

Implementasi Data *Mining* Prediksi Penjualan Produk Semen Menggunakan Metode *Linear Regression* (Studi Kasus PT. Toyo Mortar Indonesia)

Rosyiqoh Badzlin ¹, Kiki Setiawan ^{2*}

^{1,2*} Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: rosyiqohbadzlin@gmail.com ¹, ki2djoaz@gmail.com ^{2*}

Histori Artikel:

Dikirim 27 Mei 2024; Diterima dalam bentuk revisi 22 Juni 2024; Diterima 20 Juli 2024; Diterbitkan 20 September 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STM IK Indonesia Banda Aceh.

Abstrak

Industri semen di Indonesia memiliki peran vital dalam pembangunan infrastruktur dan sektor konstruksi. PT. Toyo Mortar Indonesia, yang berdiri pada tahun 2015 di kawasan industri Tangerang, berfokus pada produksi semen instan Mortar. Namun, industri semen nasional menghadapi tantangan penurunan penjualan akibat pandemi yang memengaruhi daya beli masyarakat dan memperlambat proyek properti serta infrastruktur. Untuk mencapai potensi pertumbuhan di masa depan, diperlukan strategi yang tepat dalam merencanakan penjualan. Penelitian ini mengimplementasikan data mining dengan metode *regresi linear* menggunakan aplikasi RapidMiner untuk memprediksi penjualan produk semen. Data yang digunakan mencakup data penjualan historis dan informasi produk dari PT. Toyo Mortar Indonesia. Dalam pengujian yang dilakukan, data diolah menggunakan 12 variabel, dimana 10 variabel yang berpengaruh pada hasil prediksi adalah: TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM-101, UM BIRU, bahan baku, dan penjualan lainnya. Hasil pengujian menunjukkan bahwa metode *Regresi linear* mampu memprediksi penjualan produk semen dengan nilai Root Mean Square Error (RMSE) rata-rata sebesar 493,125.701 dengan standar deviasi 126,330.525. Rata-rata Absolute Error sebesar 303,270.965 dengan standar deviasi 46,207.586, serta kesalahan relatif rata-rata sebesar 1.89% dengan standar deviasi 0.36%, yang menunjukkan akurasi prediksi yang sangat baik. Hal ini dapat membantu perusahaan dalam pengambilan keputusan terkait perencanaan penjualan menggunakan data penjualan yang tersedia.

Kata Kunci: Prediksi Penjualan Semen; *Regresi linear*; Data Mining RapidMiner.

Abstract

The cement industry in Indonesia plays a vital role in infrastructure development and the construction sector. PT. Toyo Mortar Indonesia, established in 2015 in the Tangerang industrial area, focuses on the production of instant mortar cement. However, the national cement industry faces the challenge of declining sales due to the pandemic, which has affected people's purchasing power and slowed down property and infrastructure projects. To achieve future growth potential, the right strategy is needed in sales planning. This study implements data mining using the linear regression method with the RapidMiner application to predict cement product sales. The data used includes historical sales data and product information from PT. Toyo Mortar Indonesia. In the tests conducted, the data was processed using 12 variables, of which 10 variables influenced the prediction results: TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM-101, UM BIRU, raw materials, and other sales. The test results show that the Linear Regression method can predict cement product sales with an average Root Mean Square Error (RMSE) value of 493,125.701 with a standard deviation of 126,330.525. The average Absolute Error is 303,270.965 with a standard deviation of 46,207.586, and the average relative error is 1.89% with a standard deviation of 0.36%, indicating very good prediction accuracy. This can help the company in making decisions related to sales planning using the available sales data.

Keyword: Cement Sales Prediction; Linear Regression; Data Mining RapidMiner.

1. Pendahuluan

Industri Semen di Indonesia memiliki peran vital dalam pembangunan infrastruktur dan sektor konstruksi. Pertumbuhan industri semen tercermin dari peningkatan jumlah proyek konstruksi skala besar di berbagai wilayah Indonesia. PT. Toyo Mortar Indonesia sebagai salah satu perusahaan yang berdiri pada tahun 2015, bergerak di bidang industri bahan bangunan, khususnya produksi semen instan yang dikenal dengan nama semen Mortar. PT. Toyo Mortar Indonesia terletak strategis di kawasan industri Tangerang. PT. Toyo Mortar menyediakan produk semen mortar dalam berbagai kode produk yaitu TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM BIRU, UM-101 dan Bahan baku. Meskipun PT. Toyo Mortar Indonesia telah menunjukkan komitmennya dalam menyediakan pasokan semen untuk proyek-proyek konstruksi, tantangan yang dihadapi industri semen secara keseluruhan tidak bisa diabaikan. Industri semen nasional mengalami penurunan penjualan yang signifikan akibat berbagai faktor, termasuk dampak pandemi yang mempengaruhi daya beli masyarakat dan perlambatan proyek-proyek properti serta infrastruktur. Industri semen memang memiliki potensi pertumbuhan di masa depan, namun tantangan yang ada menuntut strategi yang tepat dalam merencanakan produksi dan penjualan. Prediksi penjualan menjadi sangat penting untuk membantu dalam perencanaan strategi penjualan. PT. Toyo Mortar Indonesia memiliki data penjualan yang cukup besar. Namun berdasarkan wawancara dan observasi yang dilakukan pada PT. Toyo Mortar Indonesia, data ini belum diolah untuk memprediksi penjualan semen sehingga jumlah produk yang terjual sebagai variabel dependen dan nilai penjualan sebagai variabel independen yang akan digunakan dalam prediksi belum diidentifikasi secara jelas. Tanpa pengolahan data yang tepat, perusahaan sulit untuk membuat keputusan yang berdasarkan informasi akurat mengenai penjualan yang akan datang. Untuk mengatasi hal itu, maka penelitian ini bertujuan untuk menerapkan data mining dalam memprediksi penjualan bagi PT. Toyo Mortar Indonesia menggunakan data penjualan pada tahun sebelumnya. Data *mining* adalah bidang pengetahuan yang memfokuskan pada teknik-teknik untuk mengekstrak pengetahuan atau mengidentifikasi pola dari kumpulan data. Salah satu topik penelitian yang umum dalam data mining adalah prediksi (Rajab Mudatsir & Melangi, 2022).

Metode *regresi linear* adalah salah satu metode yang dapat diterapkan sebagai pendekatan untuk menghasilkan prediksi informasi (Revaldi *et al.*, n.d.). Dengan menerapkan metode ini, perusahaan dapat mengidentifikasi hubungan antara variabel penjualan dengan faktor-faktor lain yang memengaruhi penjualan tersebut. Namun, belum teridentifikasi apakah metode *regresi linear* efektif dalam memprediksi dengan akurat. Metode *regresi linear* dapat membantu PT. Toyo Mortar Indonesia mengatasi tantangan penurunan penjualan dengan menyediakan analisis tren penjualan yang mendalam dari data historis, sehingga perusahaan dapat mengidentifikasi pola penurunan atau peningkatan penjualan. *Regresi linear* juga membantu mengidentifikasi faktor-faktor kunci yang mempengaruhi penjualan, memungkinkan perusahaan untuk fokus pada variabel-variabel signifikan seperti jenis produk atau periode waktu tertentu. Penelitian terdahulu yang menggunakan algoritma *regresi linear* sebagai metode prediksi telah dilakukan oleh berbagai peneliti. Nuning Rusmilawati, Putri Taqwa Prasetyaningrum (Rusmilawati & Prasetyaningrum, n.d.) melakukan analisis Data Mining terhadap hasil produksi kelapa sawit agar dapat menjadi informasi yang berguna bagi PT Borneo Ketapang Indah. Dalam analisis ini, peneliti menggunakan perangkat *Rapidminer*, dengan metode *Linier Regression*. Pieter Chang Hartono, Albertus Dwiyo Widiantoro (Chang Hartono & Dwiyo Widiantoro, 2024) memprediksi harga saham Unilever menggunakan metode analisis *regresi linear* menunjukkan kinerja yang kurang memuaskan, dengan nilai *root mean squared* sebesar 57.072. Hal ini mengindikasikan bahwa metode tersebut memiliki keterbatasan dalam memprediksi dengan akurasi yang tinggi. Yusuf Syakir, Teguh Iman Hermanto, Yudhi Raymond Ramadhan (Syakir *et al.*, 2022) Memprediksi perkiraan penjualan berdasarkan total pengunjung dan total pesanan, digunakan metode Data Mining SEMMA. Setelah melakukan pengujian dengan menggunakan *RapidMiner*, diperoleh hasil berikut: *Mean Square Error* (MSE) sebesar 5.172.628.212.404, *Root Mean Square Error* (RMSE) sebesar 2.274.341, dan *Mean Absolute Percentage Error* (MAPE) sebesar 4,34%. Agus Rizkiawan, Tri Wahyudi (Rizkiawan *et al.*, 2023) menerapkan data mining dengan metode CRISP-DM

menggunakan pemodelan algoritma Linear Regression dilakukan untuk memprediksi jumlah anggota baru yang akan bergabung di masa mendatang. Terdapat lima variabel yang mempengaruhi hasil prediksi, yaitu *Outgoing Call*, *Answer*, *Call Duration*, Paket *Gold*, dan Paket *Silver*. Hasil evaluasi performa menunjukkan bahwa model algoritma Linear Regression menghasilkan performa yang baik dan prediksi yang akurat, dengan nilai *Root Mean Squared Error* (RMSE) sebesar 0,098. Dini Rahayu, Aris Gunaryati (Rahayu & Gunaryati, 2023) Prediksi terhadap jumlah produk defect menggunakan model persamaan *Regresi linear* sederhana, setelah dibandingkan dengan hasil perhitungan secara aktual dan dengan menggunakan aplikasi RapidMiner, umumnya menunjukkan kesamaan hasil. Evaluasi dan pengujian nilai RMSE pada model *regresi linear* yang diterapkan juga menghasilkan nilai RMSE sebesar 0,98.

