

# OPTIMASI PREDIKSI OMSET PENJUALAN PADA PABRIK OLAHAN TAHU MENGGUNAKAN ALGORITMA REGRESI LINEAR

Ramdan Hermawan<sup>1\*</sup>, Nana Suarna<sup>2</sup>, Irfan Ali<sup>3</sup>, Dede Rohman<sup>4</sup>

<sup>1,2,3,4</sup>STMIK IKMI Cirebon; Jl. Perjuangan No. 10B, Majasem, Cirebon, Jawa Barat 45135, Telp. (0231)490480

Received: 30 Desember 2024  
Accepted: 14 Januari 2025  
Published: 20 Januari 2025

## Keywords:

Omset penjualan, Regresi linear sederhana, Harga, Prediksi, KDD

## Correspondent Email:

[dadanbang2@gmail.com](mailto:dadanbang2@gmail.com)

**Abstrak.** Peningkatan omset penjualan merupakan aspek krusial dalam keberlanjutan bisnis, termasuk di Pabrik Olahan Tahu Khong Jaya. Penelitian ini bertujuan untuk mengeksplorasi hubungan variabel-variabel seperti harga, jumlah produksi, distribusi, dan promosi terhadap omset penjualan, serta memprediksi potensi pendapatan pada tahun 2024 dan 2025. Metode yang digunakan adalah Knowledge Discovery in Database (KDD), meliputi pengumpulan data, preprocessing, transformasi, analisis menggunakan regresi linear sederhana, dan evaluasi model. Data historis dari tahun 2022 hingga 2023 dianalisis untuk mendapatkan wawasan prediktif. Hasil analisis menunjukkan bahwa variabel harga memiliki korelasi paling kuat dengan omset ( $r = 0,884$ ), diikuti oleh variabel tanggal ( $r = 0,841$ ) dan volume penjualan ( $r = 0,638$ ). Prediksi omset untuk tahun 2024 diperkirakan mencapai Rp1.399.036.400,75, meningkat menjadi Rp1.659.565.708,42 pada tahun 2025. Tren kenaikan omset ini menggambarkan pertumbuhan bisnis yang signifikan, dengan peningkatan tahunan hampir 50% pada tahun pertama dan 37,67% pada tahun kedua. Temuan ini menegaskan pentingnya strategi penetapan harga yang optimal dan peningkatan distribusi volume penjualan untuk mendukung pertumbuhan bisnis yang berkelanjutan. Penelitian ini memberikan kontribusi dalam memahami faktor-faktor utama yang memengaruhi omset penjualan serta mengembangkan model prediktif untuk mendukung pengambilan keputusan strategis. Hasil ini diharapkan dapat menjadi dasar untuk pengelolaan bisnis yang lebih efektif di masa depan.

**Abstract.** Increasing sales turnover is an important aspect of business sustainability, including at Khong Jaya Tofu Processed Factory. This research aims to explore the relationship of variables such as price, production quantity, distribution, and promotion to sales turnover, and predict potential revenue in 2024 and 2025. The method used is Knowledge Discovery in Database (KDD), including data collection, preprocessing, transformation, analysis using simple linear regression, and model evaluation. Historical data from 2022 to 2023 was analyzed to gain predictive insights. The analysis results show that the price variable has the strongest correlation with turnover ( $r = 0.884$ ), followed by the date variable ( $r = 0.841$ ) and sales volume ( $r = 0.638$ ). Predicted turnover in 2024 is estimated to reach Rp1,399,036,400.75, increasing to Rp1,659,565,708.42 in 2025. This upward trend in turnover illustrates significant business growth, with an annual increase of almost 50% in the first year and 37.67% in the second year. The findings confirm the importance of an optimal pricing strategy and increased sales volume distribution to support sustainable business growth. This research contributes to understanding the key factors that influence sales turnover and develops a predictive model to support strategic decision-making. The results of this

*study are expected to form the basis for more effective business management in the future.*

## 1. PENDAHULUAN

Perkembangan industri pangan di Indonesia semakin pesat, seiring dengan meningkatnya permintaan konsumen terhadap produk-produk berbasis pangan tradisional, salah satunya adalah tahu [1]. Pabrik Olahan Tahu Khong Jaya, sebagai salah satu pemain dalam industri ini, menghadapi tantangan dalam mempertahankan kestabilan produksi dan penjualan. Fluktuasi omset penjualan menjadi salah satu tantangan terbesar yang harus dihadapi oleh manajemen pabrik. Hal ini dipengaruhi oleh banyak faktor, seperti tren konsumsi, harga bahan baku kedelai, persaingan pasar, serta kondisi ekonomi lokal. Untuk menjaga daya saing dan efisiensi operasional, penting bagi Pabrik Tahu Khong Jaya untuk memiliki kemampuan prediksi penjualan yang akurat.

