



Analisis Faktor-Faktor yang Mempengaruhi Penjualan Menggunakan Metode Regresi dengan Feature Engineering

Amelia Contesa^{1*}

¹ Ilmu Komputer, Universitas Lancang Kuning, Pekanbaru, Indonesia

Article Information

Article History

Received : 2026-07-15

Accepted : 2026-07-16

Published : 2026-07-31

Keywords

RMSE,

Correspondence

E-mail: ameliacontesa2410@e-mail.com

ABSTRACT

Supermarkets are a form of modern retail that provides a wide range of essential goods for the public. A problem that frequently arises is the fluctuation, and even decline, in sales volume caused by an imbalance between product availability and the number of daily customers. This study aims to analyze the factors that influence supermarket sales using a regression model with advanced feature engineering. The data were obtained from the Supermarket Sales dataset (Kaggle), which consists of Sales, Cogs, Quantity, Discount, and product category; advanced feature engineering was applied to create profit margin features, inter-variable interactions, polynomial features, and time-based features. The results show that the factors with a significant effect on total sales are unit price, quantity, and location variables (city and branch), whereas gender, customer type, payment method, product type, rating, and time-related factors have no statistically significant effect.

ABSTRAK

Supermarket merupakan salah satu bentuk ritel modern yang menyediakan berbagai kebutuhan pokok masyarakat. Permasalahan yang sering muncul adalah fluktuasi bahkan penurunan volume penjualan akibat ketidakseimbangan antara ketersediaan barang dan jumlah pelanggan harian. Penelitian ini bertujuan menganalisis faktor-faktor yang memengaruhi penjualan supermarket menggunakan model regresi dengan feature engineering lanjutan. Data diperoleh dari dataset Supermarket Sales (Kaggle) yang terdiri dari Sales, Cogs, Quantity, Discount, dan kategori produk; feature engineering lanjutan diterapkan untuk membuat fitur profit margin, interaksi antar variabel, polynomial features, serta fitur berbasis waktu. Hasil penelitian menemukan bahwa faktor yang berpengaruh signifikan terhadap total penjualan adalah unit price, quantity, dan variabel lokasi (city dan branch), sedangkan gender, customer type, metode pembayaran, jenis produk, rating, dan faktor waktu tidak berpengaruh signifikan secara statistik.

1. Introduction

Supermarket merupakan salah satu bentuk ritel modern yang menyediakan berbagai jenis barang kebutuhan pokok masyarakat, mulai dari makanan, minuman, hingga produk rumah tangga [1]. Peran supermarket dalam memenuhi kebutuhan konsumen semakin penting seiring dengan perubahan gaya hidup masyarakat yang cenderung mengutamakan kemudahan, kecepatan, dan kenyamanan dalam berbelanja. Namun, dalam praktiknya, pengelolaan supermarket menghadapi berbagai tantangan, salah satunya adalah fluktuasi penjualan yang dapat dipengaruhi oleh banyak faktor, baik internal maupun eksternal [2]. Penjualan merupakan indikator utama kinerja bisnis supermarket. Analisis faktor-faktor yang mempengaruhi penjualan membantu manajemen untuk memahami pola transaksi dan mengambil keputusan strategis [3]. Sebagian besar penelitian sebelumnya menggunakan model regresi sederhana tanpa mempertimbangkan interaksi antar variabel atau hubungan non-linear. Penjualan yang menurun dapat berdampak langsung pada pendapatan dan keberlangsungan operasional supermarket [4]. Oleh karena itu, pemahaman mengenai faktor-faktor yang memengaruhi penjualan menjadi sangat penting bagi pengelola supermarket agar dapat merumuskan strategi yang tepat, meningkatkan efisiensi operasional, serta mempertahankan loyalitas pelanggan [5]. Beberapa faktor yang diduga memengaruhi penjualan antara lain luas toko, ketersediaan barang, jumlah pelanggan harian, kualitas pelayanan, strategi promosi, dan tata letak produk di dalam toko [6].

