

## ANALISIS DAN PREDIKSI TOTAL PENJUALAN PRODUK DI SUPERMARKET MENGGUNAKAN *LINEAR REGRESSION*

Natasya Aprilia<sup>1</sup>, Indah Fenriana<sup>2\*</sup>

<sup>1,2</sup> Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma

\*Corresponding Author, email: [indah.fenriana@ubd.ac.id](mailto:indah.fenriana@ubd.ac.id)

### ABSTRAK

Perkembangan teknologi informasi mendorong sektor ritel, khususnya supermarket, untuk memanfaatkan data transaksi penjualan sebagai dasar pengambilan keputusan yang lebih akurat dan efisien. Volume data transaksi yang terus meningkat menuntut adanya metode analisis prediktif guna memperkirakan total penjualan secara objektif dan terukur. Penelitian ini bertujuan untuk menganalisis serta memprediksi total penjualan produk di supermarket menggunakan algoritma Linear Regression, serta mengevaluasi kinerja model prediksi yang dihasilkan. Dataset yang digunakan berupa data transaksi supermarket yang mencakup variabel harga satuan (unit price) dan jumlah pembelian (quantity) sebagai variabel independen, serta total penjualan sebagai variabel dependen. Metode penelitian yang diterapkan meliputi tahap preprocessing data, pembagian data menjadi data latih dan data uji, pelatihan model Linear Regression menggunakan metode Ordinary Least Squares (OLS), serta evaluasi performa model menggunakan metrik Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ). Hasil penelitian menunjukkan bahwa algoritma Linear Regression mampu memodelkan hubungan antara harga satuan, jumlah pembelian, dan total penjualan dengan tingkat akurasi yang baik serta kesalahan prediksi yang relatif rendah. Selain itu, model yang dibangun memiliki kestabilan yang cukup baik dalam memprediksi total penjualan pada data uji. Kesimpulan dari penelitian ini adalah bahwa Linear Regression efektif digunakan sebagai metode prediksi total penjualan supermarket dan dapat dimanfaatkan sebagai alat bantu analisis penjualan serta pendukung pengambilan keputusan berbasis data.

**Kata kunci:** *Linear Regression*, Prediksi Penjualan, Streamlit, Supermarket, Transaksi Penjualan.

### I. PENDAHULUAN

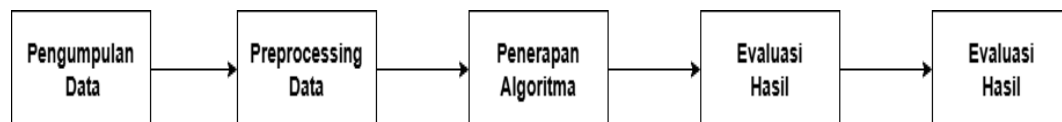
Persaingan bisnis ritel yang semakin ketat menuntut supermarket memiliki kemampuan analisis dan prediksi penjualan yang akurat. Ketidaktepatan prediksi total penjualan dapat menyebabkan ketidakseimbangan stok, meningkatnya biaya operasional, serta menurunnya kepuasan pelanggan, sehingga pemanfaatan data transaksi penjualan menjadi penting dalam pengambilan keputusan operasional (Hutagalung, Simatupang, and Simatupang 2023). Seiring meningkatnya volume data transaksi, pendekatan data-driven analysis banyak diterapkan dalam sektor ritel. Salah satu metode yang umum digunakan adalah Linear Regression karena

kesederhanaan dan kemampuannya dalam memodelkan hubungan linear antar variabel (Yola and Budianto 2013). Penelitian sebelumnya menunjukkan bahwa harga satuan dan jumlah pembelian berpengaruh signifikan terhadap total penjualan dalam satu transaksi (Aria Bima, Susanti, and Nugrahanti 2022).

Meskipun demikian, penerapan Linear Regression dalam aplikasi web interaktif serta evaluasinya menggunakan metrik RMSE, MAE, dan  $R^2$  masih terbatas (Amansyah et al. 2024). Oleh karena itu, penelitian ini bertujuan membangun model prediksi total penjualan supermarket menggunakan Linear Regression dan mengimplementasikannya dalam aplikasi berbasis web dengan batasan variabel harga satuan dan jumlah pembelian.