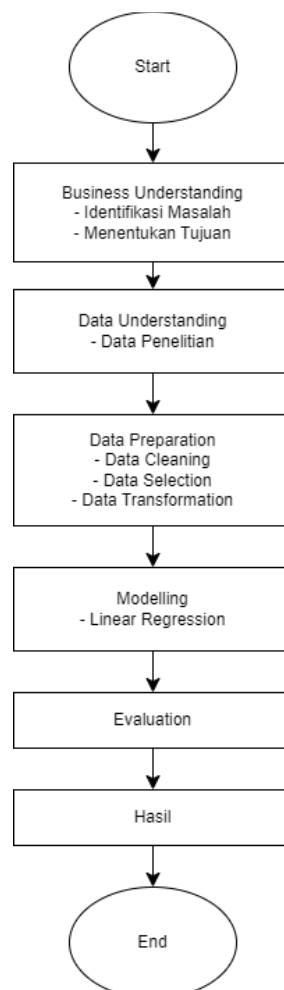
Lalu Erwin, Kartarina, Rini Anggriani, Widia Febriana, Ni Ketut Sriwinart (Erwin *et al.*, n.d.) Menggunakan metode *Regresi Linier* Berganda dan Ordinary Least Squares untuk memperkirakan penjualan berdasarkan fluktuasi rata-rata transaksi barang dalam rentang waktu tertentu. Evaluasi pola menunjukkan bahwa model yang dihasilkan dapat menjelaskan 85% dari variasi variabel terikat, dengan kekuatan hubungan antara variabel bebas dan terikat yang sangat kuat, mencapai 0,923. Ratih Dewi Putri, Andri (Dewi Putri *et al.*, 2022) menggunakan data penjualan produk elektronik untuk memprediksi produk terlaris dan hasil prediksi untuk periode tahun 2019 hingga tahun 2021 berdasarkan produk telah diketahui, dan nilai kinerja menunjukkan hasil yang baik. Nilai *Root Mean Square Error* (RSME) sebesar 1,083, dan nilai kesalahan absolut sebesar 1,065. Dichi Setiawan, Nurhadi Surojudin, Wahyu Hadikristanto (Setiawan *et al.*, 2022) menggunakan pendekatan metode prediksi model algoritma *Regresi linear* sederhana untuk prediksi terhadap kondisi penjualan obat per bulan di masa yang akan datang. Nilai *Root Mean Squared Error* (RMSE) yang didapat saat melakukan evaluasi performa model adalah sebesar $0,452 \pm 0,000$ melalui aplikasi *RapidMiner Studio*. Nurrika Riskya, Selvira Yuliana memprediksi perilaku pelanggan terkait dengan keinginan untuk membeli produk Abaya dapat dilakukan dengan mempertimbangkan faktor-faktor seperti model, jenis kelamin, berat badan, SKU, panjang baju (PB), lebar dada (LD), dan harga. Penelitian ini menggunakan Algoritma Regresi Linier Berganda, yang hasilnya menunjukkan akurasi 100%. Dengan demikian, model tersebut mampu memprediksi perilaku pelanggan dengan tingkat keakuratan yang sangat tinggi. Miftahuljannah, Aswan Supriyadi Sunge, Ahmad Turmudi (Riskya & Yuliana, 2023) menggunakan algoritma *regresi linear* untuk memprediksi penjualan dan hasilnya menunjukkan bahwa variabel atau atribut yang digunakan dalam penelitian ini, yaitu produk, aktual, dan rencana, tidak memiliki dampak signifikan pada hasilnya. RMSE sebesar 36241.241 ± 0.000 , dan Squared Error sebesar $1313427569.481 \pm 5882150128.134$.

Selanjutnya metode *regresi linear* juga dapat digunakan untuk memprediksi penjualan kendaraan menggunakan analisis sentimen dari berbagai tempat di internet. Kehadiran online sebuah kendaraan, serta mereknya, memainkan peran kunci dalam penjualan kendaraan tersebut. Prediksi penjualan di pasar saat ini tidak hanya bermanfaat bagi produsen tetapi juga bagi berbagai perusahaan lain yang memproduksi suku cadang atau aksesoris kendaraan (*Vaigai College of Engineering & Institute of Electrical and Electronics Engineers*, n.d.). Penelitian lainnya juga dilakukan oleh Gopalakrishnan, Ritesh Choudhary, dan Sarada Prasad yang menganalisis penjualan sebuah superstore besar dan memprediksi penjualan mereka di masa depan untuk membantu mereka meningkatkan keuntungan dan membuat merek mereka lebih baik dan kompetitif sesuai dengan tren pasar dengan juga menghasilkan kepuasan pelanggan, Teknik yang digunakan untuk prediksi penjualan adalah Algoritma *Regresi linear*. Data penjualan berasal dari tahun 2011-2013 dan penelitian menghasilkan nilai error 16.168% atau tingkat akurasi sebesar 84% (*Galgotias University. School of Computing Science and Engineering et al.*, n.d.). Berdasarkan rujukan penelitian-penelitian terdahulu yang telah berhasil menerapkan algoritma *Regresi linear* dalam berbagai penelitian, maka penulis bermaksud untuk melakukan penelitian pada PT. Toyo Mortar Indonesia dengan menerapkan algoritma *Regresi linear* dalam memprediksi penjualan produk semen dengan harapan metode *regresi linear* dapat membantu PT. Toyo Mortar Indonesia mengatasi tantangan penurunan penjualan dengan menyediakan analisis tren penjualan yang mendalam dari data historis, sehingga perusahaan dapat mengidentifikasi pola

penurunan atau peningkatan penjualan. *Regresi linear* juga membantu mengidentifikasi faktor-faktor kunci yang mempengaruhi penjualan, memungkinkan perusahaan untuk fokus pada variabel-variabel signifikan seperti jenis produk atau periode waktu tertentu.

2. Metode Penelitian

Pada penelitian ini metode yang digunakan adalah *regresi linear*. Metode regresi adalah suatu pendekatan statistik yang digunakan untuk memproyeksikan atau memprediksi nilai dari suatu variabel tergantung (variabel dependen) berdasarkan hubungannya dengan satu atau lebih variabel yang mempengaruhi (variabel independen) (Najla *et al.*, n.d.). Pada tahapan penelitian ini akan dilakukan dalam penerapan metode CRISP-DM. CRISP-DM, yang merupakan singkatan dari Cross-Industry Standard Process for Data Mining, dikembangkan pada tahun 1996 oleh sekelompok analis dari industri yang berbeda, termasuk Daimler Chrysler, SPSS, dan NCR. CRISP-DM bertujuan untuk menyediakan standar proses data mining yang dapat digunakan sebagai strategi umum untuk memecahkan masalah di berbagai bidang bisnis atau unit penelitian (Feblian & Daihani, 2017). Untuk itu, penulis telah merancang dan menyusun prosedur kegiatan dalam bentuk alur desain penerapan metodologi sebagai berikut:



Gambar 1. Model Proses Penelitian

Tahapan awal dalam proses data mining dimulai dengan menetapkan tujuan, menerjemahkan tujuan tersebut menjadi definisi permasalahan, serta mengidentifikasi masalah yang ada. Setelah itu, data dikumpulkan, dilakukan analisis awal untuk mengidentifikasi informasi penting, mengevaluasi kualitas data, dan memilih subset data yang relevan untuk digunakan. Selanjutnya, data disiapkan untuk analisis dengan cara memilih kasus dan variabel yang sesuai, melakukan transformasi data jika diperlukan, serta mempersiapkan data tersebut untuk pemodelan. Setelah data siap, teknik pemodelan yang sesuai diterapkan, model dikalibrasi untuk meningkatkan kinerja, dan teknik yang digunakan dievaluasi. Kemudian, kualitas model yang dihasilkan dievaluasi untuk memastikan apakah tujuan awal telah tercapai, serta membuat keputusan terkait penggunaan hasil data mining tersebut. Pada tahap akhir, model yang dihasilkan dapat digunakan untuk divisualisasikan dalam bentuk laporan atau diagram yang membantu memahami hasilnya secara lebih jelas.