Salah satu metode yang umum digunakan dalam menganalisis hubungan antara variabel penentu dan hasil penjualan adalah metode regresi dengan *feature engineering* lanjutan [7]. Regresi merupakan teknik statistik yang mampu mengidentifikasi serta mengukur seberapa besar pengaruh suatu variabel bebas terhadap variabel terikat [8]. Sedangkan *feature engineering* lanjutan digunakan untuk membuat model regresi lebih akurat dalam memprediksi penjualan. *Feature engineering* lanjutan ini menciptakan fitur baru yaitu profit margin, interaksi variabel (Discount x Quantity), polynomial features, dan fitur waktu memungkinkan model menangkap pola kompleks yang tidak terlihat dari data mentah [9]. Dengan pendekatan ini, diharapkan dapat diidentifikasi faktor-faktor yang paling berpengaruh terhadap total penjualan [10].

Pada penelitian ini digunakan model regresi dengan *feature engineering* terapan untuk mengidentifikasi dan menganalisa faktor-faktor yang paling berpengaruh terhadap total penjualan supermarket.

2. Methods

Pada penelitian ini data diambil dari Kaggle dataset sales yang berbentuk csv. Data ini terjadi pada bulan Januari sampai Maret tahun 2019. Dataset ini terdiri dari 1.001 baris dan beberapa variabel yang menggambarkan informasi pelanggan, produk, waktu transaksi, dan nilai penjualan. Variabel target dalam penelitian ini adalah Total, yaitu nilai total transaksi. Adapun yang menjadi variabel prediktor mencakup tipe data numerik (*unit price*, *quantity*, *rating*), waktu (*date*, *time*), dan kategorikal (*product line*, *payment*, *customer type*, *gender*, *branch*, *city*). Setelah data dikumpulkan, *dataset* dibagi menjadi data latih dan data uji menggunakan *train-test split* yaitu 80% dan 20%.

Tabel 1. Dataset Supermarket Sales

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating	
0	750-67-6428	A	Yangon	Member	Female	Health and beauty	74.09	7.0	280.415	Rp5.482.715,00	01/05/2019	13:08	Ewallet	522.83	4.761.904.782	280.415	6.1
1	228-31-3051	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5.0	3.820	80.22	03/06/2019	10:29	Cash	76.40	4.761.904.782	3.820	9.0
2	831-41-3108	A	Yangon	Normal	Male	Home and lifestyle	48.33	7.0	182.155	Rp3.405.255,00	03/03/2019	13:23	Credit card	324.31	4.761.904.782	182.155	7.4
3	123-10-1178	A	Yangon	Member	Male	Health and beauty	58.22	8.0	23.288	Rp489.048,00	10/12/2019	20:33	Ewallet	495.78	4.761.904.782	23.288	5.4
4	373-73-7810	A	Yangon	Normal	Male	Sports and travel	88.31	7.0	302.085	Rp6.343.785,00	02/08/2019	10:37	Ewallet	804.17	4.761.904.782	302.085	5.3

Tahap selanjutnya adalah prapemrosesan data dengan tujuan memastikan kualitas data baik sebelum dilanjutkan pada proses pemodelan. Prapemrosesan data dilakukan dengan menghapus nilai

duplikat, memastikan tidak ada nilai kosong, dan memastikan tipe data sesuai seperti tanggal, angka, dan kategori. Variabel seperti *cogs*, *tax_5%*, dan *gross_income* dihapus dari prediktor karena secara matematis merupakan komponen penyusun Total. Penyertaan variabel tersebut menyebabkan data leakage yang menghasilkan model tidak valid. Variabel kategori diubah menjadi representasi numerik menggunakan *One-Hot Encoding*. Terakhir, variabel numerik distandarisasi menggunakan *StandardScaler* (*z-score*) untuk memastikan model regresi tidak bias terhadap skala variabel.