## II. METODOLOGI

Penelitian ini bertujuan untuk membangun dan mengevaluasi model prediksi total penjualan produk di supermarket menggunakan algoritma Linear Regression (Ramdhani and Setiawan 2024). Proses penelitian melibatkan beberapa tahapan utama, yaitu pengumpulan data, preprocessing data, penerapan algoritma, evaluasi hasil, serta visualisasi dan interpretasi hasil prediksi.



Gambar 1. Tahapan Penelitian

### 2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data transaksi penjualan supermarket yang diperoleh dari dataset publik. Dataset tersebut berisi data historis transaksi penjualan yang mencakup beberapa atribut, di antaranya harga satuan (unit price), jumlah pembelian (quantity), dan total penjualan. Data ini digunakan sebagai dasar dalam membangun model prediksi total penjualan karena merepresentasikan aktivitas transaksi pelanggan di supermarket. (Wardhana et al. 2021).

## 2.2 Preprocessing Data

Tahap preprocessing data dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Pada tahap ini, dilakukan pembersihan data dari nilai kosong atau tidak valid, penyesuaian format data, serta pemilihan atribut yang relevan. Atribut yang digunakan dalam penelitian ini adalah harga satuan dan jumlah pembelian sebagai variabel independen, serta total penjualan sebagai variabel dependen. Proses preprocessing bertujuan untuk meningkatkan akurasi dan kestabilan model prediksi yang akan dibangun.

## 2.3 Penerapan Algoritma Linear Regression

Setelah tahap preprocessing, algoritma Linear Regression diterapkan untuk memprediksi total penjualan supermarket. Algoritma ini digunakan untuk memodelkan hubungan antara variabel harga satuan (*unit price*) dan jumlah pembelian (*quantity*) terhadap total penjualan. Proses pelatihan model dilakukan menggunakan metode Ordinary Least Squares (OLS) untuk meminimalkan selisih antara nilai aktual dan nilai prediksi (Ramadhan and Safitri 2019). Dataset dibagi menjadi data latih dan data uji untuk menguji kinerja model, sedangkan hasil prediksi dievaluasi menggunakan metrik RMSE, MAE, dan koefisien determinasi ( $R^2$ ) guna menilai tingkat akurasi model (Hidayat et al. 2025).

## 2.4 Evaluasi Hasil Mining

Evaluasi hasil mining dilakukan untuk mengukur kinerja model prediksi total penjualan yang dihasilkan oleh algoritma Linear Regression. Evaluasi dilakukan dengan membandingkan nilai prediksi dengan nilai aktual pada data uji. Metrik evaluasi yang digunakan dalam penelitian ini adalah Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ), yang umum digunakan untuk menilai performa model regresi. Root Mean Square Error (RMSE) digunakan untuk mengukur rata-rata kesalahan prediksi dengan memberikan penalti lebih besar pada kesalahan yang bernilai besar. RMSE dihitung menggunakan persamaan berikut:

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

Mean Absolute Error (MAE) digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi. MAE dihitung dengan persamaan berikut:

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

Koefisien determinasi ( $R^2$ ) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data total penjualan. Nilai  $R^2$  berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan model memiliki kinerja yang baik.  $R^2$  dihitung menggunakan persamaan:

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

## 2.5 Visualisasi dan Interpretasi

Hasil prediksi total penjualan divisualisasikan untuk mempermudah analisis dan interpretasi kinerja model Linear Regression. Visualisasi yang digunakan berupa grafik perbandingan nilai aktual dan nilai prediksi serta grafik kesalahan prediksi, yang bertujuan untuk menilai akurasi model dan memahami pola penjualan. Hasil visualisasi ini digunakan sebagai dasar dalam mendukung pengambilan keputusan terkait pengelolaan stok dan perencanaan penjualan (Lestari 2023).

## III. HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan ini disusun dengan menyesuaikan seluruh tahapan pada metodologi penelitian, sehingga setiap proses yang dilakukan dapat terjawab secara sistematis dan terukur.

### 3.1 Penggunaan Dataset

Data yang digunakan dalam penelitian ini berasal dari dataset transaksi penjualan supermarket yang berjumlah 1.000 data transaksi. Dataset ini memuat 17

atribut, di antaranya unit price, quantity, tax 5%, gross income, dan total penjualan. Data transaksi ini merepresentasikan aktivitas pembelian pelanggan dalam periode tertentu dan digunakan sebagai dasar analisis serta pemodelan prediksi total penjualan.