3. Hasil dan Pembahasan

3.1 Hasil

Pada tahap ini, proses implementasi dan pengujian metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) akan dilakukan. Metodologi ini terdiri dari enam tahapan, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Berikut adalah tahapan-tahapan implementasi dan pengujiannya:

3.1.1 *Business Understanding* (Pemahaman Bisnis)

Tahap ini berfokus pada pemahaman awal mengenai masalah dan tujuan penelitian. Penulis telah mengidentifikasi beberapa permasalahan dalam penelitian ini, antara lain:

- 1) PT. Toyo Mortar memiliki data penjualan yang besar, namun data tersebut belum diolah untuk memprediksi penjualan semen.
- 2) Variabel independen dan variabel dependen yang akan digunakan dalam prediksi belum diidentifikasi secara jelas.
- 3) Belum teridentifikasi apakah metode *regresi linear* efektif dalam melakukan prediksi dengan akurat.

Tujuan dari penelitian ini adalah untuk mengatasi permasalahan tersebut dengan mengolah data penjualan yang ada, mengidentifikasi variabel yang relevan, dan mengevaluasi efektivitas metode *regresi linear* dalam memprediksi penjualan semen pada PT. Toyo Mortar Indonesia.

3.1.2 *Data Understanding* (Pemahaman Data)

Dalam pengumpulan dan eksplorasi data yang tersedia, penelitian ini menggunakan data penjualan semen tahun 2023 yang diperoleh langsung dari PT. Toyo Mortar Indonesia, yang berlokasi di Jl. Padat Karya No.63, PLP Curug KM 4,5, Tangerang 15810, Banten, Indonesia. Data tersebut diekspor dari database perusahaan dalam format file *xlsx*. Dataset yang digunakan dalam penelitian ini adalah data kualitatif yang dikumpulkan melalui tiga teknik: studi pustaka, observasi, dan wawancara. Sampel yang diambil terdiri dari 315 data dengan 15 atribut.

Tabel 1. Atribut Dataset

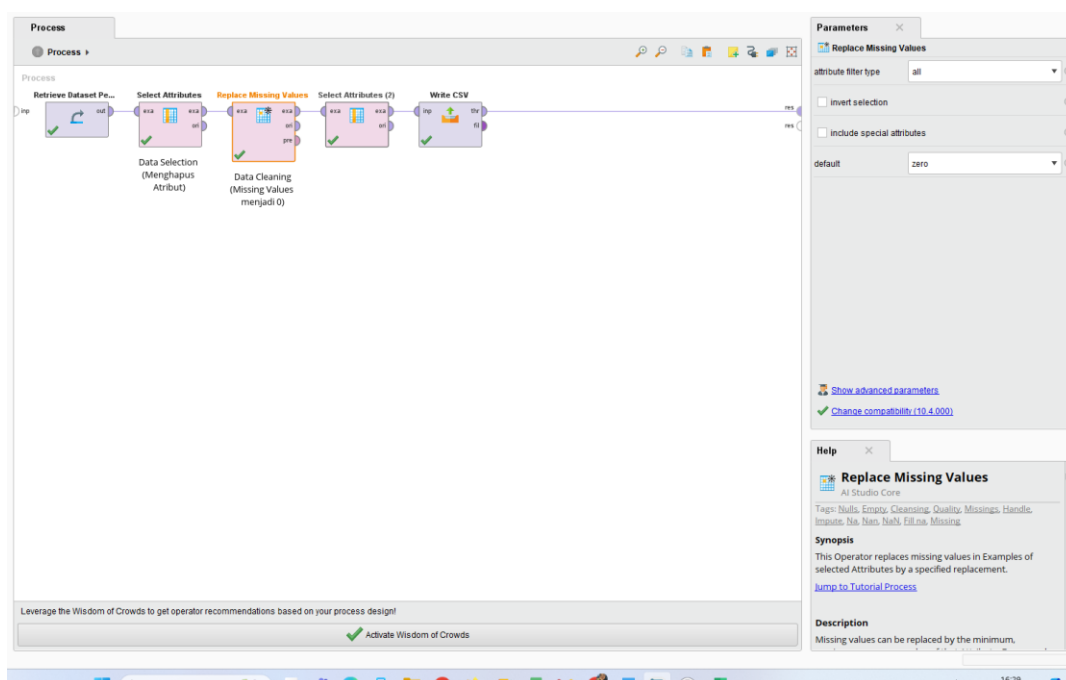
Atribut	Keterangan
Tanggal	Berisi data waktu dilakukannya transaksi penjualan
TM-168	Jumlah perekat bata ringan terjual
TM-178	Jumlah pasang bata dan plester terjual
TM-188	Jumlah acian plesteran dan beton terjual
TM-189	Jumlah acian plesteran dan beton putih terjual

TM-198	Jumlah perekat keramik dinding terjual
UM-100	Jumlah ultra Mix 100 terjual
UM-101	Jumlah ultra Mix 100 Biru terjual
UM BIRU	Jumlah ultra Mix Acian terjual
Bahan Baku	Jumlah bahan baku terjual
Penjualan lain-lain	Jumlah penjualan lainnya terjual
Penjualan	Total Penjualan
Jangka Tempo	Periode waktu pembayaran
Jatuh Tempo	Tanggal akhir ketika pembayaran
Kode	Kode marketing

3.1.3 Data Preparation (Persiapan Data)

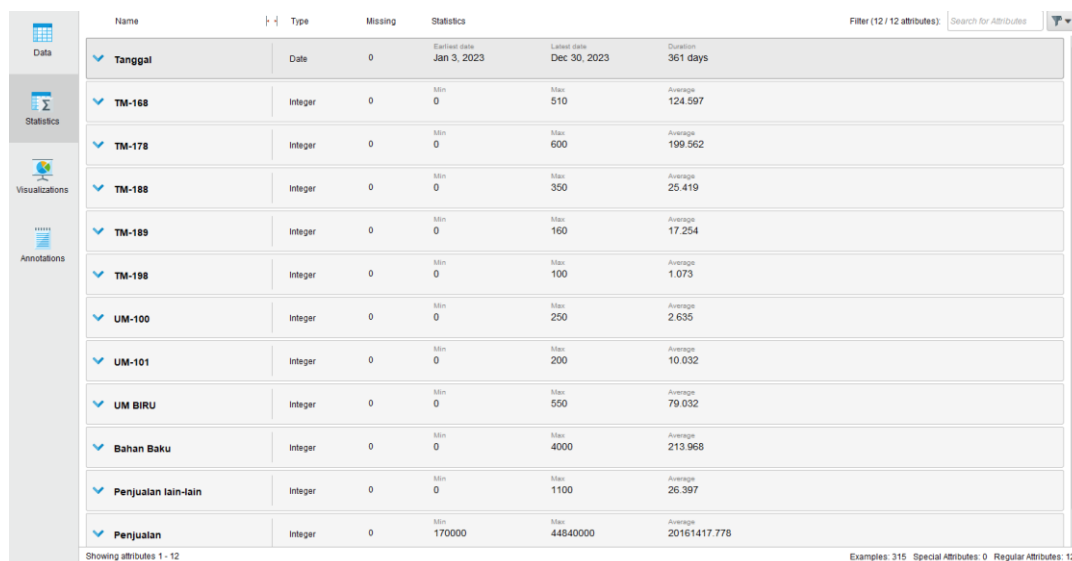
Pada tahap ini, data dipersiapkan dan diubah agar siap digunakan dalam pemodelan.

- 1) Tahap ini melibatkan pembersihan data dengan mengidentifikasi, memperbaiki, dan menghapus kesalahan dalam data. Data Cleaning yang dilakukan yaitu menghapus missing values. Proses ini dilakukan menggunakan RapidMiner seperti pada gambar berikut:



Gambar 2. Data Cleaning

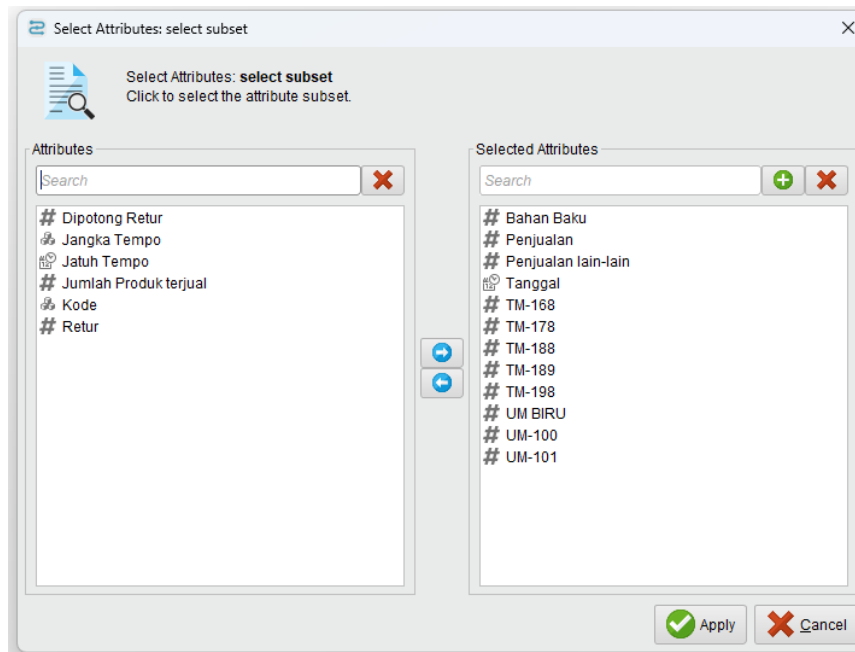
Missing values yang dihapus akan berubah menjadi 0 sehingga hasilnya sebagai berikut:



