

## IMPLEMENTASI DATA MINING DALAM MENGANALISIS POLA DATA SUPPLY CHAIN UNTUK OPTIMALISASI JUMLAH STOK BARANG MENGGUNAKAN REGRESI LINIER

Nasywa Sakha Ningrum<sup>1</sup>, Saprudin<sup>2</sup>

<sup>12</sup> Universitas Pamulang, Tangerang Selatan, 15320, Indonesia

### Info Artikel

#### Riwayat Artikel:

Received october 21, 2025

Revised october 22, 2025

Accepted november 02, 2025

#### Corresponding Author:

Nasywa Sakha Ningrum

Email: nasywa3032@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

**Abstract** – In the era of big data, managing large and complex datasets has become a challenge in supporting operational efficiency within companies. This research aims to optimize inventory management of the Darts Live Credit Card (LCC) at PT. Mitra Media Integrasi by implementing data mining techniques using a linear regression algorithm. The study addresses issues related to discrepancies between the number of received and returned cards in the supply chain, which affect the distribution process and the availability of LCC cards across various darts outlets. A quantitative approach is used with the CRISP-DM methodology, which includes business understanding, data understanding, data preparation, modelling, evaluation, and deployment. Data collection was conducted through literature review, interviews, and historical data observation from 2024 to 2025. The linear regression analysis indicates that for every one-unit increase in the Received variable, the Returned variable is predicted to decrease by 1.2178 units, with an intercept value of 1.0199. The regression model yields an MSE of 10.8358, RMSE of 3.2918, and an  $R^2$  of 0.7308, indicating a reasonably good fit in explaining the dependent variable. These results help the company accurately predict LCC inventory needs, reduce the risk of shortages or overstock, and improve the efficiency of ordering and distribution processes.

**Keywords:** Data Mining, Stock Prediction, Linear Regression, LCC Card

**Abstrak Indonesia** – Dalam era big data, pengelolaan data yang kompleks dan besar menjadi tantangan dalam mendukung efisiensi operasional perusahaan. Penelitian ini bertujuan untuk mengoptimalkan manajemen persediaan kartu Darts Live Credit Card (LCC) pada PT. Mitra Media Integrasi melalui penerapan data mining dengan pendekatan algoritma regresi linier. Permasalahan yang diangkat mencakup ketidaksesuaian antara jumlah kartu yang diterima dan dikembalikan dalam rantai pasok, yang berdampak pada proses distribusi dan ketersediaan kartu di berbagai outlet permainan darts. Penelitian ini menggunakan pendekatan kuantitatif dengan metode CRISP-DM, yang meliputi proses pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Data diperoleh melalui studi literatur, wawancara, serta observasi data historis dari tahun 2024–2025. Hasil analisis regresi linier menunjukkan bahwa setiap peningkatan satu unit pada variabel Received, akan diikuti penurunan sebesar 1.2178 unit pada variabel Returned, dengan nilai intercept sebesar 1.0199. Model regresi ini menghasilkan nilai MSE sebesar 10.8358, RMSE 3.2918, dan  $R^2$  sebesar 0.7308, yang menunjukkan model cukup baik dalam menjelaskan variabel dependen. Dengan hasil ini, perusahaan dapat memprediksi kebutuhan stok kartu LCC secara lebih akurat, mengurangi risiko kekurangan maupun kelebihan stok, serta meningkatkan efisiensi dalam proses pemesanan dan distribusi.

**Kata Kunci:** Data Mining, Prediksi Stok, Regresi Linier, Kartu LCC

### 1. PENDAHULUAN

Era big data telah membawa pertumbuhan data dalam jumlah besar di berbagai bidang, sehingga diperlukan metode baru untuk mengolah dan

memperoleh wawasan yang bermanfaat. Salah satu metode yang digunakan adalah data mining, yang mampu mengatasi tantangan terkait volume, kecepatan, dan keragaman data. Dalam dunia

usaha, data mining berperan penting dalam memprediksi persediaan barang agar perusahaan dapat memenuhi permintaan pelanggan secara tepat waktu, sekaligus meningkatkan perencanaan dan efisiensi operasional.