Algoritma linear regression menjadi salah satu metode yang dapat digunakan untuk memprediksi omset penjualan dengan mengidentifikasi pola hubungan antara variabel-variabel bebas [2], seperti jumlah produksi, harga, promosi, dan distribusi, dengan variabel terikat, yaitu omset penjualan. Dalam regresi linier, hubungan antara variabel-variabel tersebut diwakili oleh garis lurus, yang memungkinkan untuk memperkirakan nilai omset di masa depan berdasarkan data historis [3]. Algoritma ini dianggap efektif dalam memprediksi tren penjualan jangka pendek dan mengoptimalkan proses pengambilan keputusan terkait strategi bisnis

Pendapat ahli mendukung penggunaan regresi linier sebagai salah satu alat prediksi dalam dunia bisnis. Menurut [4] regresi linier mampu memberikan wawasan yang signifikan tentang hubungan antar variabel dalam konteks analisis bisnis, terutama jika hubungan tersebut bersifat linear dan dapat dijelaskan oleh data historis. Regresi linier juga memungkinkan manajer bisnis untuk memprediksi hasil dengan akurasi yang cukup baik selama asumsi-asumsi dasar dari model tersebut terpenuhi, seperti linearitas dan homoskedastisitas [5].

Sebagai tambahan, Jeffrey Wooldridge, seorang ahli dalam bidang econometrics, berpendapat bahwa meskipun regresi linier sederhana, ia memiliki kelebihan dalam hal interpretasi dan kemampuan memberikan estimasi yang jelas tentang pengaruh variabel independen terhadap variabel dependen. Wooldridge juga menekankan bahwa regresi linier menjadi alat prediksi yang sangat populer dalam industri karena efisiensinya dalam memproses data skala besar dan memberikan hasil yang mudah dipahami [6].

Penerapan algoritma linear regression di Pabrik Tahu Khong Jaya mampu memberikan prediksi yang lebih baik tentang omset penjualan. Dengan model prediksi yang akurat, perusahaan dapat mengoptimalkan rantai pasokan, menyesuaikan volume produksi dengan permintaan pasar, serta meminimalisir biaya operasional yang disebabkan oleh ketidakpastian dalam penjualan. Selain itu, model prediksi yang kuat dapat membantu perusahaan untuk menentukan strategi pemasaran yang lebih efektif, seperti kapan waktu yang tepat untuk melakukan promosi atau menyesuaikan harga produk.

Namun, perlu dicatat bahwa penggunaan regresi linier juga memiliki keterbatasan, terutama jika terdapat faktor-faktor lain yang berpotensi memengaruhi penjualan secara non-linear. Oleh karena itu, dalam penelitian ini, penulis juga akan mengeksplorasi sejauh mana algoritma regresi linier dapat dioptimalkan dalam konteks pabrik tahu, serta apakah ada kebutuhan untuk mengombinasikan algoritma ini dengan metode lain yang lebih kompleks untuk menghasilkan model prediksi yang lebih akurat.

Dengan demikian, penelitian ini akan mengkaji penerapan algoritma linear regression dalam meningkatkan model prediksi omset penjualan di Pabrik Olahan Tahu Khong Jaya, serta mengevaluasi efektivitas metode ini dalam konteks lingkungan bisnis yang dinamis.

## 2. TINJAUAN PUSTAKA

**2.1. Data Mining**

Data mining merupakan sebuah metode yang digunakan untuk menggali informasi dari basis data. Dalam penerapannya, data mining memanfaatkan berbagai pendekatan, seperti kecerdasan buatan, pembelajaran mesin (machine learning), dan teknik statistik, yang bertujuan untuk mengekstrak serta mengidentifikasi data guna memenuhi kebutuhan informasi dari database berukuran besar [7].

**2.2. Regresi Linear Sederhana**

Regresi linier merupakan metode statistik yang digunakan untuk mempelajari serta memodelkan hubungan antara dua atau lebih variabel. Jenis yang paling dasar dan sering digunakan adalah regresi linier sederhana [8]. Dalam pengamatan, sebuah variabel seringkali dipengaruhi oleh variabel lain. Sebagai contoh, tinggi badan dan berat badan seseorang dapat saling memengaruhi, di mana pada tinggi tertentu terdapat berat badan yang sesuai, begitu pula sebaliknya. Contoh lain adalah produksi padi, yang dipengaruhi oleh luas lahan yang ditanami, jenis pupuk yang digunakan, jumlah pupuk yang diaplikasikan, dan faktor lainnya.

Regresi linier sederhana, yang sering disingkat SLR (Simple Linear Regression), merupakan salah satu metode statistik yang digunakan dalam produksi untuk meramalkan atau memprediksi karakteristik kualitas maupun kuantitas. Persamaan umum dari metode regresi linier sederhana dalam penelitian ini adalah:

$$y = a + bx \tag{1}$$

Keterangan:

a = Konstanta

b = Koefisien regresi

Y = Variabel dependen (variable tak bebas)

X = Variabel independen (variable bebas)

Menghitung konstanta (a):

$$a = \frac{(\sum y) (\sum X^2) - (\sum X) (\sum Xy)}{n(\sum y) - (\sum)^2} \tag{2}$$

Menghitung koefisien (b)

$$b = \frac{n(\sum Xy) (\sum X^2) - (\sum X) (\sum Xy)}{n(\sum X^2) - (\sum X)^2} \tag{3}$$