Name	Type	Missing	Statistics
Tanggal	Date	0	Earliest date: Jan 3, 2023 Latest date: Dec 30, 2023 Duration: 361 days
TM-168	Integer	0	Min: 0 Max: 510 Average: 124.597
TM-178	Integer	0	Min: 0 Max: 600 Average: 199.562
TM-188	Integer	0	Min: 0 Max: 350 Average: 25.419
TM-189	Integer	0	Min: 0 Max: 160 Average: 17.254
TM-198	Integer	0	Min: 0 Max: 100 Average: 1.073
UM-100	Integer	0	Min: 0 Max: 250 Average: 2.635
UM-101	Integer	0	Min: 0 Max: 200 Average: 10.032
UM BIRU	Integer	0	Min: 0 Max: 550 Average: 79.032
Bahan Baku	Integer	0	Min: 0 Max: 4000 Average: 213.968
Penjualan lain-lain	Integer	0	Min: 0 Max: 1100 Average: 26.397
Penjualan	Integer	0	Min: 170000 Max: 44840000 Average: 20161417.778

Gambar 3. Hasil Data *Cleaning*

- 2) Pada tahap ini, data dipilih dari sekumpulan dataset yang ada sebelum melanjutkan ke tahap berikutnya. Proses ini penting karena tidak semua data diperlukan untuk proses data mining. Setelah melakukan identifikasi pada dataset, ditemukan beberapa atribut yang perlu dihilangkan. Atribut yang ditambahkan meliputi tanggal, TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM-101, UM BIRU, bahan baku, penjualan lain-lain, dan penjualan. Sementara itu, beberapa atribut dihapus karena tidak diperlukan, seperti retur, jangka tempo, jatuh tempo, dan kode. Dalam proses analisis dataset untuk memprediksi penjualan, beberapa variabel dipilih dan beberapa dihapus berdasarkan relevansi dan dampaknya terhadap penjualan. Variabel seperti tanggal dipertahankan karena memungkinkan analisis tren dan pola musiman, membantu dalam mengidentifikasi fluktuasi penjualan berdasarkan waktu. Variabel TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM-101, dan UM BIRU mewakili kategori atau fitur produk tertentu yang mempengaruhi penjualan. Memasukkan variabel ini memungkinkan analisis mendalam mengenai kontribusi masing-masing kategori produk terhadap penjualan total. Selain itu, variabel bahan baku juga dipilih karena ketersediaan atau biaya bahan baku dapat memengaruhi produksi dan harga jual produk, berdampak langsung pada jumlah penjualan. Penjualan lain-lain memberikan gambaran tentang sumber pendapatan tambahan, sedangkan penjualan itu sendiri adalah variabel utama yang ingin diprediksi. Sebaliknya, variabel seperti retur, jangka tempo, jatuh tempo, dan kode dihapus karena dianggap kurang relevan untuk model prediksi penjualan. Retur mungkin penting untuk analisis kepuasan pelanggan, tetapi tidak langsung mempengaruhi prediksi penjualan saat ini. Jangka tempo dan jatuh tempo lebih berkaitan dengan manajemen piutang daripada penjualan produk, sehingga kurang berpengaruh pada analisis penjualan langsung. Kode mungkin tidak memberikan informasi tambahan yang relevan, sehingga dihilangkan untuk menyederhanakan model. Proses pembersihan data ini masih dilakukan menggunakan RapidMiner. Untuk lebih jelasnya dapat dilihat pada gambar 4.



Gambar 4. Data Selection

Setelah melakukan data *cleaning* dan data *selection*, maka data yang dapat digunakan untuk proses selanjutnya dapat dilihat pada gambar 5.

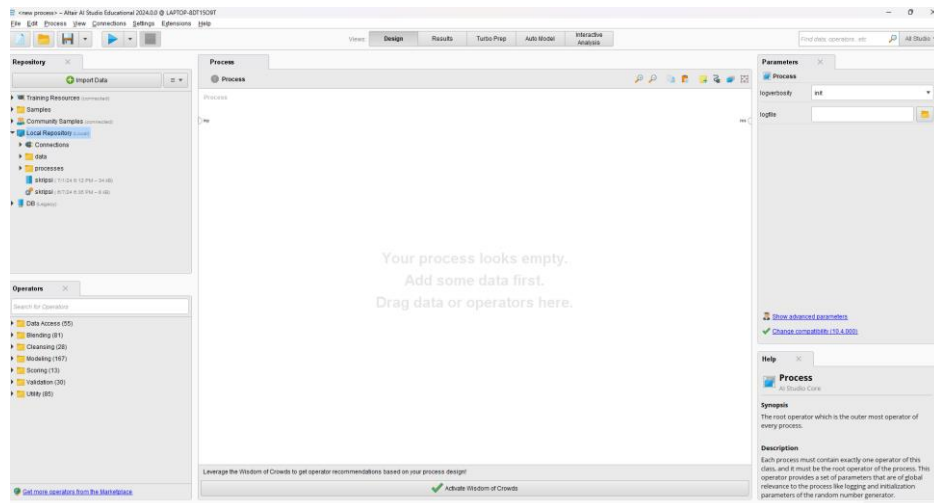
Row No.	Tanggal	TM-168	TM-178	TM-188	TM-189	TM-198	UM-100	UM-101	UM BIRU	Bahan Baku	Penjualan la...	Penjualan
1	Jan 3, 2023	65	0	50	0	0	0	0	200	0	0	13815000
2	Jan 4, 2023	65	0	50	15	0	0	0	200	0	0	14240000
3	Jan 5, 2023	0	200	50	20	0	0	0	220	0	0	21090000
4	Jan 6, 2023	50	0	0	10	0	0	30	240	0	0	14035000
5	Jan 7, 2023	100	0	0	0	0	0	0	100	0	0	8455000
6	Jan 9, 2023	130	250	0	100	0	0	0	150	0	0	28055000
7	Jan 10, 2023	5	500	10	50	0	0	0	50	0	0	25835000
8	Jan 11, 2023	210	250	0	30	0	0	0	170	0	0	28310000
9	Jan 12, 2023	85	500	10	30	0	0	0	0	4000	0	29760000
10	Jan 13, 2023	150	200	0	50	0	0	0	130	0	0	22870000
11	Jan 14, 2023	300	100	100	0	0	0	0	80	0	0	26210000
12	Jan 16, 2023	50	0	0	0	0	0	50	0	0	0	4750000
13	Jan 17, 2023	100	0	0	100	0	0	0	280	0	0	21520000
14	Jan 18, 2023	114	0	0	0	0	250	80	250	0	0	29967500
15	Jan 19, 2023	105	250	0	0	0	0	0	0	0	0	14575000
16	Jan 20, 2023	120	500	0	10	0	0	0	0	0	0	26010000
17	Jan 21, 2023	75	250	75	50	0	0	0	0	0	0	18536000
18	Jan 23, 2023	65	250	0	15	0	0	0	50	0	0	15980000
19	Jan 24, 2023	0	400	10	30	0	0	0	100	40	0	24520000
20	Jan 25, 2023	0	450	150	0	0	0	0	100	0	0	31454000
21	Jan 26, 2023	150	250	0	30	0	0	0	150	0	770	25468000
22	Jan 27, 2023	320	0	150	0	0	0	0	0	0	0	22765000
23	Jan 28, 2023	0	600	0	0	0	0	0	30	0	0	25170000

ExampleSet (315 examples, 0 special attributes, 12 regular attributes)