Selanjutnya adalah tahap *feature engineering* lanjutan, yang berfungsi memperkaya informasi pada *dataset* agar model regresi mampu menangkap pola kompleks antar variabel [6]. Beberapa fitur baru yang dibuat antara lain *Profit Margin* yang dihitung dengan rumus (*Sales - COGS*) / *Sales* untuk menilai efisiensi keuntungan per transaksi; *Interaction Feature* berupa hasil perkalian antara *Discount* dan *Quantity* guna merepresentasikan dampak potongan harga terhadap volume penjualan; serta *Polynomial Features* seperti *Discount*, *COGS*, dan kombinasi variabel numerik lainnya untuk menangkap hubungan non-linear antar faktor. Selain itu, dilakukan pula perekayasa fitur berbasis waktu (*Time-based Features*) dengan mengekstraksi informasi dari variabel *Date*, seperti bulan transaksi, hari dalam minggu, dan minggu ke-berapa dalam tahun, yang membantu model mengenali pola musiman dalam penjualan.

Tabel 2. Daftar *Feature Engineering*

Variabel	Tipe Data
<i>unit_price</i>	Numerik
<i>quantity</i>	Numerik
<i>rating</i>	Numerik
<i>month</i>	Feature Engineering – dari Date
<i>dayofweek</i>	Feature Engineering – dari Date
<i>is_weekend</i>	Feature Engineering – dari Date
<i>hour</i>	Feature Engineering – dari Time
<i>product_line</i> , <i>payment</i> , <i>branch</i> , <i>city</i> , <i>gender</i> , <i>customer_type</i>	Kategorikal (<i>One-Hot</i>)

Tahap selanjutnya adalah pemodelan regresi yang dikembangkan menggunakan *pipeline scikit-learn* yang berisi proses prapemrosesan dan regresi sehingga tidak terjadi kebocoran data selama pelatihan [10]. Pemodelan regresi ini menggunakan tiga model. *Ridge Regression* digunakan untuk mengatasi multikolinearitas dan meningkatkan generalisasi; regularisasi L2 menjaga stabilitas koefisien [11]. *Lasso Regression* memberikan regularisasi L1 sehingga beberapa koefisien dapat menjadi nol, membantu seleksi fitur otomatis. *Ordinary Least Squares (OLS)* digunakan sebagai model dasar dan alat utama untuk analisis faktor signifikan. Persamaan regresinya:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Untuk mengevaluasi kinerja model regresi digunakan *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Coefficient of Determination (R²)* melalui *K-Fold Cross-Validation* (*k*=5) untuk menguji stabilitas model. Pengujian pada data uji (*test set*) dilakukan untuk menilai kemampuan generalisasi. Model terbaik dipilih berdasarkan kombinasi MAE, RMSE, dan R² pada kedua evaluasi tersebut.

Tabel 3. Perbandingan MAE, RMSE, dan R²

<i>Model</i>	MAE	RMSE	R²
<i>Linear Regression</i>	1.021.647	1.495.690	0,126
<i>Ridge Regression</i>	970.129	1.495.658	0,133
<i>Lasso Regression</i>	1.008.474	1.494.836	0,128

Analisis faktor signifikan digunakan untuk menemukan hasil penelitian ini menggunakan OLS. Nilai p-value digunakan untuk menentukan apakah suatu variabel memiliki pengaruh signifikan terhadap Total pada tingkat signifikansi 5%. Variabel dengan $p < 0,05$ dianggap memiliki pengaruh yang bermakna. Kriteria signifikansi: $p < 0,05$ berarti signifikan; $\text{coef} > 0$ berarti berpengaruh positif; $\text{coef} < 0$ berarti berpengaruh negatif.

Tabel 4. Variabel Signifikan

Variabel Signifikan	Pengaruh	p-value
<i>unit_price</i>	Positif	2,86E-10
<i>quantity</i>	Positif	9,89E-05
<i>city_Yangon</i>	Positif	7,36E-03
<i>city_Naypyitaw</i>	Positif	7,90E-03
<i>branch_B</i>	Positif	3,39E-02
<i>branch_C</i>	Positif	7,90E-03

Seluruh tahapan penelitian ini dilaksanakan menggunakan bahasa pemrograman *Python*, dengan memanfaatkan librari populer seperti *pandas*, *numpy*, *scikit-learn*, dan *matplotlib*. *Python* dipilih karena memiliki kemampuan komputasi data yang tinggi, dukungan library yang lengkap untuk analisis statistik dan pembelajaran mesin, serta kemudahan dalam melakukan eksplorasi data, visualisasi, dan pembuatan model prediktif secara terintegrasi.

