

ALGORITMA REGRESI LINIER UNTUK MENINGKATKAN MODEL PREDIKSI PENJUALAN PADA TOKO DEVANJAYABAN

Hardika¹, Martanto², Arif Rinaldi Dikananda³, Mulyawan⁴

¹Teknik Informatika, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135

²Manajemen Informatika, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135

³Rekayasa Perangkat Lunak, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135

⁴Program Studi Sistem Infomasi, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135

Received: 1 Maret 2025
Accepted: 27 Maret 2025
Published: 14 April 2025

Keywords:

prediksi penjualan, regresi linear, KDD, akurasi model, industri otomotif.

Correspondent Email:

hard44840@gmail.com

Abstrak. Penjualan ban mobil menghadapi tantangan akibat volatilitas pasar dan pola permintaan yang kompleks, sehingga diperlukan model prediksi yang andal. Penelitian ini mengembangkan model prediksi penjualan di Toko Devan Jaya Ban menggunakan regresi linear. Data historis penjualan bulanan (Januari–April 2024) dianalisis dengan metode Knowledge Discovery in Databases (KDD), mencakup seleksi data, preprocessing, transformasi, pemodelan, dan evaluasi. Model dibangun menggunakan RapidMiner dan dievaluasi dengan Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Relative Error (RE). Hasil menunjukkan performa baik dengan RMSE 1.778, MAE 1.478 ± 0.989 , dan RE $6.81\% \pm 5.09\%$. Preprocessing, seperti normalisasi data dan pemilihan variabel relevan, meningkatkan akurasi model. Regresi linear terbukti efektif dalam memprediksi penjualan serta mendukung optimalisasi stok, perencanaan pemasaran, dan pengambilan keputusan bisnis. Pengembangan lebih lanjut dapat mencakup variabel eksternal seperti tren pasar, musim, dan faktor ekonomi, serta membandingkan regresi linear dengan algoritma pembelajaran mesin lain untuk model yang lebih adaptif.

Abstract. The automotive tire sales sector faces challenges due to market volatility and complex demand patterns, requiring a reliable predictive model. This study develops a sales prediction model for Devan Jaya Ban Store using linear regression. Monthly sales data from January to April 2024 were analyzed using the Knowledge Discovery in Databases (KDD) method, including data selection, preprocessing, transformation, modeling, and evaluation. The model was built using RapidMiner and evaluated with Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Relative Error (RE). The results indicate good performance, with an RMSE of 1.778, MAE of 1.478 ± 0.989 , and RE of $6.81\% \pm 5.09\%$. Preprocessing, such as data normalization and relevant variable selection, significantly improved model accuracy. Linear regression proved effective in predicting sales, optimizing inventory management, marketing planning, and business decision-making. Further development may integrate external variables such as market trends, seasonality, and economic factors, as well as compare linear regression with other machine learning algorithms for a more adaptive model.

1. PENDAHULUAN

Perkembangan teknologi informasi telah memfasilitasi analisis data dalam berbagai sektor, termasuk industri otomotif, yang memanfaatkan data untuk meningkatkan efisiensi operasional dan pengambilan keputusan strategis. Salah satu aplikasi yang signifikan adalah prediksi penjualan, di mana model matematis seperti algoritma regresi linear digunakan untuk memproyeksikan tren berdasarkan data historis dan faktor-faktor relevan lainnya. [1] menunjukkan bahwa algoritma regresi linear berganda dapat diterapkan secara efektif untuk memperkirakan penjualan mobil di Astra Isuzu dengan akurasi yang memadai, khususnya ketika model ini memperhitungkan variabel internal seperti promosi dan harga, serta variabel eksternal seperti kondisi ekonomi. Temuan ini menggarisbawahi pentingnya pemanfaatan model prediktif yang memadukan variabel-variabel kontekstual dalam mendukung perencanaan bisnis. Dalam konteks industri ban mobil, pendekatan serupa dapat memberikan manfaat yang signifikan, terutama dalam mengatasi tantangan volatilitas pasar dan pergeseran preferensi konsumen. Oleh karena itu, penelitian ini berfokus pada pengembangan model prediksi penjualan yang relevan untuk sektor ini, dengan memanfaatkan algoritma regresi linear yang telah terbukti efektif [2].