Metode evaluasi peramalan dilakukan dengan menghitung jumlah kesalahan absolut. Mean Absolute Deviation (MAD) adalah metrik yang digunakan untuk mengukur akurasi peramalan dengan menghitung rata-rata nilai absolut dari setiap kesalahan. MAD bermanfaat karena mengukur kesalahan peramalan menggunakan satuan yang sama dengan data aslinya. Sebagai salah satu ukuran awal untuk mengevaluasi kesalahan keseluruhan dalam sebuah model, MAD memiliki rumus perhitungan sebagai berikut:

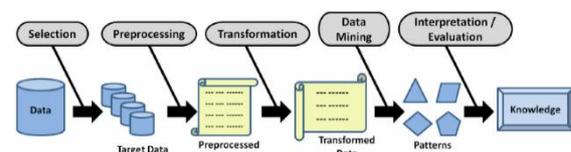
$$MAD = \frac{\sum |Y_t - F_t|}{n} \tag{4}$$

**2.3. Prediksi**

Peramalan atau prediksi adalah proses memperkirakan nilai suatu variabel di masa depan berdasarkan analisis data historis [9]. Informasi yang digunakan dalam peramalan biasanya berupa data yang dapat dihitung. Meskipun ramalan mengandung ketidakpastian terkait kondisi masa depan, tujuannya adalah memberikan estimasi yang paling mendekati. Istilah prognosis sering digunakan sebagai sinonim untuk prediksi [10].

**3. METODE PENELITIAN**

KDD (Knowledge Discovery in Databases) adalah sebuah metodologi yang digunakan dalam penelitian ini untuk mengekstraksi informasi dari dataset. Proses KDD terdiri dari lima langkah yang dilakukan secara sistematis [11], yaitu sebagai berikut:



Gambar 1. Tahapan KDD

**3.1. Data Selection**

Pada tahap awal ini, fokus utamanya adalah memilih dataset yang sesuai dan relevan dengan tujuan penelitian.

**3.2. Preprocessing Data**

Tahap kedua dalam proses pra-pemrosesan data melibatkan pembersihan dan penghapusan data yang tidak relevan,

menghilangkan nilai yang tidak sesuai, serta melakukan normalisasi data jika diperlukan.

### 3.3. Transformation Data

Pada tahap ini, data diubah menjadi format yang sesuai untuk digunakan dalam proses data mining, dengan tujuan memudahkan pengolahan data oleh alat dan algoritma yang akan digunakan.

### 3.4. Data Mining

Data mining adalah proses untuk mengidentifikasi pola atau informasi dalam kumpulan data selama tahap transformasi data. Dalam hal ini, metode Regresi Linear Sederhana digunakan sesuai dengan tujuan penelitian.

### 3.5. Evaluation

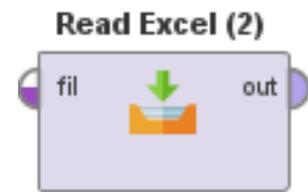
Langkah terakhir adalah merangkum dan menyimpulkan analisis terkait omset penjualan tahu dengan menerapkan metode regresi linear menggunakan RapidMiner, sekaligus memberikan gambaran mengenai prediksi omset penjualan di masa depan.

## 4. HASIL DAN PEMBAHASAN

Penelitian ini mengikuti tahapan yang sesuai dengan metodologi KDD (Knowledge Discovery in Databases).

### 4.1. Selection

Pada tahap seleksi data, data yang akan digunakan untuk prediksi dipilih menggunakan tools RapidMiner. Proses seleksi ini dilakukan melalui Microsoft Excel, di mana dalam penelitian ini, data penjualan periode 2 tahun terakhir dipilih untuk digunakan dalam proses data mining guna memprediksi penjualan tahu. Hasil dari seleksi data diketahui pada data penjualan terdapat 5 atribut yaitu atribut No, tanggal, harga, volume penjualan, omset penjualan. Proses pengambilan data awal menggunakan operator Read excel di tools RapidMiner. Operator ini berfungsi untuk membaca dataset yang telah disiapkan sebelumnya dalam format Excel, sehingga data dapat digunakan untuk proses lebih lanjut dapat dilihat pada gambar 2.



Gambar 2. Operator Read Excel

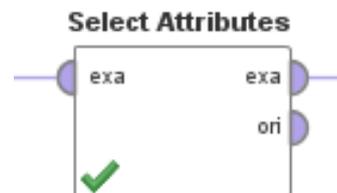
Berdasarkan hasil pembacaan data penjualan oleh oprator read exel dalam statistik dapat dilihat pada Gambar 3. di bawah ini.

