	455	5.000.000	380
18-Sep-25	455	5.240.000	400
25-Sep-25	450	6.850.000	407
02-Oct-25	450	5.450.000	399
09-Oct-25	450	6.900.000	420
16-Oct-25	450	6.870.000	420
23-Oct-25	450	6.890.000	427
30-Oct-25	455	6.500.000	410

4.2 Pembersihan Data

Pembersihan data mencakup pemeriksaan nilai kosong, duplikasi periode, serta konsistensi format angka dan mata uang. Peneliti menyeragamkan format pendapatan ke bentuk numerik (rupiah) untuk memudahkan perhitungan statistik dan pemodelan. Jika ditemukan nilai tidak wajar, peneliti melakukan verifikasi ke sumber data sebelum memasukkan data ke proses analisis.

4.3 Dataset

Dataset menyajikan penjualan mingguan yang digunakan dalam penelitian. Dataset ini memuat tanggal periode mingguan, stok dan pendapatan dan penjualan. Berikut dataset yang digunakan dapat dilihat pada Tabel 4 berikut.

Tabel 4 Dataset penjualan periode november 2024 s/d 30 oktober 2025

Periode (Tanggal)	Stok (Unit)	Pendapatan (Rp)	Penjualan (Unit)
07-Nov-24	450	5.200.000	390
14-Nov-24	450	5.950.000	410
21-Nov-24	440	5.950.000	410
28-Nov-24	420	4.900.000	370
05-Dec-24	450	5.800.000	408
12-Dec-24	450	6.100.000	416
19-Dec-24	500	6.750.000	435
26-Dec-24	500	7.155.000	460
02-Jan-25	500	5.950.000	410
09-Jan-25	450	5.155.000	395
16-Jan-25	450	5.100.000	395
23-Jan-25	500	7.750.000	460
30-Jan-25	450	5.150.000	380
06-Feb-25	500	7.900.000	458
13-Feb-25	450	5.250.000	400

Periode (Tanggal)	Stok (Unit)	Pendapatan (Rp)	Penjualan (Unit)
20-Feb-25	450	5.110.000	395
27-Feb-25	420	5.110.000	390
06-Mar-25	425	5.155.000	393
13-Mar-25	600	8.150.000	470
20-Mar-25	600	8.525.000	482
27-Mar-25	600	7.000.000	450
03-Apr-25	450	5.455.000	423
10-Apr-25	450	5.310.000	410
17-Apr-25	450	5.285.000	408
24-Apr-25	450	6.950.000	448
01-May-25	450	6.735.000	430
08-May-25	450	6.870.000	447
15-May-25	450	5.370.000	398
22-May-25	460	6.810.000	435
29-May-25	500	8.710.000	473
05-Jun-25	450	5.550.000	410
12-Jun-25	450	5.185.000	387
19-Jun-25	450	4.990.000	370
26-Jun-25	450	5.870.000	381
03-Jul-25	450	6.820.000	407
10-Jul-25	450	6.750.000	398
17-Jul-25	450	6.595.000	405
24-Jul-25	450	5.390.000	395
31-Jul-25	450	6.885.000	419
07-Aug-25	450	5.900.000	389
14-Aug-25	500	8.790.000	470
21-Aug-25	480	7.100.000	430
28-Aug-25	450	6.770.000	418
04-Sep-25	430	6.600.000	425
11-Sep-25	455	5.000.000	380
18-Sep-25	455	5.240.000	400
25-Sep-25	450	6.850.000	407
02-Oct-25	450	5.450.000	399
09-Oct-25	450	6.900.000	420
16-Oct-25	450	6.870.000	420
23-Oct-25	450	6.890.000	427
30-Oct-25	455	6.500.000	410

4.3.1 Penentuan Nilai Minimum dan Maksimum

Nilai minimum dan maksimum dihitung hanya dari data training untuk mencegah kebocoran data. Nilai min dan maks dapat dilihat pada Tabel 5 berikut:

Tabel 5 Nilai minimum dan maksimum

	Stok	Pendapatan
MIN	420	4900000
MAX	600	8790000

4.3.2 Normalisasi Data

Normalisasi dilakukan menggunakan metode Min–Max untuk menyamakan skala antar variabel prediktor menggunakan persamaan 2.