PT. Mitra Media Integrasi, distributor terbesar mesin, aksesoris, dan suku cadang *DARTSLIVE* di Indonesia sejak 2024, juga aktif mempromosikan olahraga dart melalui turnamen, namun menghadapi masalah ketidaksesuaian data stok kartu LCC antara outlet dan gudang di tengah tingginya permintaan dari outlet lain. Observasi awal menunjukkan adanya ketidaksesuaian antara data stok kartu LCC yang masuk dan keluar dari outlet ke gudang, meskipun terdapat permintaan tinggi dari outlet lain. Setiap kartu memiliki identitas unik yang memerlukan pembaruan data sebelum dapat digunakan kembali, sehingga memperlambat proses distribusi. Kondisi ini berpotensi memicu pemesanan langsung dari Singapura yang memakan waktu dan mengganggu ketersediaan kartu secara optimal. Untuk mengatasi masalah tersebut, peneliti menggunakan teknik data mining dengan pendekatan CRISP-DM dan algoritma regresi linier guna memprediksi kebutuhan kartu LCC pada bulan-bulan berikutnya.

## 2. PENELITIAN YANG TERKAIT

Secara metodologis, literatur yang dikaji didominasi oleh penggunaan algoritma Regresi Linear, yang diaplikasikan dalam enam studi untuk memprediksi produksi dan stok berbagai komoditas. Regresi Linear, baik sederhana maupun berganda, digunakan untuk memprediksi produksi beras (Studi et al., 2024), produksi pakaian renang (Wahyudi et al., 2024), dan stok barang secara umum (Rizky et al., 2019). Analisis kinerja model Regresi Linear memperlihatkan akurasi yang variatif. Mayoritas studi melaporkan performa yang memuaskan, ditunjukkan oleh nilai *RMSE* yang rendah, seperti 0,566 untuk stok madu dan 0,089% untuk produksi pakaian renang. Namun, penelitian prediksi stok suku cadang HP (Aryudha & Hasibuan, 2024) mengindikasikan adanya deviasi yang signifikan dengan nilai *MAPE* 42,64%, menggarisbawahi bahwa akurasi model sangat sensitif terhadap karakteristik data. (Sarimole & Hakim, 2024) mengimplementasikan *Clustering K-Means* untuk klasifikasi persediaan, menghasilkan pengelompokan yang akurat (*SSE* 0,0041). Di sisi lain, (Tadayonrad & Ndiaye, 2023) mengembangkan model Indikator Kinerja Utama

(*KPI*) baru yang mengintegrasikan faktor keandalan rantai pasok dan musiman dalam peramalan permintaan, bertujuan untuk optimasi *safety stock*. Sementara itu, penelitian kualitatif oleh (Sari, 2022) menegaskan bahwa efektivitas gudang ditingkatkan melalui penerapan prosedur baku seperti metode *FIFO*, pencatatan yang sistematis, dan *stock opname* rutin. Secara komprehensif, ulasan terhadap penelitian-penelitian ini menyimpulkan bahwa meskipun algoritma Regresi Linear merupakan alat prediksi kuantitatif yang dominan dan seringkali efektif, keberhasilan penerapannya sangat ditentukan oleh konteks produk, kualitas data, dan implementasi model yang cermat.

## 3. METODE PENELITIAN

### 3.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui observasi, wawancara, studi pustaka, dan studi literatur, dengan fokus pada data historis *DARTSLIVE* dari PT. Mitra Media Integrasi periode 2024–2025. observasi dan wawancara secara offline dengan Bapak Ridwan Bintang pada tanggal yang sama di Cartel Billiards bertujuan memahami proses bisnis dan pengelolaan data historis *DARTSLIVE* melalui tanya jawab lisan (hasil terdapat pada lampiran), serta dilengkapi dengan studi pustaka yang mempelajari referensi buku, jurnal, dan situs terkait, dan studi literatur yang mengacu pada jurnal serta skripsi untuk menentukan metode optimalisasi stok barang.

### 3.2 CRISP-DM

Penelitian mengikuti tahapan CRISP-DM sebagai berikut:

#### 3.2.1 Business Understanding

Fase ini mencakup penentuan tujuan bisnis, penilaian situasi, penentuan tujuan *data mining*, dan perencanaan. Perlu adanya pemahaman mengenai masalah ketidaksesuaian stok *Live Credit Card* (LCC) di PT. Mitra Media Integrasi, yang menyebabkan risiko kekurangan atau kelebihan stok. Oleh karena itu, diperlukan prediksi jumlah kartu masuk dan keluar bulanan menggunakan *data mining*. Penilaian situasi dilakukan dengan mengidentifikasi sumber daya, kendala, dan risiko, menggunakan data transaksi stok dalam format CSV dari gudang ke outlet yang diperoleh dari database PT. Mitra Media Integrasi,

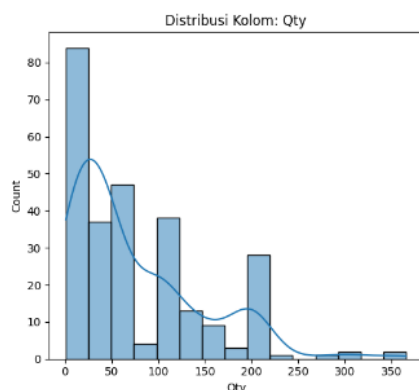
dengan fokus pada prediksi *Qty Received* dan *Qty Returned*, diikuti oleh penetapan tujuan *data mining* untuk membangun model regresi dalam memprediksi stok berdasarkan pola historis, mengidentifikasi faktor signifikan yang memengaruhi perputaran stok, dan menyediakan prediksi akurat untuk mendukung pengambilan keputusan berbasis data.