Row No.	NO	Tanggal	Harga Produ...	Volume Penj...	OMSET
1	1	Jan 1, 2022	200	12000	2400000
2	2	Jan 2, 2022	200	10000	2000000
3	3	Jan 3, 2022	200	8000	1600000
4	4	Jan 4, 2022	200	12000	2400000
5	5	Jan 5, 2022	200	10000	2000000
6	6	Jan 6, 2022	200	8000	1600000
7	7	Jan 7, 2022	200	12000	2400000
8	8	Jan 8, 2022	200	10000	2000000
9	9	Jan 9, 2022	200	8000	1600000
10	10	Jan 10, 2022	200	12000	2400000
11	11	Jan 11, 2022	200	10000	2000000
12	12	Jan 12, 2022	200	8000	1600000
13	13	Jan 13, 2022	200	12000	2400000
14	14	Jan 14, 2022	200	10000	2000000
15	15	Jan 15, 2022	200	8000	1600000
16	16	Jan 16, 2022	200	12000	2400000
17	17	Jan 17, 2022	200	10000	2000000

ExampleSet (729 examples, 0 special attributes, 5 regular attributes)

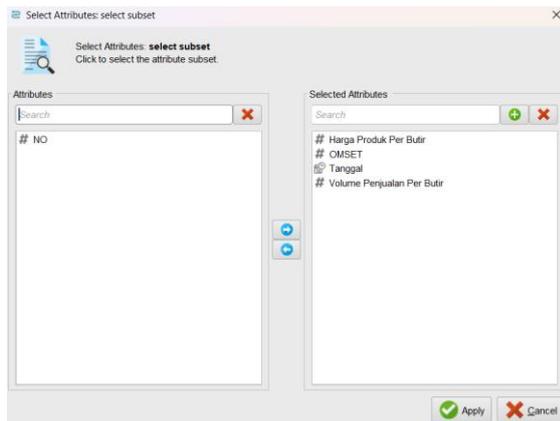
Gambar 3. Hasil Read Dataset

Langkah selanjutnya adalah menggunakan operator *Select Attributes*, yang berfungsi untuk memilih atribut-atribut tertentu yang akan digunakan dalam analisis. Operator ini memungkinkan penulis untuk menyaring data sehingga hanya variabel yang relevan yang disertakan dalam proses pengelompokan. Gambar 4. menunjukkan tampilan operator *Select Attributes* yang digunakan dalam pemilihan atribut.



Gambar 4. Operator Select Atribut

Berdasarkan hasil select atribut dapat dilihat pada gambar 4. di bawah.



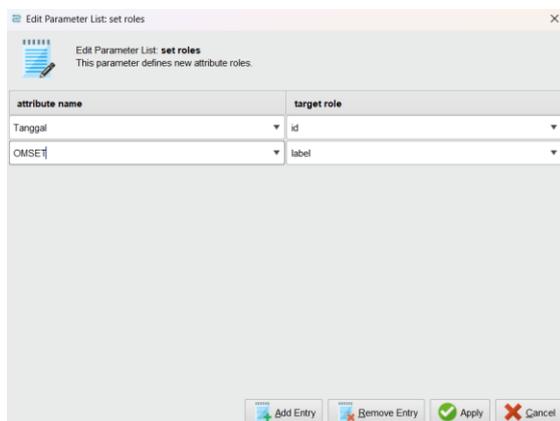
Gambar 5. Hasil dari proses select atribut

Langkah selanjutnya adalah menggunakan operator *Set Role*, yang berfungsi untuk menentukan peran atribut, seperti menetapkan atribut yang berfungsi sebagai *id* atau *label*. Dengan mengatur peran atribut ini, kita dapat memastikan bahwa atribut yang tidak relevan dengan analisis seperti *id*, tidak akan ikut diproses dalam langkah selanjutnya. Gambar 6. menunjukkan tampilan operator *Set Role* yang digunakan dalam proses ini.



Gambar 6. Operator Set Role

Berdasarkan hasil set role dapat dilihat pada gambar 7. di bawah.



Gambar 7. Parameter set role

Pada parameter *Set Role*, dipilih untuk mengedit daftar atribut yang ada dan

menentukan peran setiap atribut dalam analisis prediksi. Atribut seperti Tanggal penulis ditetapkan sebagai *id*, karena atribut ini yang akan digunakan untuk prediksi omset penjualan tahu. Sementara itu, atribut OMSET, yang berisi hasil penjualan tahu, penulis ditetapkan sebagai *label*, karena atribut ini menjadi target yang akan di prediksi melalui model.

Row No.	Tanggal	OMSET	Harga Produ...	Volume Penj...
1	Jan 1, 2022	2400000	200	12000
2	Jan 2, 2022	2000000	200	10000
3	Jan 3, 2022	1600000	200	8000
4	Jan 4, 2022	2400000	200	12000
5	Jan 5, 2022	2000000	200	10000
6	Jan 6, 2022	1600000	200	8000
7	Jan 7, 2022	2400000	200	12000
8	Jan 8, 2022	2000000	200	10000
9	Jan 9, 2022	1600000	200	8000
10	Jan 10, 2022	2400000	200	12000
11	Jan 11, 2022	2000000	200	10000
12	Jan 12, 2022	1600000	200	8000
13	Jan 13, 2022	2400000	200	12000
14	Jan 14, 2022	2000000	200	10000
15	Jan 15, 2022	1600000	200	8000
16	Jan 16, 2022	2400000	200	12000
17	Jan 17, 2022	2000000	200	10000

ExampleSet (729 examples, 2 special attributes, 2 regular attributes)

Gambar 8. Hasil set role

Hasil dari Gambar 4.8 pada operator *Set Role* menunjukkan bahwa atribut dalam dataset telah berhasil diberi peran sesuai dengan fungsinya, kemudian terdapat 2 spesial atribut dan 2 regular atribut. Atribut-atribut seperti *harga produk per butir* dan *volume penjualan per butir* diperlakukan sebagai regular, yang berarti atribut - atribut ini berfungsi sebagai fitur input untuk memprediksi omset penjualan tahu. Sementara itu, atribut OMSET yang menunjukkan hasil penjualan diperlakukan sebagai label, yang menjadi target atau hasil yang ingin diprediksi oleh model. Model proses dalam langkah Selection di RapidMiner (AI Studio 2024) ditampilkan pada Gambar 4.9 di bawah ini.