Salah satu tantangan utama dalam industri ban mobil adalah bagaimana memanfaatkan data historis untuk menghasilkan prediksi penjualan yang akurat, terutama dalam menghadapi kompleksitas pola penjualan yang dipengaruhi oleh berbagai faktor, seperti musim, tren pasar, dan perubahan preferensi konsumen. Data historis yang meliputi pola penjualan, fluktuasi harga, dan faktor eksternal lainnya sering kali belum dimanfaatkan secara maksimal karena kurangnya pendekatan analitis yang tepat. [3] menunjukkan bahwa metode regresi linear sederhana dapat digunakan untuk memprediksi harga beras di Kota Padang dengan mengandalkan data historis yang terstruktur. Hal ini menunjukkan bahwa teknik serupa memiliki potensi untuk diterapkan dalam konteks industri ban mobil, asalkan variabel yang relevan turut diperhitungkan [4]. Namun, berbeda dengan komoditas seperti beras, pola penjualan ban mobil memiliki tingkat kompleksitas yang lebih tinggi karena melibatkan variabel tambahan, seperti kondisi geografis dan pergeseran permintaan musiman. Selain itu, sebagian besar pelaku usaha kecil dan menengah di sektor ini belum memiliki infrastruktur analisis data yang memadai, sehingga pengambilan keputusan sering kali didasarkan pada intuisi atau pengalaman semata. Akibatnya, sering terjadi inefisiensi seperti overstocking atau kekurangan stok yang dapat berdampak pada biaya operasional dan kepuasan pelanggan. Penelitian ini bertujuan untuk mengatasi

tantangan tersebut dengan mengembangkan model prediksi berbasis algoritma regresi linear yang mampu mengolah data historis secara efektif. Dengan pendekatan ini, diharapkan akurasi prediksi penjualan dapat ditingkatkan, sehingga mendukung pengambilan keputusan strategis yang lebih berbasis data. Temuan dari penelitian ini berpotensi memberikan solusi praktis bagi pelaku usaha dalam menghadapi dinamika pasar yang terus berkembang. Penelitian terdahulu terkait penggunaan algoritma regresi linear untuk prediksi penjualan telah banyak dilakukan dengan hasil yang signifikan. [5] menunjukkan bahwa regresi linear efektif dalam memprediksi penjualan produk di sektor FMCG, namun tantangan muncul dalam mengintegrasikan data yang tidak terstruktur. [3] menerapkan regresi linear sederhana untuk memprediksi harga beras di Kota Padang, yang membuktikan keberhasilan model sederhana dalam memanfaatkan data historis, meskipun pola komoditas tersebut relatif stabil. Selain itu, [6] menyoroti bahwa akurasi prediksi dapat ditingkatkan dengan memperhitungkan hubungan antarvariabel dalam sektor ritel. Namun, penelitian sebelumnya belum banyak mengeksplorasi prediksi penjualan di industri ban mobil, yang memiliki pola dinamis akibat faktor eksternal seperti musim dan perubahan preferensi konsumen. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengembangkan model prediksi berbasis algoritma regresi linear yang lebih terintegrasi dan relevan dengan kompleksitas pasar ban mobil.

Tujuan utama dari penelitian ini adalah mengembangkan model prediksi penjualan yang dapat menghasilkan estimasi yang lebih akurat dan andal untuk mendukung pengambilan keputusan strategis dalam industri ban mobil. Model yang diusulkan diharapkan mampu mengintegrasikan data historis secara optimal dan memberikan hasil prediksi yang relevan dengan kebutuhan pasar yang dinamis. Penelitian ini juga bertujuan untuk mengatasi keterbatasan pendekatan sebelumnya yang kurang mampu menangkap kompleksitas pola penjualan di industri ini. Dengan pendekatan yang lebih terstruktur, model prediksi yang dihasilkan diharapkan tidak hanya bermanfaat bagi peningkatan efisiensi operasional, tetapi juga mampu mendukung pengembangan strategi bisnis jangka panjang. Secara keseluruhan, penelitian ini berkontribusi pada peningkatan kemampuan analitis dalam memanfaatkan data historis penjualan sebagai dasar pengambilan keputusan di sektor otomotif.