### 3.2.2 Data Understanding

Fase ini meliputi pengumpulan data awal, deskripsi data, eksplorasi data, dan verifikasi kualitas data. Collecting Initial Data: Data transaksi operasional DARTSLIVE, contoh dataset awal:

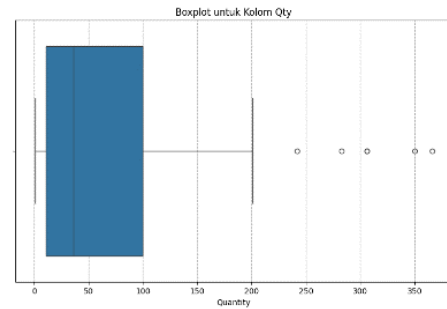
index	email	Accessories & Spareparts	Order	Status	Status Date	Qty	Data Entry	Remarks	Column 1
0	Dewa Sport & Bar	[DL2-7011_TBGI] Throwline Sticker	Shop Request	Received	23-Dec-2024	4	NaN	NaN	NaN
1	Gemu	[DL2-7011_TBGI] Throwline Sticker	Shop Request	Received	23-Dec-2024	3	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
464	Six pocket Billiard (sipp)	Tent Card Qris	Shop Request	Received	20-Mar-2025	1	Lody	new shop	NaN
465	Buddy Pool	Thermal Paper	Shop Request	Received	12-Dec-2024	2	NaN	By Hizkia	NaN

Gambar 3. 1 Dataset Awal



Gambar 3. 2 Histogram Distribusi Qty

Histogram ini menunjukkan sebaran nilai pada kolom Qty. Terlihat bahwa sebagian besar transaksi memiliki kuantitas yang relatif rendah, dengan distribusi yang condong ke kanan (right-skewed) yang mengindikasikan keberadaan nilai-nilai kuantitas yang sangat tinggi atau outlier.

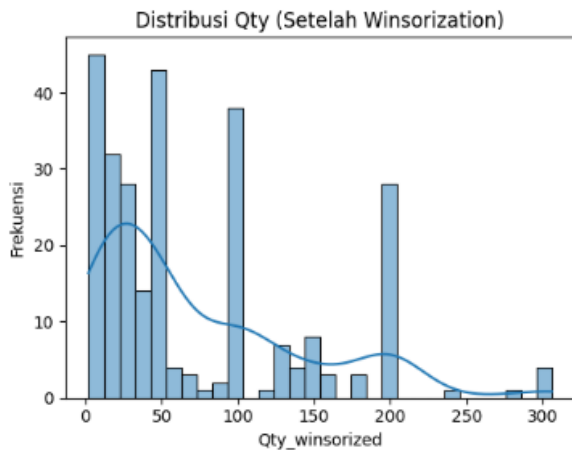


Gambar 3. 3 Boxplot Outlier

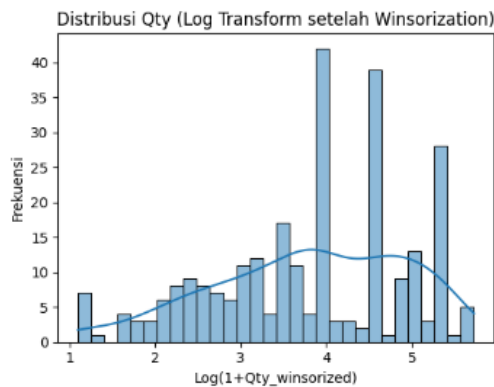
Pada gambar box plot pada atribut Qty terdapat outlier yang menunjukkan distribusi titik-titik di luar whiskers adalah outlier. Dari perhitungan statistik yang ditampilkan, diperoleh nilai Kuartil 1 (Q1) sebesar 11.0, Kuartil 3 (Q3) sebesar 100.0, dan Interquartile Range (IQR) sebesar 89.0. Batas bawah outlier diidentifikasi pada -122.5 dan batas atas pada 233.5. Terdapat 6 outlier yang teridentifikasi, dengan rincian tanggal status, jenis aksesoris.