Gambar 9. Proses selection 1

#### 4.1.2. Pre-processing

*Pre-processing* atau pembersihan data adalah tahap awal dalam analisis data, di mana data yang telah dipilih disiapkan dan dibersihkan untuk memastikan kualitas data yang akan dianalisis. Tahap ini sangat penting untuk menghilangkan noise, ketidakkonsistenan, serta masalah lainnya dalam data yang dapat memengaruhi hasil analisis. Karena dalam dataset ini tidak ditemukan nilai yang hilang atau missing values, proses pembersihan tidak memerlukan penggunaan operator tambahan untuk mengisi atau menghapus nilai yang hilang. Hasil dari pemeriksaan ini dapat dilihat pada Gambar 10, yang menunjukkan bahwa tidak ada data yang hilang atau *missing*. Dengan demikian, data siap digunakan untuk langkah-langkah analisis selanjutnya.

Name	Type	Missing	Statistics	Filter (5/5 attributes)	Search for Attributes
NO	Integer	0	Min: 1, Max: 729, Average: 364.108		
Tanggal	Date-time	0	Lowest date: Jan 1, 2022, Highest date: Dec 31, 2023, Standard: 729 days		
Harga Produk Per Butir	Integer	0	Min: 200, Max: 300, Average: 254.458		
Volume Penjualan Per Butir	Integer	0	Min: 8000, Max: 13000, Average: 11524.417		
OMSET	Integer	0	Min: 18500000, Max: 39000000, Average: 2945012.830		

Gambar 10. Hasil Preprocessing

**4.1.3. Transformation**

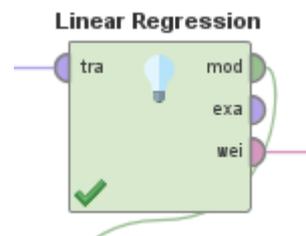
Pada tahap ini, dilakukan proses transformasi data menjadi format yang tepat untuk dapat diproses dalam pemodelan data mining. Bertujuan untuk memudahkan koordinasi data yang diproses oleh algoritma dan tools yang digunakan dalam penelitian yaitu rapidminer. Pada tahap transformasi pada data penjualan terdapat 5 atribut yaitu no, tanggal, volume penjualan dan omset penjualan seperti yang terlihat pada gambar tabel sebagai berikut.

NO	Harga Produk Per Butir	Volume Penjualan Per Butir	OMSET
1	300	13000	3900000
2	300	10000	3000000
3	300	8000	2400000
4	300	13000	3900000
5	300	10000	3000000
6	300	8000	2400000
7	300	13000	3900000
8	300	10000	3000000
9	300	8000	2400000
10	300	13000	3900000
.....			
720	500	11000	5500000
721	500	11000	5500000
722	500	11000	5500000
723	500	11000	5500000
724	500	11000	5500000
725	500	13000	6500000
726	500	13000	6500000
727	500	13000	6500000
728	500	13000	6500000
729	500	13000	6500000

Gambar 11. Hasil Transformation

**4.1.4. Data Mining**

Proses data mining yang diterapkan dalam penelitian ini yaitu prediksi menggunakan regresi linear. Pada penelitian ini, akan dimuat perhitungan prediksi omset penjualan per hari. Setelah data diolah, perhitungan akan dilakukan menggunakan RapidMiner. Proses dimulai dengan mengonfigurasi data ke dalam Rapid Miner, yang melibatkan penyesuaian dataset dan pembentukan model, sebagaimana terlihat pada Gambar 4.12 di bawah ini.



Gambar 12. Operator Linear Regression

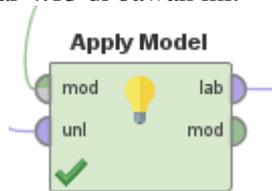
Adapun parameter yang digunakan dalam operator K-Nearest Neighbors seperti pada Tabel 1.

Tabel 1. Parameter regresi linear

No	Parameter	Isi
1	Feature selection	T-Test
2	Alpha	0.05
3	Eliminate colinear feature	✓
4	Min tolerance	0.05
5	Use bias	✓
6	ridge	1.0E-8