$$X_{\min} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

$$X_1 = 450$$

$$X1_{Norm} = \frac{450 - 420}{600 - 420} = \frac{30}{180} = 0.17$$

$$X_2 = 5200000$$

$$X2_{Norm} = \frac{5200000 - 8790000}{8790000 - 4900000} = \frac{-3590000}{3890000} = 0.08$$

Normalisasi hanya diterapkan pada variabel X_1 (Stok) dan X_2 (Pendapatan), sedangkan variabel Y (Penjualan) tidak dinormalisasi. Hasil normalisasi data dapat dilihat pada Tabel 6.

Tabel 6 Normalisasi data

Stok (X_1)	Pendapatan (X_2)	Penjualan (Y)	X_1 (Norm)	X_2 (Norm)
450	Rp5.200.000	390	0,17	0,08
450	Rp5.950.000	410	0,17	0,27
440	Rp5.950.000	410	0,11	0,27
420	Rp4.900.000	370	0,00	0,00
450	Rp5.800.000	408	0,17	0,23
450	Rp6.100.000	416	0,17	0,31
500	Rp6.750.000	435	0,44	0,48
500	Rp7.155.000	460	0,44	0,58
500	Rp5.950.000	410	0,44	0,27
450	Rp5.155.000	395	0,17	0,07
450	Rp5.100.000	395	0,17	0,05
500	Rp7.750.000	460	0,44	0,73
450	Rp5.150.000	380	0,17	0,06
500	Rp7.900.000	458	0,44	0,77
450	Rp5.250.000	400	0,17	0,09
450	Rp5.110.000	395	0,17	0,05
420	Rp5.110.000	390	0,00	0,05
425	Rp5.155.000	393	0,03	0,07
600	Rp8.150.000	470	1,00	0,84
600	Rp8.525.000	482	1,00	0,93
600	Rp7.000.000	450	1,00	0,54
450	Rp5.455.000	423	0,17	0,14
450	Rp5.310.000	410	0,17	0,11
450	Rp5.285.000	408	0,17	0,10
450	Rp6.950.000	448	0,17	0,53
450	Rp6.735.000	430	0,17	0,47
450	Rp6.870.000	447	0,17	0,51
450	Rp5.370.000	398	0,17	0,12
460	Rp6.810.000	435	0,22	0,49
500	Rp8.710.000	473	0,44	0,98

Stok (X1)	Pendapatan (X2)	Penjualan (Y)	X1 (Norm)	X2 (Norm)
450	Rp5.550.000	410	0,17	0,17
450	Rp5.185.000	387	0,17	0,07
450	Rp4.990.000	370	0,17	0,02
450	Rp5.870.000	381	0,17	0,25
450	Rp6.820.000	407	0,17	0,49
450	Rp6.750.000	398	0,17	0,48
450	Rp6.595.000	405	0,17	0,44
450	Rp5.390.000	395	0,17	0,13
450	Rp6.885.000	419	0,17	0,51
450	Rp5.900.000	389	0,17	0,26
500	Rp8.790.000	470	0,44	1,00
480	Rp7.100.000	430	0,33	0,57
450	Rp6.770.000	418	0,17	0,48
430	Rp6.600.000	425	0,06	0,44
455	Rp5.000.000	380	0,19	0,03
455	Rp5.240.000	400	0,19	0,09
450	Rp6.850.000	407	0,17	0,50
450	Rp5.450.000	399	0,17	0,14
450	Rp6.900.000	420	0,17	0,51
450	Rp6.870.000	420	0,17	0,51
450	Rp6.890.000	427	0,17	0,51
455	Rp6.500.000	410	0,19	0,41

4.4 Decision Tree

Regresi Linier digunakan untuk memodelkan hubungan antara variabel dependen yaitu penjualan mingguan (Y) dengan dua variabel independen yaitu stok (X1) dan produk terjual (X2). Metode ini mengasumsikan hubungan linier antar variabel sehingga dapat menunjukkan seberapa besar pengaruh perubahan stok dan jumlah terjual terhadap hasil penjualan.

4.4.1 Defenisi Operasional Variabel

Penelitian ini menggunakan satu variabel terikat dan dua variabel bebas, dengan definisi operasional sebagaimana ditunjukkan pada Tabel 7 berikut.