### 3.2.3 Data Preparation

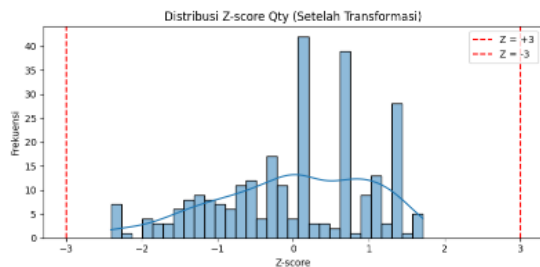
Fase persiapan data merupakan tahap penting untuk memastikan data siap digunakan dalam proses analisis. Tahap ini mencakup pemilihan data yang relevan, seperti hanya menggunakan data DARTSLIVE Live Credit Card sebanyak 270 baris, serta menghapus baris dan kolom yang tidak sesuai dengan tujuan penelitian. Pembersihan data dilakukan dengan mengganti nilai kosong menggunakan imputasi rata-rata dan penghapusan data yang tidak lengkap, serta mengganti nama kolom agar lebih mudah dipahami, misalnya email diubah menjadi outlet. Nilai pencilan (outlier) pada kolom Qty yang berasal dari transaksi Returned dengan jenis pesanan Collection tidak dihapus karena dianggap sebagai bagian dari kejadian bisnis yang valid. Sebagai gantinya, nilai-nilai tersebut ditangani menggunakan tiga metode: winsorization untuk membatasi pengaruh nilai ekstrem, log transform untuk mengurangi ketimpangan distribusi data, dan z-score untuk menstandarkan data ke dalam skala yang seragam. Proses reduksi ini difokuskan untuk menyesuaikan distribusi data agar lebih mendekati normal tanpa menghilangkan informasi penting.



Gambar 3. 4 Distribusi Qty Winsorization



Gambar 3. 5 Distribusi Qty Transformasi Log



Gambar 3. 6 Distribusi Qty Z-Score

Proses winsorization terbukti efektif dalam mengurangi pengaruh nilai-nilai outlier pada data. Meskipun demikian, distribusi data setelah winsorization masih menunjukkan ketidakseimbangan atau kemiringan (skewness) yang signifikan, sehingga diperlukan transformasi lanjutan. Penerapan transformasi log setelah winsorization berhasil memperbaiki bentuk distribusi dengan menyamakan sebaran data dan menurunkan pengaruh dari nilai-nilai besar.

### 3.2.4 Modeling

Fase pemodelan merupakan inti dalam proses data mining, di mana data yang telah dipersiapkan digunakan untuk membangun model guna menjawab permasalahan bisnis. Pada tahap ini digunakan teknik regresi linier, yang bertujuan memodelkan hubungan antara fitur numerik dan variabel target yang bersifat kontinu. Dataset yang terdiri dari variabel numerik (Qty), waktu (Status Date), dan kategorikal (Order, Status, serta Accessories & Spareparts) diolah melalui proses *feature engineering* menjadi bentuk yang sesuai untuk pemodelan. Nilai Qty yang awalnya memiliki distribusi miring ke kanan ditransformasi menjadi tiga bentuk: *Qty\_winsorized*, *Qty\_log\_transformed*, dan *Qty\_log\_zscore*, guna mengurangi pengaruh pencilan dan mendekatkan distribusi ke bentuk normal. Data waktu diubah menjadi fitur numerik seperti tahun dan bulan, sementara data kategorikal dikonversi menjadi variabel biner agar dapat digunakan dalam analisis regresi. Seluruh proses ini menghasilkan *dataframe* baru yang telah dibersihkan dan ditransformasi, siap digunakan untuk membangun model dengan asumsi hubungan linier antara fitur dan target.

Bulan	Status	Total_Qty
2024-09-01	Received	-0.999557
2024-09-01	Returned	-0.455787
2024-10-01	Received	1.592727
2024-10-01	Returned	-3.560385
2024-11-01	Received	0.238674
2024-11-01	Returned	1.708846
2024-12-01	Received	3.163136
2024-12-01	Returned	-5.010884
2025-01-01	Received	6.273638
2025-01-01	Returned	-1.249034
2025-02-01	Received	2.018377
2025-02-01	Returned	-0.335425
2025-03-01	Received	10.252924
2025-03-01	Returned	-9.193463
2025-04-01	Received	13.087314
2025-04-01	Returned	-20.491910
2025-05-01	Received	6.561341
2025-05-01	Returned	-3.555532