Tabel 4.1 menampilkan sejumlah parameter yang digunakan untuk mengatur algoritma Regresi Linear dalam penelitian ini. Parameter Feature Selection berfungsi memilih

variabel independen yang relevan untuk model, dengan tujuan menghilangkan fitur yang tidak relevan atau redundan, sehingga dapat mengurangi overfitting, meningkatkan interpretasi model, dan mempercepat proses pelatihan. Parameter Alpha digunakan untuk mengatur kekuatan regularisasi dalam regresi ridge atau lasso, membantu mencegah overfitting dengan menambahkan penalti pada koefisien regresi. Parameter Eliminate Colinear Feature bertugas menghapus fitur dengan multikolinieritas tinggi (korelasi antar-fitur yang sangat kuat). Parameter Min Tolerance menetapkan batas toleransi minimum untuk menentukan konvergensi atau menghentikan iterasi, sehingga pelatihan dihentikan ketika perubahan pada fungsi loss berada dalam ambang toleransi, mencegah overtraining. Parameter Use Bias diaktifkan untuk menambahkan bias atau intercept ke dalam model, yang memungkinkan model membuat prediksi yang akurat ketika semua nilai fitur adalah nol. Parameter Ridge menambahkan regularisasi L2 untuk mengurangi overfitting dengan memberikan penalti pada besar koefisien regresi. Regularisasi ini memungkinkan semua fitur tetap dipertahankan, tetapi dengan koefisien yang lebih kecil. Proses selanjutnya dapat dilihat pada Gambar 4.13 di bawah ini.



Gambar 13 Operator Apply Model

Model regresi linear yang telah dikembangkan diterapkan pada data uji melalui operator Apply Model. Operator ini berfungsi sebagai sarana untuk menerapkan model yang telah dilatih, memungkinkan prediksi nilai omset berdasarkan data dari atribut lain, seperti Volume penjualan dan harga produk. Dengan penerapan ini, model dapat menguji kemampuannya dalam menghasilkan prediksi yang akurat sesuai dengan pola yang telah dipelajari sebelumnya. Adapaun parameter yang digunakan seperti Tabel dibawah ini.

Tabel 2. Parameter Apply Model

No	Parameter	Isi
1	Application	default

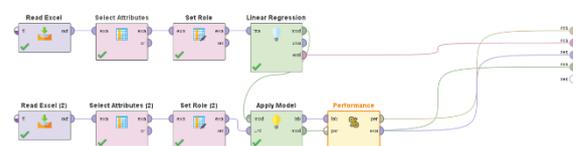
Di Tabel 2 penelitian ini menggunakan parameter application parameter sesuai default yang tersedia di operator apply modelnya. parameter yang digunakan adalah default dari operator apply model itu sendiri. Setelah model selesai dilatih menggunakan data latih, langkah selanjutnya adalah mengevaluasi kinerjanya dalam memprediksi keluaran berdasarkan data yang tidak pernah dilihat sebelumnya. Proses ini penting untuk memastikan bahwa model tidak hanya menghafal data latih (overfitting), tetapi juga mampu menggeneralisasi pola yang relevan dari data tersebut. Adapaun hasilnya terdapat pada Gambar 4.14 di bawah ini.

Row No.	Tanggal	OMSET	prediction(O...	Harga Produ...	Volume Penj...
1	Jan 1, 2024	?	3819809.664	300	13000
2	Jan 2, 2024	?	3083509.496	300	10000
3	Jan 3, 2024	?	2592642.717	300	8000
4	Jan 4, 2024	?	3819809.664	300	13000
5	Jan 5, 2024	?	3083509.496	300	10000
6	Jan 6, 2024	?	2592642.717	300	8000
7	Jan 7, 2024	?	3819809.664	300	13000
8	Jan 8, 2024	?	3083509.496	300	10000
9	Jan 9, 2024	?	2592642.717	300	8000
10	Jan 10, 2024	?	3819809.664	300	13000
11	Jan 11, 2024	?	3083509.496	300	10000
12	Jan 12, 2024	?	2592642.717	300	8000
13	Jan 13, 2024	?	3819809.664	300	13000
14	Jan 14, 2024	?	3083509.496	300	10000
15	Jan 15, 2024	?	2592642.717	300	8000
16	Jan 16, 2024	?	3819809.664	300	13000

ExampleSet (729 examples, 3 special attributes, 2 regular attributes)

Gambar 14. Hasil Prediksi

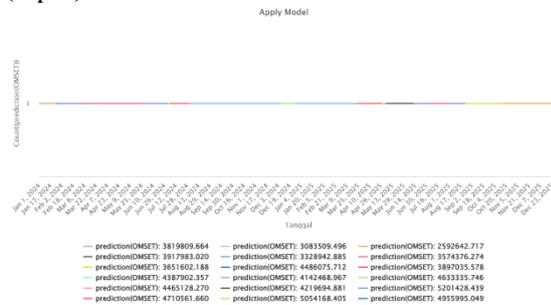
Adapun model pengujian algoritma Regresi Linear ditampilkan pada Gambar 4.15.



Gambar 15. Proses Linear Regression

Model pengujian pada workflow di atas menggunakan algoritma Regresi Linear untuk memprediksi berapa jumlah omset penjualan pada tahun 2024 dan 2025. Proses dimulai dengan membaca dataset menggunakan operator Read CSV, lalu memilih atribut relevan seperti tanggal, harga produk per butir,

volume penjualan per butir, dan omset melalui Select Attributes. Operator Set Role kemudian menetapkan peran atribut: kolom omset dipilih sebagai label (target) untuk prediksi, sedangkan kolom lainnya dijadikan *regular attributes* (input).