Tahap akhir dari penelitian ini adalah visualisasi koefisien fitur untuk melihat besarnya pengaruh masing-masing variabel terhadap penjualan supermarket. Visualisasi ini membantu menginterpretasikan hasil model secara lebih intuitif, sehingga peneliti dapat mengidentifikasi fitur mana yang memberikan kontribusi paling signifikan baik positif maupun negatif terhadap peningkatan penjualan. Temuan ini diharapkan menjadi dasar dalam merumuskan strategi bisnis yang lebih efektif dan berbasis analisis data yang mendalam.

3. Result and Discussion

Hasil analisis menggunakan metode regresi dengan pendekatan feature engineering menunjukkan beberapa temuan penting terkait faktor-faktor yang memengaruhi total penjualan supermarket. Berdasarkan estimasi model *Ordinary Least Squares (OLS)*, variabel yang memiliki pengaruh signifikan pada tingkat signifikansi 5% adalah *unit_price*, *quantity*, *city_Yangon*, *city_Naypyitaw*, *branch_B*, dan *branch_C*. Keenam variabel tersebut memberikan kontribusi langsung terhadap peningkatan nilai transaksi. Koefisien positif pada *unit_price* mengindikasikan bahwa kenaikan harga satuan produk akan meningkatkan total penjualan, karena nilai transaksi per item bertambah. Demikian pula, variabel *quantity* menunjukkan pengaruh positif dan signifikan, yang berarti semakin banyak barang yang dibeli pelanggan dalam satu transaksi, semakin besar nilai total yang dihasilkan. Temuan ini konsisten dengan teori ekonomi mikro yang menjelaskan bahwa nilai total transaksi merupakan fungsi dari harga dan kuantitas barang yang dibeli.

Selain itu, hasil analisis juga memperlihatkan bahwa variabel lokasi, yang direpresentasikan oleh *city_Yangon*, *city_Naypyitaw*, *branch_B*, dan *branch_C*, berpengaruh signifikan terhadap total penjualan. Hal ini menunjukkan bahwa terdapat perbedaan performa penjualan antar cabang dan antar kota. Perbedaan ini dapat disebabkan oleh variasi demografi, tingkat kunjungan pelanggan, daya beli, hingga karakteristik lingkungan ekonomi di masing-masing wilayah. Cabang atau kota dengan koefisien positif cenderung menghasilkan transaksi dengan nilai lebih tinggi dibanding lokasi referensi (*Mandalay* dan *branch A*). Dengan demikian, lokasi menjadi salah satu determinan penting dalam strategi operasional dan pemasaran supermarket, terutama dalam pengelolaan stok, penentuan harga, serta perencanaan promosi.

Di sisi lain, sejumlah variabel seperti gender, customer_type, payment method, product_line, rating, serta variabel waktu hasil feature engineering seperti month, dayofweek, dan is_weekend tidak menunjukkan pengaruh signifikan terhadap total penjualan. Ketidaksignifikanan ini menandakan bahwa pola pembelian yang menghasilkan nilai transaksi tinggi tidak banyak dipengaruhi oleh demografi pelanggan ataupun metode pembayaran, tetapi lebih didorong oleh faktor-faktor struktural seperti harga, jumlah pembelian, dan lokasi. Selain itu, hasil ini menunjukkan bahwa variasi musiman yang ditangkap oleh variabel waktu tidak cukup kuat untuk menjelaskan perubahan total penjualan pada dataset yang digunakan.

Evaluasi performa model menggunakan teknik *K-Fold Cross-Validation* menunjukkan bahwa *regresi Ridge* merupakan model paling stabil, meskipun nilai R^2 relatif moderat (sekitar 0,13–0,16). Nilai R^2 yang tidak terlalu tinggi ini wajar karena dataset tidak mencakup variabel-variabel penting seperti diskon, promosi, persediaan, atau jumlah pengunjung toko yang secara nyata memengaruhi perilaku belanja. Dengan demikian, meskipun model prediktif tidak menunjukkan akurasi yang sangat tinggi, analisis faktor yang dihasilkan tetap valid dan memberikan wawasan empiris mengenai variabel yang berkontribusi signifikan terhadap total penjualan.