Gambar 3. 7 Agregasi Atribut Status dan Qty

Bulan	Received (x)	Returned (y)	x.y	x <sup>2</sup>
2024-09	-0.999557	-0.455787	0.455585	0.999114
2024-10	1.592727	-3.560385	-5.670722	2.536781
2024-11	0.238674	1.708846	0.407858	0.056965
2024-12	3.163136	-5.010884	-15.850107	10.005427
2025-01	6.273638	-1.249034	-7.835986	39.358533
2025-02	2.018377	-0.335425	-0.677013	4.073844
2025-03	10.252924	-9.193463	-94.259881	105.122460
2025-04	13.087314	-20.491910	-268.184060	171.277785
2025-05	6.516341	-3.555532	-23.169059	42.462697
Total	$\Sigma x =$ 42.143574	$\Sigma y =$ -42.139574	$\Sigma xy =$ -414.983385	$\Sigma x^2 =$ 375.8936059

**Gambar 3. 8 Agregasi Bulanan Nilai Z-Score Qty**

Status bulan	Received	Returned	Received_x_Returned	Returned_Prediksi
2024-09	-0.999557	-0.455787	0.455585	2.237177
2024-10	1.592727	-3.560385	-5.670722	-0.919726
2024-11	0.238674	1.708846	0.407858	0.729249
2024-12	3.163136	-5.010884	-15.850107	-2.832181
2025-01	6.273638	-1.249034	-7.835986	-6.620173
2025-02	2.018377	-0.335425	-0.677013	-1.438085
2025-03	10.252924	-9.193463	-94.259881	-11.466178
2025-04	13.087314	-20.491910	-268.184060	-14.917918
2025-05	6.516341	-3.555532	-23.169059	-6.915739

**Gambar 3. 9 Hasil Perhitungan Prediksi Regresi Linier dengan Python**

Model regresi linier digunakan untuk memetakan hubungan antara variabel input dan output, dengan bentuk umum persamaan sebagai berikut:

Y = jumlah kartu yang diprediksi akan dikembalikan ke gudang (unit)

X = kebutuhan kartu oleh outlet (unit)

a = konstanta (intersep)

b = koefisien regresi (pengaruh X terhadap Y)

### Menentukan Persamaan Regresi Linier

Berikut ini merupakan proses menghitung nilai a (konstanta) dan b (koefisien) variabel.

Formula untuk menentukan nilai b (koefisien regresi):

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Terlebih dahulu mencari nilai b (koefisien):

$$b = \frac{9(-414.983385) - (42.143574)(-42.139574)}{9(375.8936059) - (42.143574)^2} = \frac{-1956.97}{1606} = -1.21781$$

Formula untuk menentukan nilai a (konstanta):

$$a = \frac{\Sigma y - b \Sigma x}{n} = \frac{-42.139574 - (-1.21781)(42.143574)}{9}$$

$$a = 1.0199$$

Persamaan Regresi Linier :

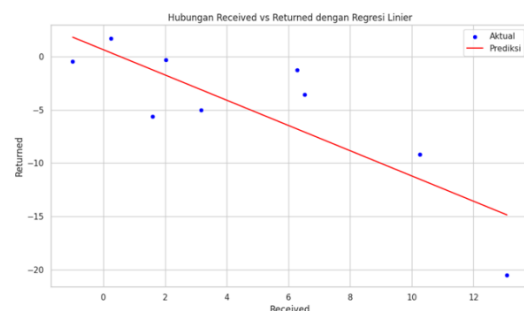
$$Y = a + bX$$

$$Y = 1.0199 + -1.2178 X$$

Berdasarkan analisis regresi linear yang dilakukan, diperoleh persamaan  $Y = 1,0199 + (-1,2178)X$ , yang menunjukkan bahwa setiap kenaikan satu unit pada variabel X (Received) akan menyebabkan penurunan sebesar 1,2178 unit pada variabel Y (Returned), dengan nilai intersep 1,0199 saat X bernilai nol. Model ini berhasil mengidentifikasi adanya defisit stok kartu LCC di PT. Mitra Media Integrasi, yang ditunjukkan melalui hasil prediksi jumlah kartu yang dikembalikan ke gudang sebesar -608 unit. Nilai negatif tersebut bukan mencerminkan pengembalian aktual, melainkan mengindikasikan kekurangan stok sebesar 608 kartu untuk memenuhi kebutuhan outlet sebanyak 500 kartu. Temuan ini memperlihatkan adanya ketidakseimbangan signifikan antara jumlah kartu yang diterima dan dikembalikan, serta memberikan dasar kuantitatif yang jelas bagi perencanaan dan pengendalian persediaan. Berdasarkan hasil prediksi regresi, perhitungan stok optimal dilakukan dengan rumus: Stok Optimal = Kebutuhan Outlet + (Kebutuhan Outlet - Returned Prediksi) Dengan substitusi nilai: Stok Optimal = 500 + (500 - (- 608)) = 500 + 1108 = 1608