Tabel 7 Operasional variabel

Kode	Variabel	Jenis	Definisi Operasional
Y	Penjualan	Terikat	Jumlah produk yang terjual pada minggu berjalan (unit).
X1	Stok	Bebas	Jumlah produk yang tersedia untuk dijual pada minggu berjalan (unit).
X2	Pendapatan	Bebas	Total pendapatan penjualan pada satu minggu (rupiah).

4.4.2 Model Regresi

Setelah variabel diidentifikasi, dilakukan proses pelatihan model regresi linear untuk mencari hubungan matematis antara variabel input dan output. Model regresi menggunakan pendapatan yang telah ditransformasikan ke skala jutaan rupiah untuk menghindari perbedaan skala numerik dan menghasilkan koefisien yang lebih stabil. Model regresi yang tercoipta dapat dilihat pada Tabel 8 berikut:

Tabel 8 Model regresi

X1²	X2²	Y²	X1Y	X2Y	X1X2
202500	27	152100	175500	2028	2340
202500	35	168100	184500	2440	2678
193600	35	168100	180400	2440	2618
176400	24	136900	155400	1813	2058
202500	34	166464	183600	2366	2610
202500	37	173056	187200	2538	2745
250000	46	189225	217500	2936	3375
250000	51	211600	230000	3291	3578
250000	35	168100	205000	2440	2975
202500	27	156025	177750	2036	2320
202500	26	156025	177750	2015	2295
250000	60	211600	230000	3565	3875
202500	27	144400	171000	1957	2318
250000	62	209764	229000	3618	3950
202500	28	160000	180000	2100	2363
202500	26	156025	177750	2018	2300
176400	26	152100	163800	1993	2146
180625	27	154449	167025	2026	2191
360000	66	220900	282000	3831	4890
360000	73	232324	289200	4109	5115
360000	49	202500	270000	3150	4200
202500	30	178929	190350	2307	2455
202500	28	168100	184500	2177	2390
202500	28	166464	183600	2156	2378
202500	48	200704	201600	3114	3128
202500	45	184900	193500	2896	3031
202500	47	199809	201150	3071	3092
202500	29	158404	179100	2137	2417
211600	46	189225	200100	2962	3133
250000	76	223729	236500	4120	4355
202500	31	168100	184500	2276	2498
202500	27	149769	174150	2007	2333
202500	25	136900	166500	1846	2246
202500	34	145161	171450	2236	2642
202500	47	165649	183150	2776	3069
202500	46	158404	179100	2687	3038
202500	43	164025	182250	2671	2968
202500	29	156025	177750	2129	2426
202500	47	175561	188550	2885	3098
202500	35	151321	175050	2295	2655
250000	77	220900	235000	4131	4395
230400	50	184900	206400	3053	3408
202500	46	174724	188100	2830	3047
184900	44	180625	182750	2805	2838
207025	25	144400	172900	1900	2275

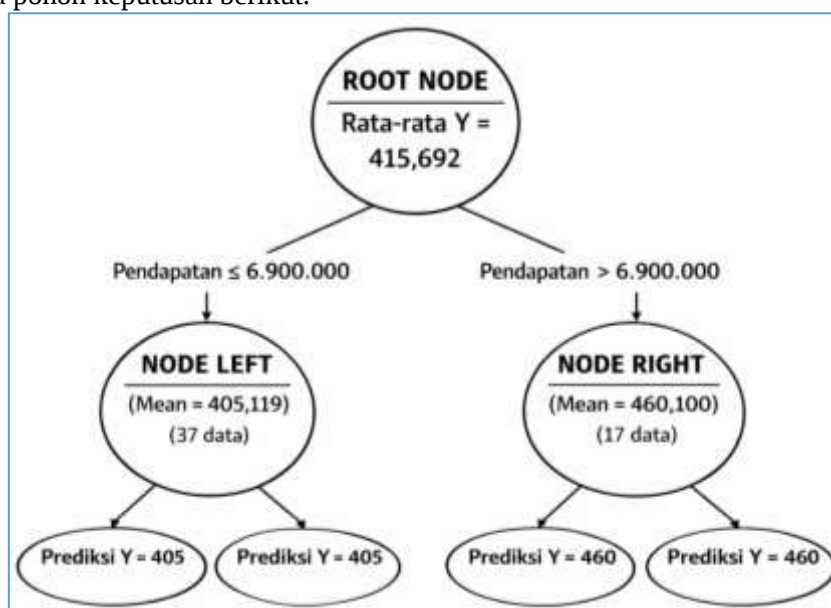
$X1^2$	$X2^2$	Y^2	$X1Y$	$X2Y$	$X1X2$
207025	27	160000	182000	2096	2384
202500	47	165649	183150	2788	3083
202500	30	159201	179550	2175	2453
202500	48	176400	189000	2898	3105
202500	47	176400	189000	2885	3092
202500	47	182329	192150	2942	3101
207025	42	168100	186550	2665	2958
11285000	2093	9024564	10073775	136625	152421