Gambar 5. Data yang siap diolah

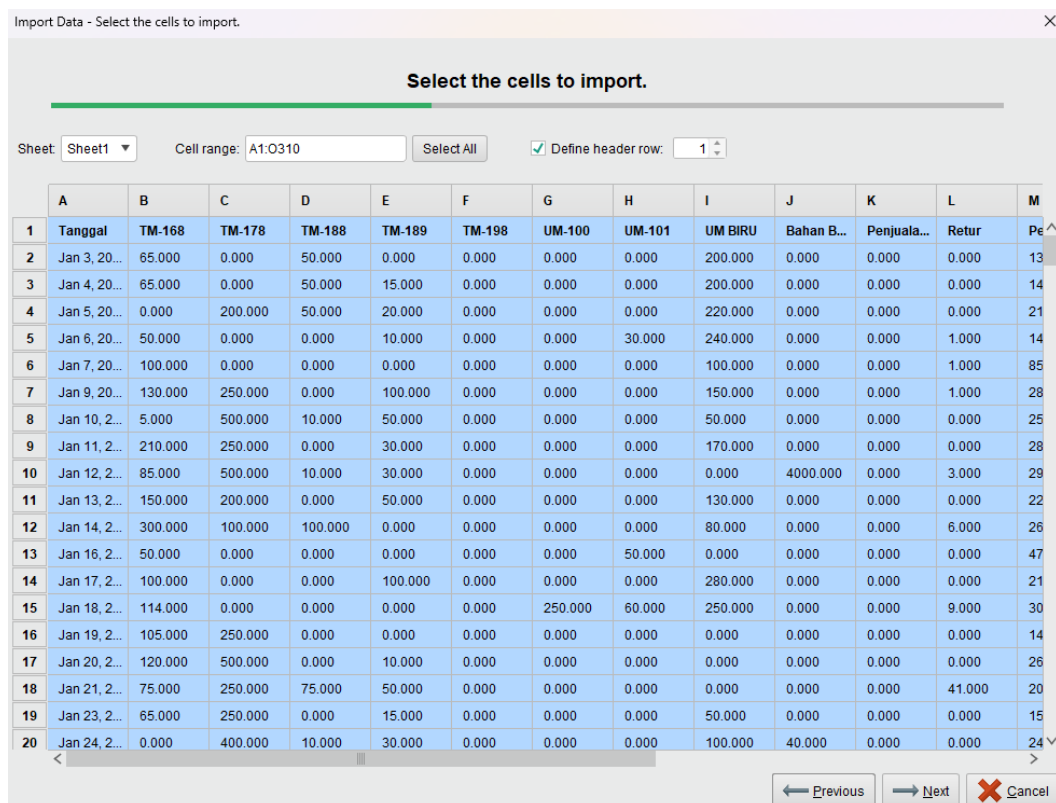
- 3) Data Transformation pada tahap ini, dilakukan perubahan format, skala, atau struktur data agar sesuai dengan kebutuhan analisis atau model yang akan digunakan. Atribut "Tanggal" dijadikan ID, sementara atribut "Penjualan" dijadikan label yang akan digunakan dalam aplikasi RapidMiner.

- 4) Import Data pada RapidMiner yang digunakan sebagai alat pembelajaran dalam ilmu data mining. Berikut adalah tampilan awal aplikasi RapidMiner saat pertama kali dibuka, yang dapat dilihat pada gambar berikut:



Gambar 6. Tampilan Awal Aplikasi RapidMiner

Langkah pertama yang dilakukan adalah mengimpor data. Dataset yang telah disiapkan kemudian dimasukkan ke dalam *Repository RapidMiner*. *Repository* ini berfungsi sebagai lokasi penyimpanan terpusat untuk data dalam aplikasi RapidMiner.



Gambar 7. Import Data

Setelah itu, tentukan tipe variabel untuk setiap atribut. Atribut "Tanggal" menggunakan tipe date, sedangkan atribut TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM-101, UM BIRU, bahan baku, penjualan lain-lain, dan penjualan menggunakan tipe variabel integer karena berisi data angka yang dapat dihitung. Atribut "Penjualan" akan dijadikan sebagai label karena atribut ini yang akan digunakan untuk menghasilkan keluaran prediksi.

Import Data - Format your columns.

Format your columns.

☐ Replace errors with missing values ⓘ

	Tanggal date id	TM-168 integer	TM-178 integer	TM-188 integer	TM-189 integer	TM-198 integer	UM-100 integer	UM-101 integer
1	Jan 3, 2023	65	0	50	0	0	0	0
2	Jan 4, 2023	65	0	50	15	0	0	0
3	Jan 5, 2023	0	200	50	20	0	0	0
4	Jan 6, 2023	50	0	0	10	0	0	30
5	Jan 7, 2023	100	0	0	0	0	0	0
6	Jan 9, 2023	130	250	0	100	0	0	0
7	Jan 10, 2023	5	500	10	50	0	0	0
8	Jan 11, 2023	210	250	0	30	0	0	0
9	Jan 12, 2023	85	500	10	30	0	0	0
10	Jan 13, 2023	150	200	0	50	0	0	0
11	Jan 14, 2023	300	100	100	0	0	0	0
12	Jan 16, 2023	50	0	0	0	0	0	50
13	Jan 17, 2023	100	0	0	100	0	0	0
14	Jan 18, 2023	114	0	0	0	0	250	60
15	Jan 19, 2023	105	250	0	0	0	0	0
16	Jan 20, 2023	120	500	0	10	0	0	0
17	Jan 21, 2023	75	250	75	50	0	0	0
18	Jan 23, 2023	65	250	0	15	0	0	0

no problems.

Previous Finish Cancel

Gambar 8. Menentukan Tipe Variabel

Penjualan

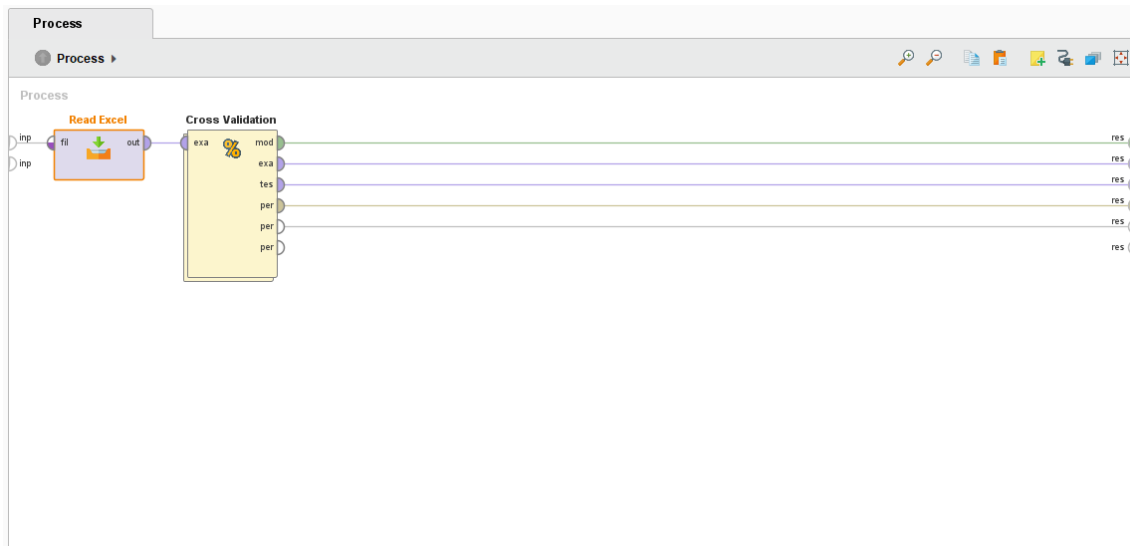
integer
label

4030000
31790000
25770000
10725000
23300000
28547000
29150000
29887500

Gambar 9. Role Label

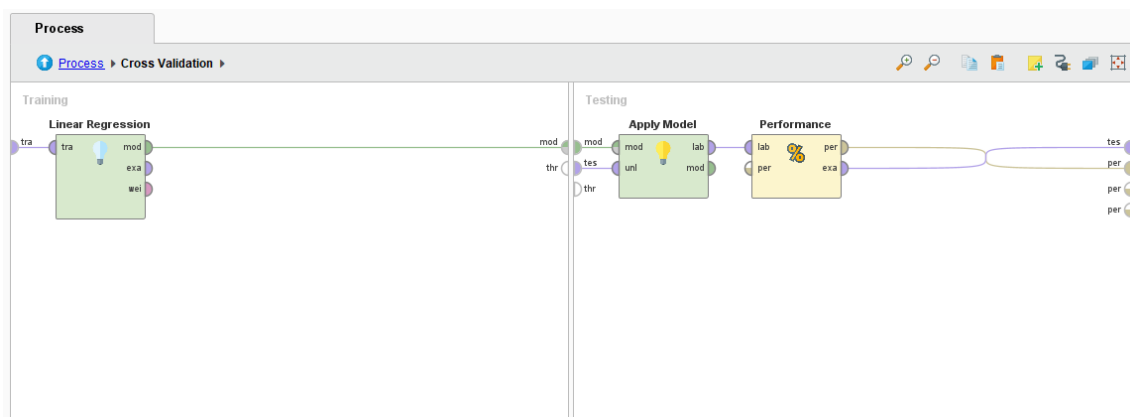
3.1.4 Modelling

Tahap pemodelan melibatkan penggunaan algoritma dan teknik pada dataset yang telah dipersiapkan. Algoritma yang digunakan adalah *Linear Regression*, di mana variabel independen digunakan sebagai fitur dan variabel dependen sebagai target. Pembagian data menjadi data training dan data uji tidak dilakukan secara langsung, melainkan menggunakan operator *Cross Validation*. Operator ini memiliki dua subproses utama: pelatihan model dan pengujian untuk mengukur kinerja model.



Gambar 10. *Cross Validation*

Pada tahap ini, dataset terkoneksi dengan operator *Cross Validation* untuk membagi data secara otomatis. Setelah itu, model *Linear Regression* diterapkan pada data training menggunakan operator *Apply Model* pada bagian pengujian. Untuk mengevaluasi kinerja model, operator *Performance* digunakan untuk memberikan nilai kriteria kinerja secara otomatis, dan memastikan koneksi kabel yang benar untuk menghindari kesalahan teknis.