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan algoritma regresi linear sebagai metode utama untuk pengembangan model prediksi penjualan. Pengolahan data dilakukan menggunakan perangkat lunak RapidMiner, yang memungkinkan proses analisis

data menjadi lebih terstruktur dan efisien. Data historis penjualan diolah melalui serangkaian tahapan, termasuk preprocessing, pemodelan, dan evaluasi kinerja model. RapidMiner dipilih karena kemampuannya dalam menangani data yang kompleks dan mendukung visualisasi hasil analisis secara intuitif. Proses evaluasi model dilakukan untuk memastikan akurasi dan reliabilitas prediksi, sehingga model yang dihasilkan dapat digunakan sebagai dasar pengambilan keputusan strategis di sektor ban mobil.

Implikasi dari penelitian ini sangat signifikan, baik dari sisi akademis maupun praktis. Secara akademis, penelitian ini dapat memperkaya literatur tentang penerapan algoritma regresi linear dalam prediksi penjualan, khususnya di sektor otomotif. Hasil dari penelitian ini diharapkan memberikan kontribusi pada pengembangan model prediksi yang lebih akurat dengan menggunakan data historis yang lebih terstruktur. Dari sisi praktis, model yang dikembangkan dapat membantu perusahaan ban mobil untuk meningkatkan efisiensi operasional, seperti dalam perencanaan stok dan pengelolaan persediaan, sehingga dapat mengurangi risiko kekurangan atau kelebihan stok. Selain itu, dengan akurasi prediksi yang lebih baik, perusahaan dapat merencanakan strategi pemasaran dan penjualan yang lebih tepat sasaran, mengoptimalkan profitabilitas. Penelitian ini juga dapat menjadi referensi bagi pelaku usaha kecil dan menengah yang belum memiliki infrastruktur analisis data yang memadai, membantu mereka dalam membuat keputusan berbasis data. Implikasi jangka panjang dari penelitian ini adalah peningkatan daya saing industri ban mobil melalui penerapan teknologi analisis data yang lebih maju dan terintegrasi.

2. TINJAUAN PUSTAKA

Prediksi penjualan merupakan topik yang banyak dibahas dalam berbagai penelitian. Penelitian [7] menyoroti penerapan regresi linear untuk memprediksi penjualan properti di PT XYZ selama periode Juni 2014 hingga Mei 2019. Dalam penelitian ini, metode regresi linear digunakan melalui tahapan pengumpulan data, preprocessing, perhitungan nilai variabel, dan evaluasi akurasi menggunakan MSE, RMSE, serta MAPE. Hasil menunjukkan persamaan regresi untuk beberapa tipe properti, seperti kavling ($Y = 26,5 + 0,9X$) dan ruko ($Y = 14,4 - 0,4X$), dapat memberikan prediksi akurat untuk beberapa tipe properti, meskipun beberapa tipe diprediksi tidak terjual.

[8] melakukan analisis pengaruh kualitas pelayanan dan kualitas produk terhadap kepuasan konsumen di Restoran Cwie Mie Malang menggunakan regresi linear berganda. Penelitian ini menekankan pentingnya uji asumsi klasik sebelum analisis regresi, termasuk uji normalitas,

multikolinearitas, autokorelasi, dan heteroskedastisitas. Hasil penelitian menunjukkan kedua variabel bebas berpengaruh signifikan terhadap kepuasan konsumen, dengan model regresi yang signifikan secara keseluruhan.

Penelitian lain oleh [9] mengembangkan sistem prediksi penjualan smartphone Lenovo menggunakan regresi linear. Data historis dari tahun 2014 hingga 2016 digunakan untuk membangun model, yang diimplementasikan dalam Visual Basic.Net dengan SQL Server 2008. Hasil evaluasi menunjukkan bahwa model mampu memberikan prediksi yang baik, dengan nilai MAPE 0,032 dan MSE 5,16, meskipun disarankan untuk menambahkan metode lain seperti Weighted Moving Averages untuk meningkatkan performa.

[10] menggunakan regresi linear berganda untuk memprediksi peningkatan omset penjualan di PT. Makmur Jaya. Penelitian ini fokus pada pengurangan kesalahan manual dengan mengotomatisasi proses prediksi menggunakan data historis. Hasil menunjukkan bahwa metode ini efektif untuk meningkatkan akurasi prediksi omset penjualan dan memberikan rekomendasi strategi bisnis berdasarkan hasil prediksi.