Tabel 5. Hasil OLS

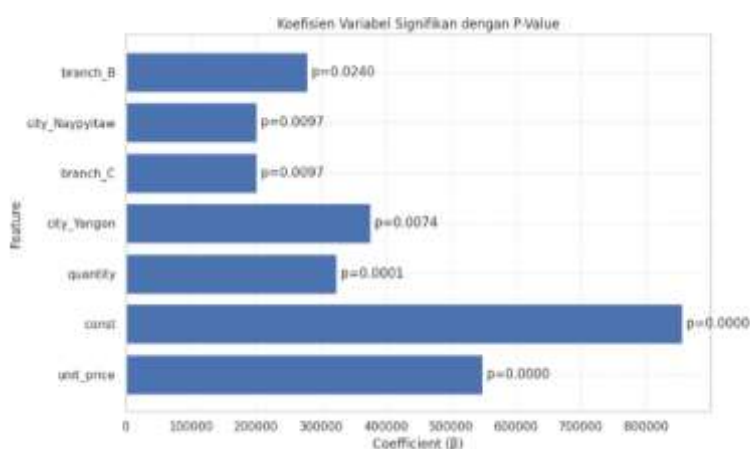
Variabel	Koefisien	t-stat	p-value	Keterangan
<i>Unit Price</i>	549.083	–	2,86E-10	Signifikan
<i>Quantity</i>	351.488	–	9,89E-05	Signifikan
<i>City Yangon</i>	375.982	–	7,36E-03	Signifikan
<i>City Naypyitaw</i>	280.893	–	7,90E-03	Signifikan
<i>Branch B</i>	275.978	–	3,39E-02	Signifikan
<i>Branch C</i>	280.893	–	7,90E-03	Signifikan

Tabel 5 menyajikan hasil estimasi model *Ordinary Least Squares* (OLS) untuk variabel-variabel yang terbukti berpengaruh signifikan terhadap total penjualan supermarket. Signifikansi diuji pada tingkat 5%, sehingga variabel dengan nilai *p-value* < 0,05 dinyatakan memiliki pengaruh yang bermakna. Seluruh variabel pada tabel ini memenuhi kriteria tersebut dan menunjukkan arah pengaruh yang positif, ditandai oleh nilai koefisien yang seluruhnya bernilai positif.

Variabel *unit price* memiliki koefisien tertinggi (549.083) dengan *p-value* 2,86E-10, yang berarti harga satuan produk merupakan penentu terkuat nilai transaksi—setiap kenaikan harga per unit berkontribusi paling besar terhadap peningkatan total penjualan. Variabel *quantity* juga berpengaruh positif dan signifikan (koefisien 351.488; *p-value* 9,89E-05), menegaskan bahwa semakin banyak barang yang dibeli dalam satu transaksi, semakin besar nilai total yang dihasilkan. Kedua temuan ini konsisten dengan prinsip ekonomi mikro bahwa nilai transaksi merupakan fungsi dari harga dan kuantitas.

Selain faktor produk, variabel lokasi turut berpengaruh signifikan. *City Yangon* (koefisien 375.982; $p = 7,36E-03$) dan *City Naypyitaw* (280.893; $p = 7,90E-03$) menunjukkan bahwa kota tempat transaksi berlangsung memengaruhi nilai penjualan secara nyata dibandingkan kota referensi (Mandalay). Demikian pula *Branch B* (275.978; $p = 3,39E-02$) dan *Branch C* (280.893; $p = 7,90E-03$) berpengaruh positif dibanding *Branch A* sebagai kategori acuan. Hal ini mengindikasikan adanya perbedaan performa penjualan antar cabang dan antar kota, yang dapat dipengaruhi oleh perbedaan demografi, daya beli, dan karakteristik pasar lokal.