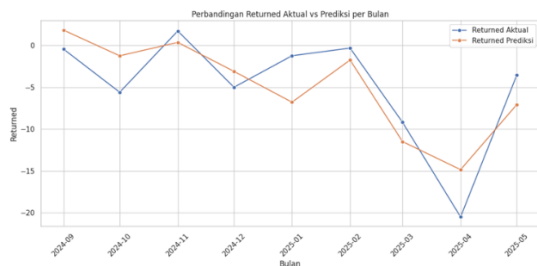
Bulan	Received	Returned Aktual	Returned Prediksi
2024-09	-0.999557	-0.455787	2.237177
2024-10	1.592727	-3.560385	-0.919726
2024-11	0.238674	1.708846	0.729249
2024-12	3.163136	-5.010884	-2.832181
2025-01	6.273638	-1.249034	-6.620173
2025-02	2.018377	-0.335425	-1.438085
2025-03	10.252924	-9.193463	-11.466178
2025-04	13.087314	-20.491910	-14.917918
2025-05	6.516341	-3.555532	-6.915739

**Gambar 3. 10 Hasil Prediksi Regresi Linier**



**Gambar 3. 11 Hubungan Regresi Linier Variabel Received dan Returned**





**Gambar 3. 12 Perbandingan Return Aktual dengan Return Prediksi**

Hubungan antara variabel *Received* dan *Returned* dianalisis menggunakan model regresi linier, di mana hasil visualisasi menunjukkan adanya korelasi negatif antara keduanya. Peningkatan nilai *Received* cenderung diikuti oleh penurunan nilai *Returned*, sebagaimana ditunjukkan oleh kemiringan garis regresi yang mengarah ke bawah. Evaluasi kinerja model dilakukan dengan membandingkan nilai aktual dan nilai prediksi dari waktu ke waktu. Meskipun terdapat perbedaan pada beberapa periode, secara umum model mampu mengikuti pola tren data aktual. Hal ini menunjukkan bahwa model memiliki kemampuan dalam memprediksi arah perubahan data, meskipun akurasi tidak selalu konsisten di setiap waktu.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1 Evaluation

Pada tahap ini, sebagian besar proses data mining telah diselesaikan. Model regresi linier berhasil dibangun dan dievaluasi, serta secara teknis mampu mempelajari hubungan antara jumlah barang yang diterima (*Received*) dan dikembalikan (*Returned*). Berdasarkan kriteria performa teknis, model telah memenuhi indikator keberhasilan minimum, seperti koefisien determinasi ( $R^2$ ), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Namun demikian, keberhasilan teknis tidak dapat menjadi satu-satunya ukuran; pencapaian terhadap tujuan bisnis yang telah ditetapkan sejak awal proyek juga menjadi pertimbangan penting.

Model regresi linier dinilai telah memenuhi tujuan bisnis, yaitu memprediksi jumlah barang yang dikembalikan berdasarkan data historis barang yang diterima. Evaluasi dilakukan menggunakan tiga jenis transformasi data: *Z-Score*, *Log Transform*, dan *Log Winsorized*, untuk mengetahui pengaruh transformasi terhadap kinerja model. Berikut adalah hasil evaluasi kuantitatif terhadap model:

**Tabel 3. 1 Matrix Evaluasi**

Matrix	Nilai
MSE	10.8358
RMSE	3.2918
$R^2$	0.7308

Nilai *Mean Squared Error* (MSE) sebesar 10.8358 menunjukkan tingkat kesalahan prediksi yang cukup rendah dan masih dapat diterima, sedangkan *Root Mean Squared Error* (RMSE) sebesar 3.2918 menunjukkan deviasi rata-rata prediksi terhadap nilai aktual sekitar 3,29 unit. Koefisien determinasi ( $R^2$ ) sebesar 0.7308 menunjukkan bahwa 73,08% variasi pada variabel *Returned* dapat dijelaskan oleh variabel *Received*, menandakan model memiliki kemampuan prediksi yang cukup baik.