Penelitian [11] mengeksplorasi penerapan regresi linear berganda untuk memprediksi harga sembako di DKI Jakarta. Penelitian ini mencakup pengolahan dan transformasi data, pembangunan model, serta pengujian asumsi klasik seperti normalitas residual, homoskedastisitas, dan multikolinearitas. Hasil penelitian menunjukkan bahwa 84,2% variasi harga dapat dijelaskan oleh variabel bebas, menunjukkan model memiliki tingkat akurasi yang tinggi.

[12] menerapkan regresi linear berganda untuk memprediksi harga saham di sektor pelayaran dengan studi kasus PT Buana Lintas Lautan, Tbk. Penelitian ini menggunakan metodologi CRISP-DM, yang mencakup enam tahapan, mulai dari pemahaman bisnis hingga evaluasi dan deployment. Hasil menunjukkan model dapat memprediksi harga saham dengan akurasi cukup baik, meskipun masih terdapat selisih kecil antara hasil prediksi dan data aktual.

Dalam penelitian ini membahas konsep Data Mining, terutama dalam penerapan Regresi Linier Berganda untuk mengidentifikasi hubungan antara variabel independen dan variabel dependen dalam memprediksi penjualan. Pendekatan ini sering digunakan dalam analisis bisnis guna meningkatkan efisiensi pengelolaan stok dan menghindari kelebihan atau kekurangan persediaan. Dalam penelitian ini, RapidMiner digunakan sebagai alat bantu dalam pengolahan data dan penerapan algoritma prediktif. Beberapa studi terdahulu menunjukkan bahwa metode regresi linier berganda telah berhasil diterapkan dalam berbagai bidang,

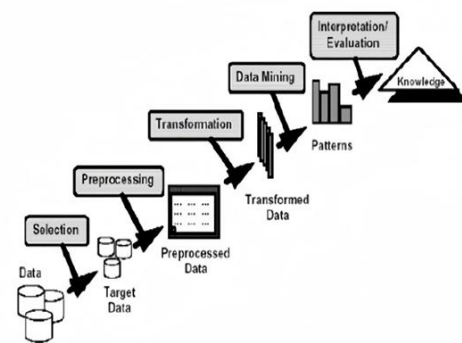
seperti peramalan permintaan barang dan analisis pola penjualan. Berdasarkan hal tersebut, penelitian ini menerapkan metode regresi linier untuk memprediksi stok produk madu di PT. Madu Pramuka Batang guna meningkatkan efektivitas perencanaan persediaan[13].

Dari tinjauan pustaka tersebut, gap analysis menunjukkan bahwa sebagian besar penelitian terdahulu berfokus pada penerapan regresi linear untuk sektor-sektor tertentu, seperti properti, restoran, smartphone, sembako, dan saham, namun belum ada yang secara spesifik mengembangkan model prediksi penjualan untuk sektor otomotif, khususnya toko penjualan ban mobil. Penelitian ini menawarkan kebaharuan dengan mengaplikasikan regresi linear dalam memprediksi penjualan di sektor otomotif, menggunakan data historis penjualan toko kecil, serta melibatkan proses preprocessing data yang mendalam. Selain itu, penelitian ini juga memberikan evaluasi performa model menggunakan metrik seperti RMSE, MAE, dan RE untuk memastikan tingkat akurasi yang optimal, sekaligus menyarankan pengembangan model yang lebih kompleks untuk meningkatkan adaptabilitas terhadap perubahan pasar.

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini menggunakan algoritma regresi linear sebagai metode utama untuk membangun model prediksi penjualan, didukung oleh penerapan metodologi Knowledge Discovery in Databases (KDD). KDD dipilih sebagai kerangka kerja karena kemampuannya dalam mengelola dan mengekstraksi informasi yang bermakna dari data yang kompleks dan besar. Tahapan KDD yang diterapkan meliputi seleksi data, preprocessing, transformasi, pemodelan, dan evaluasi. Pada tahap pemodelan, algoritma regresi linear digunakan untuk menganalisis hubungan antara data historis dengan hasil penjualan, sehingga menghasilkan model yang mampu memprediksi nilai penjualan secara akurat. Metode ini memungkinkan identifikasi pola tersembunyi dalam data dan mendukung pengambilan keputusan berbasis fakta. Evaluasi model dilakukan untuk memastikan kinerja dan akurasi prediksi dengan mengacu pada metrik statistik yang relevan, seperti nilai Mean Absolute Error (MAE) atau Root Mean Square Error (RMSE). Kombinasi antara algoritma regresi linear dan pendekatan KDD diharapkan memberikan hasil yang andal dalam membantu industri ban mobil memanfaatkan data sebagai aset strategis.