Gambar 16. Visualisasi Hasil Prediksi

Visualisasi hasil prediksi model regresi linear dapat dilihat pada grafik yang menampilkan perbandingan nilai aktual dan prediksi. Grafik tersebut menunjukkan bahwa model mampu memprediksi omset dari tahun 2024 hingga 2025 dengan pola yang hampir serupa, meskipun terdapat sedikit perbedaan pada nilai tertentu. Visualisasi ini membantu memverifikasi sejauh mana model berhasil memprediksi nilai OMSET berdasarkan data historis yang tersedia.

**4.1.5. Evaluation**

Pada hasil eksperimen diketahui nilai prediksi untuk omset pada tahun 2024 dan 2025 seperti pada Gambar 17.

Row No.	Tanggal	OMSET	prediction(O...	Harga Produ...	Volume Penj
1	Jan 1, 2024	?	3819809.664	300	13000
2	Jan 2, 2024	?	3083509.496	300	10000
3	Jan 3, 2024	?	2592642.717	300	8000
4	Jan 4, 2024	?	3819809.664	300	13000
5	Jan 5, 2024	?	3083509.496	300	10000
6	Jan 6, 2024	?	2592642.717	300	8000
7	Jan 7, 2024	?	3819809.664	300	13000
8	Jan 8, 2024	?	3083509.496	300	10000
9	Jan 9, 2024	?	2592642.717	300	8000
10	Jan 10, 2024	?	3819809.664	300	13000
11	Jan 11, 2024	?	3083509.496	300	10000
12	Jan 12, 2024	?	2592642.717	300	8000
13	Jan 13, 2024	?	3819809.664	300	13000
14	Jan 14, 2024	?	3083509.496	300	10000
15	Jan 15, 2024	?	2592642.717	300	8000
16	Jan 16, 2024	?	3819809.664	300	13000
17	Jan 17, 2024	?	3083509.496	300	10000
18	Jan 18, 2024	?	2592642.717	300	8000

ExampleSet (729 examples, 3 special attributes, 2 regular attributes)

Gambar 17. Hasil Prediksi Omset

Gambar 17. menunjukkan hasil dari prediksi omset tahun 2024-2025 yang

dilakukan menggunakan algoritma Regresi Linear pada dataset penjualan pabrik olahan tahu Khong Jaya. Hasil keseluruhan menunjukkan pada tahun 2024 omset penjualan tahu mencapai angka 1.399.036.400,75, dan pada tahun 2025 didapatkan hasil bahwa omset penjualan tahu mencapai angka 1.659.565.708,42.

**4.1.6. Variabel Berpengaruh**

Berdasarkan analisis matriks korelasi, diketahui bahwa atribut-atribut harga per butir, volume penjualan per butir, dan tanggal memiliki pengaruh signifikan terhadap omset penjualan di Pabrik Olahan Tahu Khong Jaya. Variabel harga per butir menunjukkan korelasi paling kuat dengan omset, sebesar 0.884, yang mengindikasikan bahwa perubahan harga memiliki dampak langsung dan signifikan terhadap tingkat pendapatan. Variabel volume penjualan per butir memiliki korelasi sebesar 0.638, menunjukkan hubungan positif yang sedang hingga kuat, di mana peningkatan volume penjualan turut berkontribusi terhadap kenaikan omset. Selain itu, variabel tanggal juga memiliki korelasi yang kuat sebesar 0.841, yang dapat mencerminkan adanya pola musiman atau tren tertentu dalam aktivitas penjualan. Secara keseluruhan, hasil ini menegaskan bahwa harga per butir merupakan faktor utama yang memengaruhi omset, diikuti oleh faktor waktu dan volume penjualan. Untuk memperdalam pemahaman mengenai pengaruh tersebut, analisis regresi dapat digunakan untuk mengukur dampak masing-masing variabel secara lebih detail.

**4.1.7. Hasil Prediksi Omset Penjualan**

Hasil penelitian ini menunjukkan prediksi omset penjualan tahu dari Pabrik Olahan Tahu Khong Jaya untuk tahun 2024-2025, yang dihitung menggunakan algoritma Regresi Linear Sederhana. Grafik yang ditampilkan menggambarkan proyeksi omset secara bertahap, dengan setiap segmen warna pada grafik mewakili nilai prediksi omset untuk periode tertentu. Pada tahun 2024, total prediksi omset penjualan tahu diperkirakan mencapai Rp1.399.036.400,75, sementara pada tahun 2025, angka ini diperkirakan meningkat menjadi Rp1.659.565.708,42. Hal ini menunjukkan adanya tren kenaikan omset tahunan, yang dimulai dengan omset sebesar

Rp937.615.000 pada tahun 2022, dan meningkat menjadi Rp1.205.800.000 pada tahun 2023. Kenaikan tersebut mencatatkan hampir 50% pada tahun pertama dan 37,67% pada tahun kedua, yang dapat dijadikan indikator positif untuk pertumbuhan penjualan tahu dalam periode tersebut. Angka-angka yang tertera pada legenda dan garis tren memberikan gambaran rinci mengenai prediksi untuk setiap periode yang diukur, serta mencerminkan pola fluktuasi dan perkembangan yang konsisten dengan data penjualan historis.