Proses pemodelan dilakukan secara sistematis, mulai dari pemilihan data dan fitur yang relevan, transformasi data untuk mengatasi outlier dan distribusi tidak normal, hingga pembangunan dan evaluasi model regresi linier menggunakan metrik statistik. Transformasi data terbukti berpengaruh signifikan terhadap akurasi dan stabilitas model, sehingga pemilihan metode transformasi menjadi langkah penting. Berdasarkan hasil evaluasi, jika akurasi menjadi prioritas utama, transformasi *Z-Score* dapat digunakan dengan catatan hasil prediksi perlu dikembalikan ke bentuk awal agar mudah dipahami secara bisnis. Namun, jika kemudahan interpretasi lebih diutamakan, transformasi logaritmik bisa menjadi alternatif meskipun dengan akurasi yang sedikit lebih rendah.

##### 4.2 Deployment

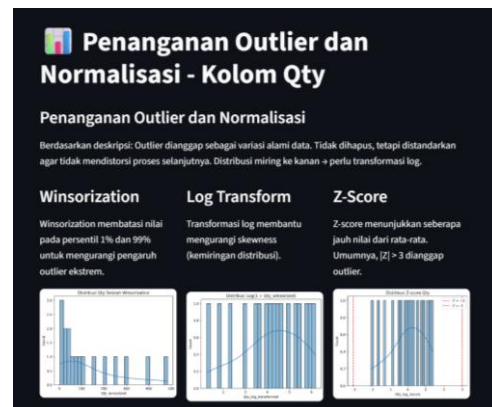
Fase implementasi merupakan tahap akhir dari metodologi CRISP-DM, di mana model dan insight yang dihasilkan digunakan untuk memberikan nilai nyata bagi perusahaan. Implementasi mencakup integrasi model ke dalam sistem informasi yang digunakan serta penyampaian hasil analisis kepada pemangku kepentingan untuk mendukung pengambilan keputusan. Sistem prediksi dikembangkan menggunakan *framework Streamlit*, dan menyajikan berbagai visualisasi interaktif untuk membantu pengguna memahami data dan hasil prediksi. Berikut adalah tampilan dashboard yang telah dikembangkan:

### Membaca data\_darts.csv

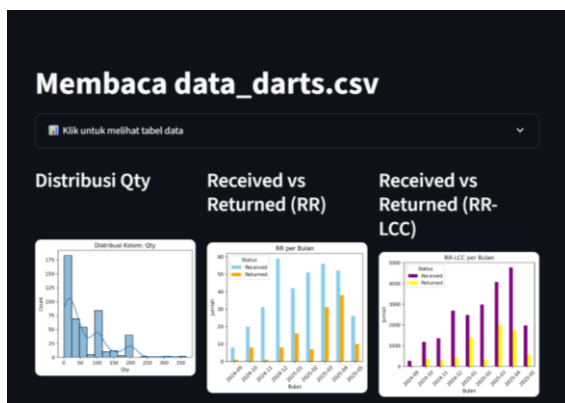
Klik untuk melihat tabel data

	email	Accessories & Spareparts	Order	Status	Status Date
0	Dewa Sport & Bar	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	23-Dec-2024
1	Gemu	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	23-Dec-2024
2	Gen's Bar & BBQ Resto	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	23-Dec-2024
3	The Cue Billiard	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	23-Dec-2024
4	FireWok Eating House	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	4-Oct-2024
5	BuddyPool	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	26-Nov-2024
6	Darts Hub Jakarta	[DL2-7011_TBQ] Throwline Sticker	Shop Request	Received	4-Oct-2024