Gambar 11. Implementasi Metode *Linear Regression*

3.1.5 Evaluation

Setelah dilakukan pengujian menggunakan aplikasi rapidminer, maka didapatkan beberapa variabel yang mempengaruhi keluaran prediksi.

LinearRegression

44127.389 * TM-168
+ 40189.110 * TM-178
+ 55747.986 * TM-188
+ 60144.776 * TM-189
+ 143987.850 * TM-198
+ 48015.941 * UM-100
+ 46725.419 * UM-101
+ 40500.331 * UM BIRU
+ 854.939 * Bahan Baku
+ 1280.638 * Penjualan lain-lain
+ 20969.583

Gambar 12. Variabel yang mempengaruhi

Berikut untuk nilai intercept dan coefficients dari variable yang mempengaruhi keluaran prediksi:

Tabel 2. Nilai *Intercept* dan *Coefficients*

Variabel	Atribut	Coefficients (β)
X1	TM-168	44127,389
X2	TM-178	40189,110
X3	TM-188	55747,986
X4	TM-189	60144,776
X5	TM-198	143987,850
X6	UM-100	48015,941
X7	UM-101	46725,419
X8	UM BIRU	40500,331
X9	Bahan Baku	854,939
X10	Penjualan lain-lain	1280,638
Intercept (β_i)		20969,583

Pada penelitian ini, menggunakan Algoritma *Regresi linear* Berganda. *Regresi linear* Berganda sebenarnya sama dengan *regresi linear* sederhana, hanya saja variable bebasnya lebih dari satu. Berikut adalah perhitungan prediksi dengan *Regresi linear* Berganda yang memiliki rumus:

$$y=\beta_1+\beta_2 X_2+\beta_3 X_3+\dots \beta_n X_n$$

Dimana:

y = Variabel dependent (terikat)

β_1 = intercept dari model

$\beta_2,\beta_3,\dots\beta_n$ = Koefisien-koefisien parsial dari variable dependen ke-1

$X_2,X_3,\dots X_n$ = Variabel-variabel independent ke-1 dengan parameternya

Kemudian nilai Intercept dan Coefficients yang sudah didapatkan pada table di atas dapat dimasukkan ke dalam rumus tersebut.

- 1) $y = 20969,583 + (65 \times 44127,389) + (0 \times 40189,110) + (50 \times 55747,986) + (0 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (200 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 13.776.715,368$
- 2) $y = 20969,583 + (65 \times 44127,389) + (0 \times 40189,110) + (50 \times 55747,986) + (15 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (200 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 14.678.887,008$
- 3) $y = 20969,583 + (0 \times 44127,389) + (200 \times 40189,110) + (50 \times 55747,986) + (20 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (220 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 20.959.159,223$
- 4) $y = 20969,583 + (50 \times 44127,389) + (0 \times 40189,110) + (0 \times 55747,986) + (10 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (30 \times 46725,419) + (240 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 13.950.628,803$
- 5) $y = 20969,583 + (100 \times 44127,389) + (0 \times 40189,110) + (0 \times 55747,986) + (15 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (100 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 9.385.913,223$
- 6) $y = 20969,583 + (130 \times 44127,389) + (250 \times 40189,110) + (0 \times 55747,986) + (100 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (150 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 27.894.334,903$
- 7) $y = 20969,583 + (5 \times 44127,389) + (500 \times 40189,110) + (10 \times 55747,986) + (50 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (50 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 25.925.896,738$
- 8) $y = 20969,583 + (210 \times 44127,389) + (250 \times 40189,110) + (0 \times 55747,986) + (30 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (170 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 28.024.398,323$
- 9) $y = 20969,583 + (85 \times 44127,389) + (500 \times 40189,110) + (10 \times 55747,986) + (30 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (0 \times 40500,331) + (4000 \times 854,939) + (0 \times 1280,638)$
 $= 29.647.931,788$
- 10) $y = 20969,583 + (150 \times 44127,389) + (200 \times 40189,110) + (0 \times 55747,986) + (50 \times 60144,776) + (0 \times 143987,850) + (0 \times 48015,941) + (0 \times 46725,419) + (130 \times 40500,331) + (0 \times 854,939) + (0 \times 1280,638)$
 $= 22.950.181,763$

Penulis juga membuat perhitungan prediksi secara manual dengan menggunakan Microsoft excel, dapat dilihat pada gambar berikut:

Tanggal	TM-168	TM-178	TM-188	TM-189	TM-198	UM-100	UM-101	UM BIRU	Bahan Baku	Penjualan lain-lai	Penjualan	Prediksi Manual
3 Jan 2023	65	0	50	0	0	0	0	200	0	0	13815000,000	13776715,368
4 Jan 2023	65	0	50	15	0	0	0	200	0	0	14240000,000	14678887,008
5 Jan 2023	0	200	50	20	0	0	0	220	0	0	21090000,000	20959159,223
6 Jan 2023	50	0	0	10	0	0	30	240	0	0	14035000,000	13950628,803
7 Jan 2023	100	0	0	0	0	0	0	100	0	0	8455000,000	8483741,583
9 Jan 2023	130	250	0	100	0	0	0	150	0	0	28055000,000	27894334,903
10 Jan 2023	5	500	10	50	0	0	0	50	0	0	25835000,000	25925896,738
11 Jan 2023	210	250	0	30	0	0	0	170	0	0	28310000,000	28024398,323
12 Jan 2023	85	500	10	30	0	0	0	0	4000	0	29760000,000	29647931,788
13 Jan 2023	150	200	0	50	0	0	0	130	0	0	22870000,000	22950181,763
14 Jan 2023	300	100	100	0	0	0	0	80	0	0	26210000	26092922,363
16 Jan 2023	50	0	0	0	0	0	50	0	0	0	4750000	4563609,983
17 Jan 2023	100	0	0	100	0	0	0	280	0	0	21520000	21788278,763
18 Jan 2023	114	0	0	0	0	250	60	250	0	0	29967500	29984085,069
19 Jan 2023	105	250	0	0	0	0	0	0	0	0	14575000	14701622,928
20 Jan 2023	120	500	0	10	0	0	0	0	0	0	26010000	26012259,023
21 Jan 2023	75	250	75	50	0	0	0	0	0	0	18536000	20566139,008
23 Jan 2023	65	250	0	15	0	0	0	50	0	0	15980000	15863715,558
24 Jan 2023	0	400	10	30	0	0	0	100	40	0	24520000	22542667,383
25 Jan 2023	0	450	150	0	0	0	0	100	0	0	31454000	30518300,083
26 Jan 2023	150	250	0	30	0	0	0	150	0	770	25468000	25552839,623

Gambar 13. Hasil Prediksi Secara Manual

Dapat dilihat pada gambar di bawah untuk perbandingan dengan perhitungan prediksi yang dihasilkan dengan *RapidMiner*.

Tanggal ↑	Penjualan	prediction(P...	TM-168	TM-178	TM-188	TM-189	TM-198	UM-100	UM-101	UM BIRU	Bahan Baku	Penjualan la...
Jan 3, 2023	13815000	13769567.665	65	0	50	0	0	0	0	200	0	0
Jan 4, 2023	14240000	14685222.039	65	0	50	15	0	0	0	200	0	0
Jan 5, 2023	21090000	20997270.884	0	200	50	20	0	0	0	220	0	0
Jan 6, 2023	14035000	13928247.211	50	0	0	10	0	0	30	240	0	0
Jan 7, 2023	8455000	8468991.935	100	0	0	0	0	0	0	100	0	0
Jan 9, 2023	28055000	27876342.821	130	250	0	100	0	0	0	150	0	0
Jan 10, 2023	25835000	25982189.738	5	500	10	50	0	0	0	50	0	0
Jan 11, 2023	28310000	28018571.314	210	250	0	30	0	0	0	170	0	0
Jan 12, 2023	29760000	29527961.986	85	500	10	30	0	0	0	0	4000	0
Jan 13, 2023	22870000	22953155.104	150	200	0	50	0	0	0	130	0	0
Jan 14, 2023	26210000	26117980.074	300	100	100	0	0	0	0	80	0	0
Jan 16, 2023	4750000	4603394.506	50	0	0	0	0	0	50	0	0	0
Jan 17, 2023	21520000	21782209.692	100	0	0	100	0	0	0	280	0	0
Jan 18, 2023	29967500	30056718.968	114	0	0	0	0	250	60	250	0	0
Jan 19, 2023	14575000	14719153.195	105	250	0	0	0	0	0	0	0	0
Jan 20, 2023	26010000	26025814.702	120	500	0	10	0	0	0	0	0	0
Jan 21, 2023	18536000	20606945.249	75	250	75	50	0	0	0	0	0	0
Jan 23, 2023	15980000	15865665.827	65	250	0	15	0	0	0	50	0	0
Jan 24, 2023	24520000	22531421.577	0	400	10	30	0	0	0	100	40	0
Jan 25, 2023	31454000	30446278.699	0	450	150	0	0	0	0	100	0	0
Jan 26, 2023	25468000	25494540.931	150	250	0	30	0	0	0	150	0	770
Jan 27, 2023	22765000	22497871.328	320	0	150	0	0	0	0	0	0	0
Jan 28, 2023	25170000	25338101.635	0	600	0	0	0	0	0	30	0	0
Jan 29, 2023	4400000	4439755.808	100	0	0	0	0	0	0	0	0	0

Gambar 14. Hasil Prediksi dengan RapidMiner

Berikut adalah tahap selanjutnya dalam analisis, yaitu pengujian dan evaluasi menggunakan metrik *Root Mean Square Error* (RMSE) untuk mengukur tingkat akurasi prediksi.