4.4.3 Evaluasi Model

Berdasarkan hasil evaluasi model regresi linear, diperoleh nilai Mean Squared Error (MSE) sebesar 125,71, Mean Absolute Error (MAE) sebesar 0,12, serta Root Mean Squared Error (RMSE) sebesar 11,21. Nilai MAE yang sangat kecil menunjukkan bahwa secara rata-rata selisih antara nilai aktual dan hasil prediksi relatif rendah, sehingga model memiliki tingkat kesalahan prediksi yang minimal. Nilai RMSE yang sedikit lebih besar dari MAE mengindikasikan adanya beberapa kesalahan prediksi, namun masih dalam batas yang dapat diterima dan tidak signifikan. Selain itu, nilai koefisien determinasi (R^2) sebesar 0,8322 menunjukkan bahwa model regresi linear mampu menjelaskan sekitar 83,22% variasi data aktual, sementara sisanya dipengaruhi oleh faktor lain di luar model. Hasil ini menunjukkan bahwa model regresi linear yang dibangun memiliki kemampuan prediksi yang sangat baik dan layak digunakan sebagai alat bantu analisis serta peramalan.

4.4.4 Analisa Decision Tree

Model Decision Tree Regresi dibentuk dengan tujuan meminimalkan nilai error menggunakan kriteria Sum of Squared Error (SSE). Pada tahap ini dilakukan evaluasi terhadap variabel Stok ($X1$) dan Pendapatan ($X2$) sebagai kandidat pemisah data. Hasil evaluasi akhir model Decision Tree Regresi. Penurunan SSE dari 38.959,077 menjadi 14.543,305 menunjukkan peningkatan akurasi yang signifikan. Nilai RMSE dan MAE yang relatif kecil mengindikasikan kesalahan prediksi yang rendah. Nilai R^2 sebesar 0,627 menandakan bahwa variabel pendapatan mampu menjelaskan 62,7% variasi penjualan, sehingga model dinilai cukup baik dan layak digunakan untuk prediksi. Hasil akhir ini dapat dilihat pada pohon keputusan berikut.



Gambar 2 Pohon keputusan

Gambar 2 di atas menunjukkan struktur Decision Tree Regresi yang terbentuk berdasarkan variabel pendapatan. Proses dimulai dari root node yang menggunakan rata-rata seluruh data

penjualan sebagai prediksi awal. Selanjutnya dilakukan pemisahan node menggunakan threshold pendapatan sebesar 6.900.000. Data dengan pendapatan di bawah atau sama dengan threshold masuk ke node kiri dengan rata-rata penjualan 405,119, sedangkan data di atas threshold masuk ke node kanan dengan rata-rata penjualan 460,100. Struktur ini menunjukkan bahwa variabel pendapatan mampu memisahkan data secara efektif dan menghasilkan model prediksi yang lebih akurat.