**Tabel 1. Isi Dataset**

Invoice ID	Branch	City	Customer	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross mar	gross inco	Rating
750-67-84	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08	Ewallet	522.83	4.761905	26.1415	9.1
226-31-30	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	3/8/2019	10:29	Cash	76.4	4.761905	3.82	9.6
631-41-31	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23	Credit card	324.31	4.761905	16.2155	7.4
123-19-11	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761905	23.288	8.4
373-73-79	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	604.17	4.761905	30.2085	5.3
699-14-30	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/25/2019	18:30	Ewallet	597.73	4.761905	29.8865	4.1
355-53-59	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2/25/2019	14:36	Ewallet	413.04	4.761905	20.652	5.8
315-22-56	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2/24/2019	11:38	Ewallet	735.6	4.761905	36.78	8
665-32-91	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	1/10/2019	17:15	Credit card	72.52	4.761905	3.626	7.2

### 3.2 Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Berdasarkan hasil pembersihan data, tidak ditemukan nilai kosong (missing value) yang signifikan pada atribut utama yang digunakan. Selanjutnya dilakukan seleksi atribut dengan menetapkan unit price dan quantity sebagai variabel independen serta total penjualan sebagai variabel dependen. Hasil preprocessing menunjukkan bahwa data telah siap digunakan untuk proses pelatihan model Linear Regression tanpa memerlukan transformasi tambahan.

### 3.3 Penerapan Algoritma *Linear Regression*

Setelah proses preprocessing, algoritma Linear Regression diterapkan menggunakan metode Ordinary Least Squares (OLS). Dataset dibagi menjadi data latih dan data uji untuk mengukur kemampuan generalisasi model. Hasil pelatihan model menunjukkan adanya hubungan linear yang positif antara harga satuan dan jumlah pembelian terhadap total penjualan. Model regresi yang dihasilkan mampu mempelajari pola hubungan antar variabel secara konsisten berdasarkan data latih. (Anggara et al. 2023).

### 3.4 Evaluasi dan Pembahasan Hasil

Evaluasi kinerja model dilakukan menggunakan metrik Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ). Berdasarkan hasil evaluasi pada data uji, diperoleh nilai RMSE dan MAE yang

relatif rendah, yang menunjukkan bahwa selisih antara nilai prediksi dan nilai aktual masih dalam batas yang dapat diterima. Selain itu, nilai  $R^2$  mendekati 1, yang mengindikasikan bahwa model Linear Regression mampu menjelaskan sebagian besar variasi data total penjualan. Hasil ini menunjukkan bahwa model memiliki tingkat akurasi dan kestabilan yang baik. (Suraya 2024).

### 3.5 Visualisasi dan Interpretasi Hasil

Visualisasi hasil prediksi dilakukan dengan membandingkan nilai aktual dan nilai prediksi total penjualan dalam bentuk grafik. Hasil visualisasi menunjukkan bahwa garis prediksi mengikuti pola data aktual dengan cukup baik. Selain itu, grafik kesalahan prediksi menunjukkan distribusi error yang relatif kecil dan tidak menyimpang secara ekstrem. Temuan ini memperkuat hasil evaluasi numerik bahwa model Linear Regression layak digunakan untuk memprediksi total penjualan supermarket. Berdasarkan keseluruhan hasil dan pembahasan tersebut, dapat disimpulkan bahwa setiap tahapan metodologi penelitian, mulai dari pengumpulan data, preprocessing, penerapan algoritma, evaluasi model, hingga visualisasi hasil, telah berhasil dilaksanakan dan saling mendukung dalam menghasilkan model prediksi total penjualan yang akurat dan dapat diimplementasikan.

## IV. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Linear Regression efektif digunakan untuk memprediksi total penjualan produk di supermarket berdasarkan data transaksi historis (Setiawan 2019). (Setiawan 2019). Model yang dibangun menunjukkan kinerja yang baik berdasarkan metrik RMSE, MAE, dan  $R^2$ , serta mampu memodelkan hubungan antara harga satuan, jumlah pembelian, dan total penjualan secara akurat (Abdurrahman 2017). Implementasi model dalam aplikasi berbasis web juga memudahkan pengguna dalam melakukan analisis dan prediksi penjualan secara otomatis. Penelitian selanjutnya disarankan untuk menambahkan variabel lain serta membandingkan Linear Regression dengan algoritma prediksi yang lebih kompleks guna meningkatkan akurasi model (Alfarizi et al. 2023).