Hasil penelitian ini mendukung temuan dari Penelitian dari Miftahuljannah yang mana membahas penggunaan metode Regresi Linear Sederhana untuk prediksi penjualan produk. Hasil prediksi data menunjukkan adanya peningkatan pesanan hingga mencapai 11.614, disertai evaluasi model dengan nilai MAE sebesar 22,89, MSE sebesar 9,13, RMSE sebesar 3,02, dan R<sup>2</sup>-Score sebesar 0,91 [12]. Namun hasil ini tidak sejalan dengan penelitian ini karena tidak menitikberatkan pada pengukuran akurasi, melainkan fokus pada hasil prediksi omset itu sendiri. Hasil penelitian ini menunjukkan tren peningkatan yang baik pada omset penjualan, yang mengindikasikan bahwa regresi linier dapat memberikan proyeksi yang sesuai dengan pola historis data penjualan.

## 5. KESIMPULAN

Berdasarkan hasil penelitian menggunakan regresi linear sederhana untuk memprediksi OMSET Penjualan Tahu di Pabrik Khong Jaya dapat disimpulkan bahwa:

### 1. karakteristik Data yang Berpengaruh:

Analisis menunjukkan bahwa fitur-fitur utama seperti tanggal, omset, volume, harga produk, merupakan atribut yang paling memengaruhi hasil prediksi. Temuan ini sejalan dengan literatur yang menyebutkan bahwa faktor-faktor tersebut merupakan indikator utama dalam menentukan peningkatan omset penjualan produk.

### 2. Model prediksi dengan Algoritma Regresi Linear Sederhana:

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Regresi Linear sederhana efektif digunakan untuk memprediksi omset penjualan tahu di Pabrik Olahan Tahu Khong Jaya. Prediksi menunjukkan tren peningkatan omset tahunan yang konsisten, dengan omset sebesar Rp937.615.000 pada tahun 2022, meningkat menjadi Rp1.205.800.000 pada tahun 2023, Rp1.399.036.400,75 pada tahun 2024, dan Rp1.659.565.708,42 pada tahun 2025. Analisis matriks korelasi menunjukkan bahwa harga per butir, volume penjualan per butir, dan tanggal merupakan faktor utama yang memengaruhi omset. Di antaranya, harga per butir memiliki korelasi paling kuat dengan omset (0,884), diikuti oleh tanggal (0,841) dan volume penjualan per butir (0,638). Temuan ini menunjukkan bahwa regresi linier dapat memberikan proyeksi omset yang selaras dengan pola historis penjualan, memberikan wawasan penting untuk perencanaan bisnis di masa depan.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

**DAFTAR PUSTAKA**

- [1] I. H. Payong, P. Jandu, "Pengolahan Limbah Ampas Tahu Menjadi Kerupuk Aneka Rasa Di Kelurahan Poco Mal," *Jurdimas (Jurnal Pengabd. Kpd. Masyarakat) R.*, vol. 6, no. 2, 2023, doi: 10.33330/jurdimas.v6i2.1887.
- [2] S. Hasibuan, L. H. Musthofa, "Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang," *JOSTECH*, doi: 10.15548/jostech.v2i1.3802.
- [3] D. Hamdanah, F. Fitriana, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah," *J. Nas. Pendidik. Teknol. Inf.*, 2021, doi: 10.23887/JANAPATI.V10I1.31035.
- [4] S. K. Aityan, *Numerical Methods Using Java*. 2022. [Online]. Available: [https://doi.org/10.1007/978-1-4842-6797-4\\_14](https://doi.org/10.1007/978-1-4842-6797-4_14)
- [5] R. O. Mensah, *Applied Linear Regression for Business Analytics with R*. 2023. [Online]. Available: [https://doi.org/10.1007/978-3-031-21480-6\\_3](https://doi.org/10.1007/978-3-031-21480-6_3)
- [6] J. M. Wooldridge, *Econometrics Introductory Modern Approach*. 2018.
- [7] Dodi Nofri Yoliad, "Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-means," *Insearch (Information Syst. Res. J.)*, vol. 3, no. 1, 2023.
- [8] H. Husdi and H. Dalai, "Penerapan Metode Regresi Linear Untuk Prediksi Jumlah Bahan Baku Produksi Selai Bilfagi," *J. Inform.*, vol. 10, no. 2, pp. 129–135, 2023, doi: 10.31294/inf.v10i2.14129.
- [9] T. C. Silva, *Models and Methods for Management Science*. 2022. [Online]. Available: [https://doi.org/10.1007/978-981-19-1614-4\\_3](https://doi.org/10.1007/978-981-19-1614-4_3)
- [10] D. H. Taufik Hidayat, Rahmi Darnis, "Algoritma Regresi Linier Berganda Untuk Analisis Efisiensi Stok Produk Di PT. Madu Pramuka Batang," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 10, no. 3, pp. 239–247, 2022, [Online]. Available: <http://dx.doi.org/10.23960/jitet.v12i3.4899%0AALGORITMA>
- [11] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [12] M. Ferdinan, "Prediksi Jumlah Penjualan Tahun 2024 Menggunakan Metode Regresi Linier," vol. 4, no. 3, pp. 202–208, 2024.