4.5 Perbandingan Metode

Kedua metode, yaitu Regresi Linear dan Decision Tree Regresi, memberikan pendekatan yang berbeda dalam menganalisis dan memprediksi penjualan, di mana Regresi Linear berfokus pada hubungan linier antara variabel independen dan variabel dependen berupa penjualan dengan persamaan matematis yang sederhana dan mudah dipahami. Berdasarkan hasil evaluasi, model regresi linear menghasilkan nilai Mean Absolute Error (MAE) sebesar 0,12, Root Mean Squared Error (RMSE) sebesar 11,21, serta koefisien determinasi (R^2) sebesar 0,8322, yang menunjukkan bahwa rata-rata kesalahan prediksi relatif kecil dan sekitar 83,22% variasi penjualan dapat dijelaskan oleh model. Hasil ini mengindikasikan bahwa model regresi linear memiliki kemampuan prediksi yang sangat baik dan layak digunakan sebagai alat bantu analisis dan peramalan, meskipun tetap memiliki keterbatasan dalam menangkap pola hubungan non-linier sehingga pada kondisi tertentu diperlukan metode pemodelan yang lebih kompleks.

Sementara itu, Decision Tree Regresi menawarkan pendekatan yang lebih fleksibel karena tidak mengasumsikan adanya hubungan linier antara variabel. Model ini bekerja dengan membagi data ke dalam beberapa node berdasarkan nilai threshold yang paling optimal. Berdasarkan hasil perhitungan error dan penurunan SSE, variabel pendapatan dipilih sebagai atribut terbaik dengan threshold sebesar 6.900.000. Model Decision Tree Regresi menghasilkan nilai MAE sebesar 13,345, RMSE sebesar 16,724, dan R^2 sebesar 0,627. Nilai R^2 tersebut menunjukkan bahwa model mampu menjelaskan sekitar 62,7% variasi penjualan, yang berarti lebih baik dibandingkan regresi linear. Meskipun struktur Decision Tree lebih kompleks, model ini mampu menangkap pola non-linier dan memberikan interpretasi keputusan yang jelas melalui pembentukan node. Untuk lebih jelas, perbandingan metode ini dapat dilihat pada Tabel 9 berikut:

Tabel 9 Perbandingan metode

Metrik Evaluasi	Regresi Linear	Decision Tree Regresi
Mean Absolute Error (MAE)	0,12 unit	13,345 unit
Root Mean Squared Error (RMSE)	11,21 unit	16,724 unit
R^2 (Koefisien Determinasi)	0,83223	0,627
Keakuratan Prediksi	Cukup baik, namun error relatif besar	Lebih akurat dengan error kecil
Kompleksitas Model	Relatif sederhana	Lebih kompleks
Kemampuan Menangani Data Non-Linier	Terbatas	Lebih baik
Interpretasi Model	Persamaan matematis	Pohon keputusan visual
Kemampuan Segmentasi Data	Tidak ada	Ada (berdasarkan threshold pendapatan)

5 Kesimpulan

Hasil penelitian menunjukkan bahwa *Regresi Linear* memiliki kinerja yang lebih baik dibandingkan *Decision Tree Regresi* dalam memprediksi penjualan. Hal ini ditunjukkan oleh nilai MAE dan RMSE yang lebih kecil serta nilai koefisien determinasi (R^2) yang lebih tinggi, sehingga model *Regresi Linear* mampu menghasilkan prediksi yang lebih akurat dan stabil. Regresi Linear lebih cocok digunakan pada kondisi data yang memiliki pola hubungan linier antara variabel pendapatan dan penjualan, dengan tingkat kesalahan prediksi yang relatif rendah serta interpretasi model yang sederhana. Sementara itu, *Decision Tree Regresi* lebih unggul dalam menangani data dengan pola kompleks dan non-linier serta mampu melakukan segmentasi data berdasarkan threshold pendapatan, meskipun tingkat akurasi masih berada di bawah *Regresi Linear*. Oleh karena itu, dalam penelitian ini Regresi Linear direkomendasikan sebagai model utama untuk memprediksi penjualan berdasarkan pendapatan.