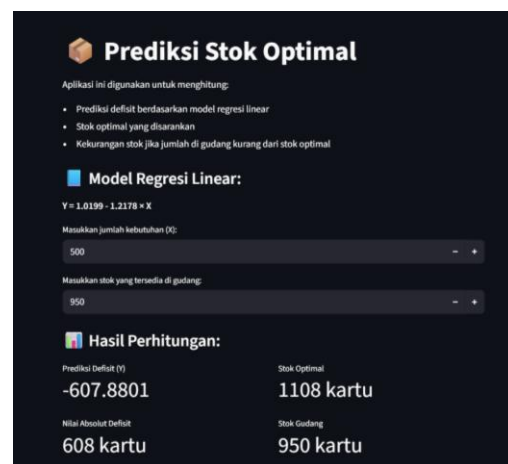
Gambar 4. 1 Dashboard Analisis Prediksi



Gambar 4. 4 Penanganan Outlier dan Normalisasi



Gambar 4. 2 Distribusi Data



Gambar 4. 5 Prediksi Stok Optimal



Gambar 4. 3 Distribusi Frekuensi Qty

Model regresi linier yang dibangun untuk memprediksi stok kartu LCC di PT. Mitra Media Integrasi telah menunjukkan performa yang cukup baik, namun tetap memerlukan pemantauan rutin, misalnya setiap bulan atau kuartal, dengan menghitung ulang metrik evaluasi seperti MSE, RMSE, dan  $R^2$  menggunakan data terbaru. Jika performa menurun ( $R^2 < 0,60$ ) atau terjadi perubahan signifikan pada pola data, perlu dilakukan pelatihan ulang model atau penambahan variabel input, dan apabila perubahan terlalu besar, pembangunan ulang model disarankan. Penelitian ini bertujuan mengatasi inefisiensi rantai pasok akibat ketidaksesuaian data stok dengan pendekatan data mining dan metodologi CRISP-DM. Hasil menunjukkan model cukup akurat, namun terdapat defisit pada beberapa periode yang mengindikasikan adanya variabel eksternal yang belum dianalisis. Oleh karena itu, model ini diusulkan sebagai alat bantu pengambilan keputusan operasional dan dasar pengembangan sistem prediksi stok, dengan rekomendasi

pengujian algoritma alternatif seperti time series forecasting dan validasi data secara rutin untuk meningkatkan akurasi di masa mendatang.

## 5. KESIMPULAN

Ketidaksesuaian antara data kartu yang diterima (*Received*) dan dikembalikan (*Returned*) menunjukkan adanya ketidakseimbangan stok kartu LCC yang berdampak pada efisiensi manajemen rantai pasok di PT. Mitra Media Integrasi. Untuk menangani hal tersebut, digunakan algoritma regresi linier dengan nilai koefisien determinasi ( $R^2$ ) sebesar 0,7308, yang menunjukkan bahwa 73,08% variasi stok dapat dijelaskan oleh data historis. Prediksi untuk bulan Juni menunjukkan potensi defisit sebanyak 608 kartu, sehingga stok optimal diperkirakan mencapai 1.608 kartu. Hasil ini membuktikan bahwa regresi linier dapat dimanfaatkan sebagai alat bantu dalam merencanakan persediaan di masa mendatang secara lebih akurat dan efisien. Evaluasi terhadap variabel data yang digunakan menunjukkan bahwa meskipun data *Received* cukup berkontribusi dalam membentuk model prediksi yang akurat (dengan MSE = 10.8358 dan RMSE = 3.2918 yang menandakan kesalahan prediksi rendah), regresi linier yang diterapkan baik melalui *Excel* maupun *Python* menghasilkan perbedaan yang sangat kecil, yakni sekitar 0,0006%, yang kemungkinan besar disebabkan oleh pembulatan angka desimal. Hal ini mengindikasikan bahwa meskipun data yang tersedia tergolong cukup dari segi periode, belum sepenuhnya mencerminkan variabel-variabel yang mempengaruhi ketersediaan stok secara signifikan, khususnya terkait pembaruan sistem dan perubahan

pola permintaan dari outlet. Oleh karena itu, akurasi prediksi dapat ditingkatkan melalui penambahan variabel yang lebih relevan serta penggunaan metode analisis yang lebih sesuai dan mampu menyesuaikan dengan dinamika kondisi operasional.

## DAFTAR PUSTAKA

- [1] Aryudha, G., & Hasibuan, W. R. (2024). *Linear Regression Analysis in Predicting the Amount of Stock of HP Sparepart Goods in GMT*. 4(1).
- [2] Rizky, F., Syahra, Y., Mariami, I., & \_ Y. (2019). Implementasi Data Mining Untuk Memprediksi Target Pemakaian Stok Barang Menggunakan Metode Regresi Linier Berganda. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 18(2), 167. <https://doi.org/10.53513/jis.v18i2.156>
- [3] Sari, N. (2022). *ISSN : 2809-7491*. 2, 85–91.
- [4] Sarimole, F. M., & Hakim, L. (2024). Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang. *Jurnal Sains Dan Teknologi*, 5(3), 846–854. <https://doi.org/10.55338/saintek.v5i3.2709>
- [5] Studi, P., Informatika, T., & Cirebon, S. I. (2024). *PENERAPAN DATA MINING DALAM PREDIKSI PRODUKSI BERAS MENGGUNAKAN METODE REGRESI LINEAR*. 8(2), 2020–2026.
- [6] Tadayonrad, Y., & Ndiaye, A. B. (2023). A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality. *Supply Chain Analytics*, 3(June), 100026. <https://doi.org/10.1016/j.sca.2023.100026>
- [7] Wahyudi, T., Bebriani, S., Teknik, P., Sekolah, I., Ilmu, T., Cipta, K., & Informatika, K. (2024). 1,2 1, 2. 7, 1588–1594.