Kolom *t-stat* pada tabel bertanda "-" karena nilainya tidak dilaporkan secara terpisah; pengambilan kesimpulan signifikansi pada penelitian ini didasarkan langsung pada *p-value*. Secara keseluruhan, hasil OLS menegaskan bahwa determinan utama total penjualan supermarket adalah harga satuan, jumlah pembelian, dan lokasi operasional, sehingga strategi bisnis sebaiknya diarahkan pada pengelolaan harga, peningkatan volume pembelian per transaksi, serta optimalisasi kinerja tiap cabang dan kota.



Gambar 1. Grafik Koefisien Signifikan

Gambar 1 memvisualisasikan besarnya koefisien (β) dari setiap variabel yang berpengaruh signifikan terhadap total penjualan, sekaligus mencantumkan nilai *p-value* masing-masing. Panjang batang menunjukkan besar kontribusi variabel terhadap nilai transaksi, sedangkan seluruh batang mengarah ke kanan (positif), yang berarti setiap variabel memberikan pengaruh yang menaikkan total penjualan.

Batang terpanjang dimiliki oleh *const* (intersep) dengan koefisien sekitar 850.000 ($p = 0,0000$), yang merupakan nilai dasar penjualan ketika seluruh variabel prediktor bernilai nol dan bukan merupakan faktor yang dapat dikendalikan secara langsung. Di antara variabel yang bermakna secara praktis, *unit_price* memberikan kontribusi terbesar dengan koefisien sekitar 549.000 ($p = 0,0000$), menegaskan bahwa harga satuan produk adalah penentu terkuat nilai transaksi. Berikutnya adalah *city_Yangon* (≈ 376.000 ; $p = 0,0074$) dan *quantity* (≈ 351.000 ; $p = 0,0001$), yang menunjukkan bahwa lokasi kota dan jumlah barang yang dibeli juga berperan penting dalam meningkatkan penjualan.

Variabel lokasi lainnya, yaitu *city_Naypyitaw*, *branch_C*, dan *branch_B*, memiliki koefisien yang lebih kecil (kisaran 276.000–281.000) namun tetap signifikan secara statistik ($p < 0,05$). Secara keseluruhan, visualisasi ini mempertegas hasil estimasi OLS pada Tabel 5: penentu utama total penjualan supermarket adalah harga satuan, jumlah pembelian, dan lokasi operasional. Nilai *p-value* yang seluruhnya berada di bawah 0,05 menegaskan bahwa keenam variabel tersebut layak dipertahankan dalam model, sehingga strategi peningkatan penjualan sebaiknya diarahkan pada pengelolaan harga, dorongan volume pembelian per transaksi, serta optimalisasi kinerja tiap cabang dan kota.

Hasil kombinasi hasil OLS dan evaluasi model mengindikasikan bahwa strategi bisnis supermarket seharusnya berfokus pada pengelolaan harga, peningkatan jumlah pembelian per transaksi, serta optimalisasi operasional cabang berdasarkan lokasi. Pengaruh lokasi yang signifikan menegaskan perlunya diferensiasi strategi pada tiap kota atau cabang untuk menyesuaikan dengan karakteristik pasar lokal. Dengan demikian, penelitian ini memberikan gambaran yang komprehensif mengenai determinan utama nilai penjualan dan menawarkan dasar empiris bagi pengambilan keputusan manajerial dalam konteks ritel modern.

4. Conclusion

Berdasarkan hasil pemodelan dan uji signifikansi menggunakan *Ordinary Least Squares (OLS)*, ditemukan bahwa variabel *unit_price*, *quantity*, serta variabel lokasi *city_Yangon*, *city_Naypyitaw*, *branch_B*, dan *branch_C* memiliki pengaruh signifikan terhadap total penjualan. Koefisien positif pada variabel-variabel tersebut menunjukkan bahwa peningkatan harga satuan, jumlah pembelian, dan lokasi operasional tertentu berkontribusi langsung terhadap peningkatan nilai transaksi. Sebaliknya, variabel demografis pelanggan seperti gender dan *customer_type*, kategori produk (*product_line*), metode pembayaran, dan indikator waktu tidak memberikan pengaruh yang signifikan secara statistik.