	PerformanceVector
Performance	PerformanceVector:
	root_mean_squared_error: 493125.701 +/- 126330.525 (micro average: 507745.542 +/- 0.000)
	absolute_error: 303270.965 +/- 46207.586 (micro average: 303226.855 +/- 407257.916)
	relative_error: 1.89% +/- 0.36% (micro average: 1.89% +/- 2.86%)
	Description
	Annotations

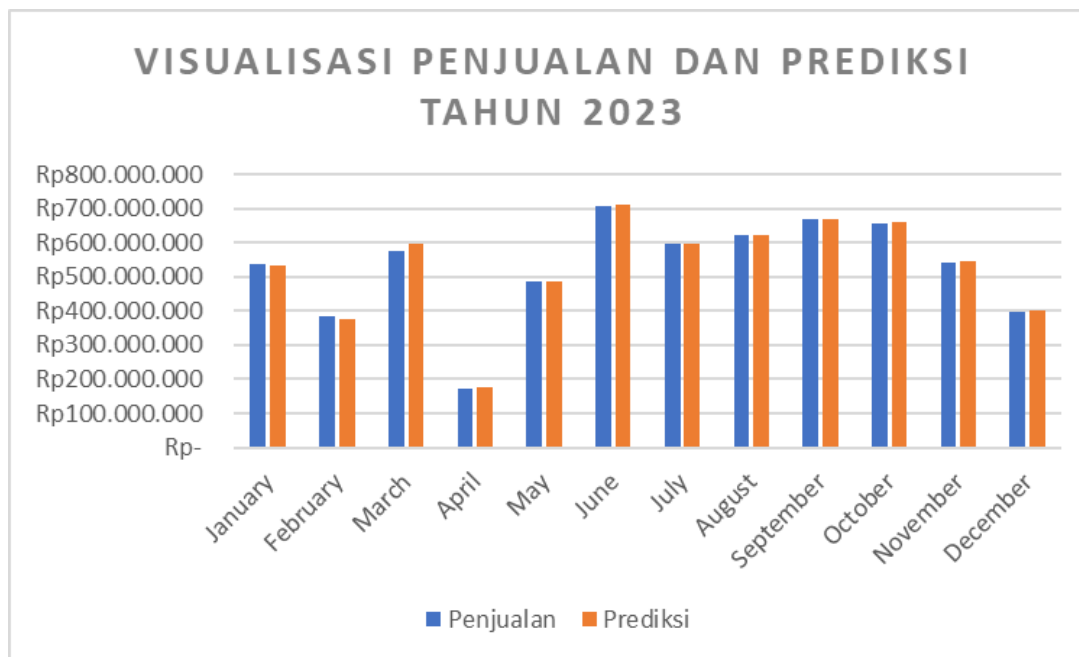
Gambar 15. Performance Vector

- Dengan menggunakan operator performa regresi pada RapidMiner, hasil evaluasi menunjukkan:
- 1) RMSE 493,125.701 +/- 126,330.525: Nilai ini menunjukkan bahwa rata-rata kesalahan kuadrat akar dari prediksi model adalah sekitar 493,125.701. Standar deviasi dari RMSE, menunjukkan variasi dalam kesalahan prediksi. Sebagian besar nilai RMSE berada dalam rentang 493,125.701 $\pm$ 126,330.525, yaitu antara 366,795.176 dan 619,456.226.
 - 2) Absolute Error 303,270.965 +/- 46,207.586: Rata-rata kesalahan absolut model adalah sekitar 303,270.965. Standar deviasi dari absolute error, menunjukkan seberapa jauh nilai-nilai individual dari absolute error menyebar dari rata-rata. Sebagian besar absolute error berada dalam rentang 303,270.965 $\pm$ 46,207.586, yaitu antara 257,063.379 dan 349,478.551.

- 3) Relative Error 1.89% +/- 0.36%: Nilai ini sangat baik karena menunjukkan bahwa rata-rata kesalahan relatif prediksi hanya sekitar 1.89% dari nilai penjualan sebenarnya.

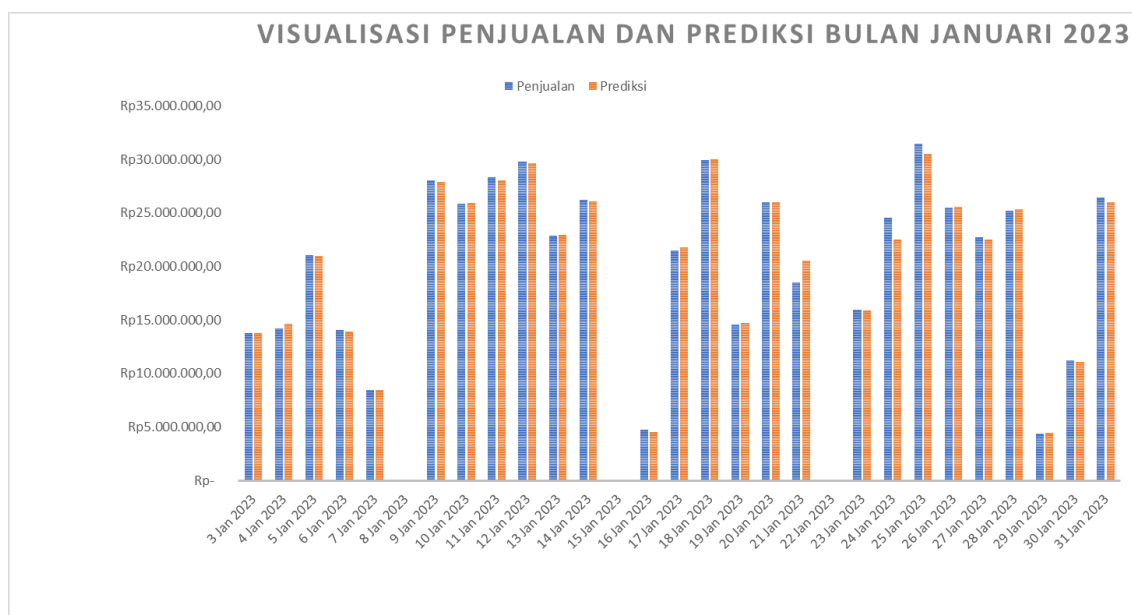
3.1.6 Deployment

Diagram *Clustered Column* menunjukkan bahwa skema data Penjualan dan Prediksi memiliki kemiripan. Hal ini menunjukkan bahwa penggunaan metode *Regresi linear* untuk memprediksi penjualan semen di PT. Toyo Mortar Indonesia berhasil.



Gambar 16. *Clustered Column Chart* Tahun 2023

Untuk lebih jelasnya, berikut adalah visualisasi data untuk bulan Januari tahun 2023.



Gambar 17. *Clustered Column Chart* Januari Tahun 2023

3.1.7 Hasil Akhir Pengujian

Dalam pengujian prediksi jumlah penjualan semen di PT. Toyo Mortar Indonesia menggunakan model algoritma *Regresi linear*, hasil yang diperoleh menunjukkan nilai *Root Mean Squared Error* (RMSE) sebesar 493.125,701, *Absolute Error* sebesar 303.270,965, dan *Relative Error* sebesar 1,89%. RMSE, sebagai ukuran kesalahan prediksi, mengukur seberapa besar deviasi antara nilai yang diprediksi dan nilai aktual dalam dataset. Nilai RMSE yang rendah menandakan bahwa model memiliki tingkat kesalahan yang kecil dan dengan demikian menunjukkan akurasi yang tinggi dalam prediksi penjualan. Dalam hal ini, nilai RMSE yang relatif rendah mengindikasikan bahwa model regresi linier yang digunakan cukup efektif dalam meramalkan penjualan dengan kesalahan yang minim. *Absolute Error*, yang menunjukkan rata-rata kesalahan absolut dalam unit yang sama dengan data penjualan, juga menunjukkan ukuran kesalahan prediksi yang terukur. Dalam konteks ini, nilai *Absolute Error* sebesar 303.270,965 menandakan bahwa rata-rata deviasi dari nilai prediksi terhadap nilai aktual berada dalam rentang yang dapat diterima. *Relative Error*, yang dinyatakan sebagai persentase, menunjukkan proporsi kesalahan relatif terhadap nilai aktual, dengan nilai 1,89% menunjukkan bahwa kesalahan prediksi model sangat kecil dibandingkan dengan nilai penjualan aktual.