Referensi

- [1] A. Mei Sarah, B. Kurniadi, and E. Warsini, "Implementasi Metode Regresi Linier dalam memprediksi Penyakit Anemia Secara Dini," *J. Tek. (Teknologi Komput. dan Sist. Informasi)*, Vol. 3, No. 1, pp. 14–23, 2023.
- [2] P. K. Illahi, A. R. Viana, N. F. Mita Peramata, and M. Yudha Pratama, "Application of Decision Tree Algorithm and Linear Regression for Breast Cancer Classification," in *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2023, pp. 86–92.
- [3] K. Oktavia Usnaat, P.K Kellen, Yoseph, J. Tey Seran, "Implementasi Metode Decision Tree dalam memprediksi Penjualan Produk Terlaris (Studi Kasus: Kreatif Mart Atambua)," *Zo. J. Sist. Inf.*, Vol. 7, No. 1, pp. 317–330, 2025.
- [4] R. A. Saputra and A. Pratama, "Implementasi Decision Tree untuk Prediksi Harga Rumah di Daerah Tebet," *J. Inf. Syst. Manag.*, Vol. 6, No. 2, pp. 4–10, 2025.
- [5] S. Hanifah, F. Akbar, and R. Pratama Santi, "Implementasi Business Intelligence dan Prediksi menggunakan Regresi Linear pada Data Penjualan dan Breakage di PT XYZ," *J. Nas. Teknol. dan Sist. Inf.*, Vol. 08, No. 03, pp. 144–152, 2022.
- [6] D. Nabila and E. Victor Haryanto, "Penerapan Metode Regresi Linear Berganda dalam Data Mining Prediksi Penjualan Produk pada PT Cipta Mebelindo berbasis Web," *JID (Jurnal Info Digit.)*, Vol. 2, No. 3, pp. 1940–1953, 2024.
- [7] A. Putra, "Penerapan Data Mining untuk Prediksi Harga Kayu Manis menggunakan Algoritma Regresi Linear Sederhana," *J. Sains Inform. Terap. (JSIT)*, Vol. 04, No. 03, pp. 574–580, 2025.
- [8] N. Aprilia and I. Fenriana, "Analisis dan Prediksi Total Penjualan Produk di Supermarket menggunakan Linear Regression," in *POTERS: Proceedings of Technology, Engineering and Computers*, 2026, pp. 79–86.
- [9] A. P. Kohan et al., "Komparasi Akurasi Regresi Linear Berganda dan Decision Tree Prediksi Harga Motor Bekas berbasis Website," *INFOTECH J.*, Vol. 11, No. 2, pp. 386–393, 2025.
- [10] A. Hurifiani, A. I. Purnamasari, and I. Ali, "Penerapan Algoritma Regresi Linear untuk Prediksi Penjualan Alat Tulis Kantor (ATK) di BUMDES," *JATI (Jurnal Mhs. Tek. Inform.)*, Vol. 8, No. 1, pp. 266–273, 2024.
- [11] O. J. Ababil, S. A. Wibowo, and H. Z. Zahro, "Penerapan Metode Regresi Linier dalam Prediksi Penjualan Liquid Vape di Toko Vapor Pandaan berbasis Website," *JATI (Jurnal Mhs. Tek. Inform.)*, Vol. 6, No. 1, pp. 186–195, 2022.
- [12] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *J. Sains dan Edukasi Sains*, Vol. 6, No. 1, pp. 34–43, 2023.
- [13] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, "Prediksi Penjualan Kendaraan menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia," *Innov. J. Soc. SCI. Res.*, Vol. 4, No. 4, pp. 1199–1216, 2024.
- [14] A. Supriyatna, A. I. Purnamasari, and I. Ali, "Analisis Penjualan Produk UMKM di Shopee pada Toko Agung0Na9 menggunakan Model Algoritma Regresi Linear," *JATI (Jurnal Mhs. Tek. Inform.)*, Vol. 8, No. 2, pp. 1911–1915, 2024, DOI: 10.36040/jati.v8i2.8372.
- [15] M. K. Effendi, S. Sriyanto, G. Goesderilidar, H. Widi Nugroho, and J. Triloka, "Penerapan Data Mining untuk Prediksi Pendaftar PPDB di SMKN3 Metro menggunakan Machine Learning," *J. Jar. Sist. Inf. Robot.*, Vol. 9, No. 1, pp. 59–63, 2025, [Online]. Available: <http://ojsamik.amikmitragama.ac.id>
- [16] P. C. Hartono and A. D. Widianoro, "Analisis Prediksi Harga Saham Unilever menggunakan Regresi Linier dengan RapidMiner," *J. Comput. Inf. Syst. Ampera*, Vol. 5, No. 3, pp. 2775–2496, 2024, [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>