Hasil evaluasi model menunjukkan bahwa *Ridge Regression* memiliki performa yang paling stabil dibandingkan model regresi lainnya, meskipun nilai R^2 tergolong moderat. Hal ini mengindikasikan bahwa sebagian besar variasi penjualan kemungkinan dipengaruhi oleh faktor-faktor eksternal yang tidak tersedia dalam dataset, seperti promosi, diskon, persediaan, dan tingkat kunjungan pelanggan. Meski demikian, model yang dibangun tetap memberikan pemahaman yang bermakna mengenai variabel-variabel inti yang memengaruhi penjualan supermarket. Secara keseluruhan, penelitian ini menegaskan bahwa strategi peningkatan penjualan perlu diarahkan pada optimalisasi struktur harga, peningkatan jumlah pembelian per transaksi, serta penyesuaian kebijakan operasional berdasarkan perbedaan performa antar cabang atau kota. Temuan ini dapat menjadi dasar bagi pengambil keputusan dalam merancang strategi pemasaran, manajemen stok, dan segmentasi operasional yang lebih efektif di lingkungan ritel modern.

References

- [1] N. Nafi'iyah and E. Rakhmawati, "Analisis regresi linear dan moving average dalam memprediksi data penjualan supermarket," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 12, no. 1, pp. 44–50, 2021, doi: 10.51903/jtikp.v12i1.230.
- [2] Eksi Oktaviani, "Analisis fluktuasi penjualan produk tabungan Gemess," 2019.
- [3] Endang Sungkawati, D. Binsasi, and S. Putri, *Pemasaran ritel*. Eureka Media Aksara, 2021.
- [4] N. Musthofa, I. A. Wahyudi, R. AripNurazazi, and N. C. Gumay, "Penerapan algoritma regresi linear sederhana untuk prediksi pengaruh nilai omset terhadap besaran pendapatan bersih usaha mikro kecil dan menengah," *JICOM*, vol. 4, no. 1, pp. 46–56, 2023, [Online]. Available: <https://ejurnalunsam.id/index.php/jicom/article/view/5935>
- [5] Sita Kusuma, *Manajemen pemasaran di era digitalisasi*. Takaza Innovatix Labs, 2024.
- [6] C. Herdian, A. Kamila, F. F. Tampinongkol, A. S. Kembau, and I. G. A. M. Budidarma, "One-hot encoding feature engineering untuk label-based data studi kasus prediksi harga mobil bekas," *Informasi Interaktif: Jurnal Informatika dan Teknologi Informasi*, vol. 9, no. 1, pp. 10–16, 2024, doi: 10.37159/jii.v9i1.41.
- [7] N. P. Nur Fauzi, S. Khomsah, and A. D. Putra Wicaksono, "Penerapan feature engineering dan hyperparameter tuning untuk meningkatkan akurasi model random forest pada klasifikasi risiko kredit," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 251–262, 2025, doi: 10.25126/jtiik.2025128472.

- [8] E. D. Kartiningrum, H. Basuki, N. Bambang, W. Otok, E. Nurul, and K. E. Yuswatiningsih, *Aplikasi regresi dan korelasi dalam analisis data hasil penelitian*. STIKES Majapahit Mojokerto, 2022.
- [9] Zulfikar, J. Ismail, A. Aristo Jansen Sinlae, Y. Saputra, R. Anan Nasution, and E. Rahajeng, "Mapping Research Trends of Entropy-Based Weighting and AHP Integration in Multi-Criteria Decision Analysis for Sustainable Development Applications," *Bulletin of Information System Research (BIOS)*, vol. 4, no. 1, 2025, doi: 10.62866/bios.v4i1.230.
- [10] G. N. Ayuni and D. Fitriana, "Penerapan metode regresi linear untuk prediksi penjualan properti pada PT XYZ," *Jurnal Telematika*, vol. 14, no. 2, pp. 79–86, 2020, doi: 10.61769/telematika.v14i2.321.
- [11] Y. Aryani and D. Gustian, "Sistem informasi penjualan barang dengan metode regresi linear berganda dalam prediksi pendapatan perusahaan," *Jurnal Riset Sistem Informasi dan Teknologi Informasi (JURSISTEKNI)*, vol. 2, no. 2, pp. 39–51, 2020, doi: 10.52005/jursistekni.v2i2.47.