3.2 Pembahasan

Penelitian ini menggunakan metode *regresi linear* untuk memprediksi penjualan produk semen di PT. Toyo Mortar Indonesia. Metode ini dipilih berdasarkan keunggulannya dalam memprediksi hubungan antara variabel independen dan variabel dependen, seperti yang juga dibuktikan dalam berbagai studi terdahulu. Misalnya, penelitian Chang Hartono dan Dwiyo Widianoro (2024) menunjukkan bahwa *regresi linear* efektif dalam memprediksi harga saham Unilever, dengan nilai akurasi yang cukup tinggi. Hal ini sejalan dengan hasil penelitian ini, di mana *regresi linear* mampu memprediksi penjualan dengan nilai *Root Mean Square Error* (RMSE) sebesar 493125.701 dan kesalahan absolut rata-rata sebesar 303270.965. Hasil yang diperoleh juga sejalan dengan penelitian oleh Erwin *et al.* (2023), yang mengaplikasikan metode *regresi linear berganda* dalam manajemen penjualan pada MPM MOTOR Dealer di Lombok Timur. Mereka menemukan bahwa *regresi linear* mampu memprediksi variabel penjualan dengan tingkat akurasi yang tinggi, serupa dengan penelitian ini yang menggunakan variabel seperti TM-168, TM-178, dan UM-100 sebagai prediktor utama penjualan.

Selain itu, metode *regresi linear* juga diterapkan oleh Feblian dan Daihani (2017) dalam model CRISP-DM untuk memprediksi *sales pipeline*, dan hasilnya menunjukkan bahwa *regresi linear* adalah alat yang efisien dalam mengidentifikasi pola penjualan historis yang berguna bagi perencanaan bisnis. Dalam penelitian ini, pendekatan CRISP-DM juga digunakan, yang membantu dalam menyusun langkah-langkah sistematis mulai dari pemahaman bisnis hingga evaluasi model. Penerapan *regresi linear* dalam berbagai bidang, seperti yang dijelaskan oleh Syakir *et al.* (2022) untuk prediksi penjualan di *marketplace* Shopee, dan oleh Rajab Mudatsir dan Melangi (2022) untuk prediksi jumlah produksi ikan asin, semakin memperkuat relevansi metode ini. Prediksi yang akurat dalam penelitian ini menunjukkan bahwa *regresi linear* memberikan hasil yang serupa dalam berbagai penerapan, terutama dalam memprediksi penjualan dan produksi. Namun, penelitian ini juga mengidentifikasi bahwa meskipun metode *regresi linear* memberikan akurasi yang baik, ada ruang untuk perbaikan. Penelitian oleh Rizkiawan *et al.* (2023) dan Rahayu dan Gunaryati (2023) menyarankan bahwa penambahan variabel atau penggunaan metode lain, seperti *regresi linear berganda* atau algoritma *machine learning* lainnya, dapat meningkatkan akurasi prediksi. Oleh karena itu, penelitian lanjutan dengan mempertimbangkan faktor-faktor tambahan atau metode prediksi yang lebih kompleks dapat memberikan hasil yang lebih optimal. Secara keseluruhan, hasil penelitian ini memperkuat temuan dari berbagai studi sebelumnya bahwa *regresi linear* adalah metode yang efektif untuk memprediksi penjualan, terutama ketika digunakan dengan data historis yang relevan. Hasil prediksi yang akurat ini dapat digunakan oleh manajemen PT. Toyo Mortar Indonesia untuk merancang strategi penjualan yang lebih efisien dan berbasis data.

4. Kesimpulan

Penelitian ini menggunakan metode *Regresi linear* untuk memprediksi jumlah penjualan produk semen di PT. Toyo Mortar Indonesia, yang dirancang dalam tahapan-tahapan metode CRISP-DM (*Cross Industry Standard for Data Mining*) sebagai model pengujian data mining. Dalam pengujian yang dilakukan, data diolah menggunakan 12 variabel. Dari 12 variabel tersebut, hanya 10 yang mempengaruhi hasil prediksi, yaitu: TM-168, TM-178, TM-188, TM-189, TM-198, UM-100, UM-101, UM BIRU, bahan baku, dan penjualan lainnya. Hasil pengujian menggunakan metode *Regresi linear* dengan alat Rapidminer menunjukkan bahwa nilai *Root Mean Squared Error* (RMSE) adalah 493.125,701, *Absolute Error* adalah 303.270,965, dan *Relative Error* adalah 1,89%. Ini menunjukkan performa yang bagus dan hasil prediksi yang cukup akurat. Dari penelitian ini, ada beberapa saran yang perlu dipertimbangkan sebagai bentuk dari pengembangan penelitian yang penulis lakukan diantaranya yaitu menambahkan atribut yang lebih spesifik dan meningkatkan jumlah data untuk prediksi penjualan semen agar prediksi semakin akurat. Semakin banyak atribut dan data yang digunakan, semakin baik pula hasil perhitungan dan penelitian ini dapat dikembangkan dengan menggabungkan metode lain sehingga menghasilkan prediksi yang lebih variatif.

5. Ucapan Terima Kasih

Peneliti mengucapkan terima kasih yang sebesar-besarnya atas bimbingan, dukungan, dan dedikasi Bapak Kiki Setiawan dalam membantu pembuatan jurnal ini. Tanpa panduan dan dorongan yang berharga dari Anda, pencapaian ini tidak akan mungkin terwujud. Saya sangat berterima kasih atas waktu dan pengetahuan yang Anda luangkan untuk membimbing saya dalam menyusun jurnal ini. Ucapan terima kasih juga saya sampaikan kepada Ibu Nadia Mutiara dari PT Toyo Mortar Indonesia, yang telah menjadi narasumber dalam penelitian ini. Terima kasih atas waktu dan informasi berharga yang telah diberikan. Dukungan dan kontribusi Anda sangat membantu dalam penyusunan jurnal ini.

6. Daftar Pustaka

- Ayuni, G. N., & Fitriana, D. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal telematika*, 14(2), 79-86. DOI: <https://doi.org/10.61769/telematika.v14i2.321>.
- Erwin, L., Kartarina, K., Anggriani, R., Febriana, W., & Sriwinarti, N. K. (2023). Analisis Metode Regresi Linier Berganda Dan Ordinary Least Squared Dalam Mengelola Manajemen Operasional Penjualan MPM MOTOR Dealer di Lombok Timur. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 7(1), 137-150. DOI: <https://doi.org/10.52362/jisamar.v7i1.1016>.
- Feblian, D., & Daihani, D. U. (2016). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X. *Jurnal Teknik Industri*, 6(1).
- Hartono, P. C., & Widianoro, A. D. (2024). Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier dengan RapidMiner. *Journal of Computer and Information Systems Ampera*, 5(3), 174-190. DOI: 10.51519/journalcisa.v5i3.481.

- Mudatsir, M. R., & Serwin, S. (2022). Prediksi Jumlah Produksi Ikan Asin Menggunakan Metode Regresi Linear Sederhana. *Jurnal Ilmiah Ilmu Komputer Bantayo Lo Komputer*, 1(2), 118-124.
- Rahayu, D., & Gunaryati, A. (2023). Memprediksi Kualitas Produk Inspeksi Dalam Meminimalisasi Resiko Produk Ng Meggunakan Algoritma Regresi Linier. *Bulletin of Information Technology (BIT)*, 4(3), 384-390. DOI: <https://doi.org/10.47065/bit.v4i3.894>.
- Revaldi, K. A., Cahyono, F. D., Hakimah, M., & Muhima, R. R. (2023, April). Penerapan Metode Regresi Linier Sederhana dalam Memprediksi Jumlah Kebutuhan Ekspor Migas dan Non-Migas di Indonesia. In *Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika (SNESTIK)* (Vol. 1, No. 1, pp. 268-275).
- Riskya, N., & Yuliana, S. (2023). Penerapan Data Mining Untuk Prediksi Perilaku Pelanggan Menggunakan Multiple Linear Regression. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(3). DOI: <http://dx.doi.org/10.23960/jitet.v11i3.3194>.
- Rizkiawan, A., & Wahyudi, T. (2023). Implementasi Data Mining Untuk Memprediksi Member Baru Menggunakan Linear Regression Pada Pt. Gsi. *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, 6(1), 118-126. DOI: <https://doi.org/10.37600/tekinkom.v6i1.707>.
- Rusmilawati, N., & Prasetyaningrum, P. T. (2021). Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunakan Metode Linier Regression. *Journal Of Information System And Artificial Intelligence*, 1(2), 78-84.
- Setiawan, D., Surojudin, N., & Hadikristanto, W. (2022). Prediksi Penjualan Obat Dengan Algoritma Regresi Linear. *Prosiding Sains dan Teknologi*, 1(1), 237-246.
- Syakir, Y., Hermanto, T. I., & Ramadhan, Y. R. (2022). Analisis Marketplace Shopee Untuk Memprediksi Penjualan dengan Algoritma Regresi Linier. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 6(2), 904-915. DOI: <http://dx.doi.org/10.30645/j-sakti.v6i2.501>.