## DAFTAR PUSTAKA

- Abdurrahman, Ginanjar. 2017. "Analisis Aturan Asosiasi Data Transaksi Supermarket Menggunakan Algoritma Apriori." *Sistem & Teknologi Informasi Indonesia* 2(2): 100–111.
- Alfarizi, M. Riziq Sirfatullah, Muhamad Zidan Al-farish, Muhamad Taufiqurrahman, Ginan Ardiansah, and Muhamad Elgar. 2023. "Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning Dan Deep Learning." *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)* 2(1): 1–6.
- Amansyah, Ilham, Jamaludin Indra, Euis Nurlaelasari, and Ayu Ratna Juwita. 2024. "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus Pada Industri Otomotif Di Indonesia." *Journal Of Social Science Research* 4: 1199–1216. <https://j-innovative.org/index.php/Innovative>.
- Anggara, Aditya, Karina Auliasari, Yosep Agus Pranoto, and Fakultas Teknologi Industri. 2023. "PENYEWAAN KAMERA DI JOE KAMERA." 7(1): 852–58.
- Aria Bima, Alim Citra, Pratiwi Susanti, and Fatim Nugrahanti. 2022. "Analisis Regresi Linear Untuk Memprediksi Kriteria Pengaruh Penjualan Mobil Honda Civic Dan Toyota Camry Pada Kota Surabaya." *Bulletin of Information Technology (BIT)* 3(4): 217–30. doi:10.47065/bit.v3i4.361.
- Hidayat, Rahmat, Haris Tri Saputra, Mirdatul Husnah, Nabila Nabila, M Bintang Hidayatullah, Muhammad Naufal Nazhmi, Jauzaa Azra, and Astri Rana. 2025. "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi Di Supermarket." *Simkom* 10(1): 101–9. doi:10.51717/simkom.v10i1.703.
- Hutagalung, DEvi Kristina, Meiyati Simatupang, and Rini Simatupang. 2023. "3 1,2,3." 3(3): 627–32.
- Lestari, Silvia. 2023. "Analisis Algoritma Regresi Linear Sederhana Dalam Memprediksi Tingkat Penjualan Album KPOP." *INSOLOGI: Jurnal Sains dan Teknologi* 2(1): 199–209. doi:10.55123/insologi.v2i1.1692.
- Ramadhan, Puji Sari, and Nurdianti Safitri. 2019. "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi

- Linier Berganda Pada BPS Deli Serdang.” 18(1): 55–61.
- Ramdhani, Fauzi, and Kiki Setiawan. 2024. “Penerapan Data Mining Untuk Prediksi Pelanggan Di PT. XYZ Menggunakan Algoritma Linear Regression.” *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 4(2): 490–97. doi:10.57152/malcom.v4i2.1217.
- Setiawan, I G A. 2019. “STRATEGI BERSAING TOKO RITEL ‘Strategi Digitalisasi Rantai Pasokan Pamella Supermarket Yogyakarta.’” *Ncab*: 472–78.
- Suraya, Marshanda. 2024. “Algoritma Pemrograman: Kunci Efisiensi Dalam Pengelolaan Data Besar.” *Jurnal Pendidikan Berkarakter* 2(1): 308–14. <https://doi.org/10.51903/pendekar.v2i1.601>.
- Wardhana, Andhika Wisnu, Endah Patimah, Alleyda Irzky Shafarindu, Yohanne Marintan Siahaan, Ballya Vicky Haekal, and Desta Sandya Prasvita. 2021. “Klasifikasi Data Penjualan Pada Supermarket Dengan Metode Decision Tree.” *Senamika* 2(1): 660–67. <https://conference.upnvj.ac.id/index.php/senamika/article/view/1389>.
- Yola, Melfa, and Duwi Budianto. 2013. “Pelayanan Dan Harga Produk Pada Supermarket Dengan Menggunakan Metode Importance Performance Analysis ( Ipa ).” *Jurnal Optimasi Sistem Industri* 12(12): 301–9.