Gambar 1. 1 KDD

3.2. Pengumpulan Data

Teknik pengumpulan data ini adalah metode pengumpulan data sekunder, data yang dikumpulkan dari pihak Toko Devan Jaya Ban merupakan data yang telah tersedia sebelumnya (data historis penjualan). Data ini tidak diperoleh langsung dari observasi atau wawancara, tetapi dari catatan atau arsip yang telah disusun sebelumnya oleh pihak toko. Pengumpulan data dalam penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk memastikan data yang diperoleh relevan dan akurat. Tahapan-tahapan tersebut meliputi:

1. Memilih Tempat Penelitian, Tahap pertama adalah menentukan lokasi penelitian, yang dalam hal ini adalah Toko Devan Jaya Ban (DJB), karena toko ini memiliki data historis penjualan yang relevan untuk penelitian terkait prediksi penjualan ban mobil.
2. Meminta Izin Melakukan Penelitian, Peneliti menghubungi pihak manajemen Toko Devan Jaya Ban untuk meminta izin melakukan penelitian di toko tersebut. Langkah ini penting untuk mendapatkan persetujuan resmi dari pihak terkait agar penelitian dapat dilaksanakan tanpa hambatan.
3. Mengirim Surat Permohonan dari Kampus, Setelah mendapatkan persetujuan awal, peneliti mengirimkan surat permohonan resmi dari kampus kepada pihak toko sebagai bentuk formalitas dan legitimasi penelitian yang akan dilakukan.
4. Meminta Data Penjualan yang Dibutuhkan, Tahap terakhir adalah pengumpulan data penjualan dari pihak toko. Data yang diminta meliputi total penjualan bulanan, jenis ban yang terjual, dan harga ban. Data ini menjadi sumber utama dalam proses analisis untuk mengembangkan model prediksi penjualan yang berbasis regresi linier.

3.3. Analisa Data

Teknik analisis data dalam penelitian ini menggunakan metode regresi linear dengan bantuan tool RapidMiner untuk membangun model prediksi.

Proses analisis dimulai dengan pengolahan data untuk memastikan kualitas data yang akan digunakan, diikuti oleh pembagian dataset menjadi data pelatihan dan pengujian. Model regresi linear kemudian dikembangkan berdasarkan data pelatihan. Evaluasi kinerja model dilakukan dengan menggunakan tiga metrik utama, yaitu Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Mean Absolute Error (MAE). MSE digunakan untuk mengukur rata-rata kuadrat kesalahan prediksi, RMSE memberikan interpretasi dalam satuan yang sama dengan data asli, dan MAE mencerminkan rata-rata kesalahan absolut. Ketiga metrik ini memberikan gambaran komprehensif mengenai tingkat akurasi dan keandalan model prediksi yang dihasilkan. RapidMiner digunakan untuk mengotomatisasi proses analisis dan memvisualisasikan hasil evaluasi model.

Regresi linear adalah teknik statistik yang digunakan untuk menganalisis hubungan antara dua variabel, yaitu variabel independen (X) dan variabel dependen (Y). Metode ini bertujuan untuk memprediksi nilai Y berdasarkan nilai X serta memahami pengaruh X terhadap Y. Persamaan dasar dari regresi linear sederhana adalah:

$$Y = a + bX$$

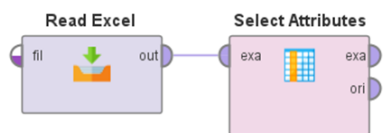
Dengan keterangan:

- a: konstanta, yaitu titik potong pada sumbu Y
- b: koefisien variabel X yang menunjukkan seberapa besar pengaruh X terhadap Y
- Y: variabel dependen
- X: variabel independen

3.4. Tahapan Regresi Linier

Berikut adalah seluruh tahapan algoritma regresi linier menggunakan tools rapid miner dengan metode penelitian Knowledge Discovery in Databases (KDD):

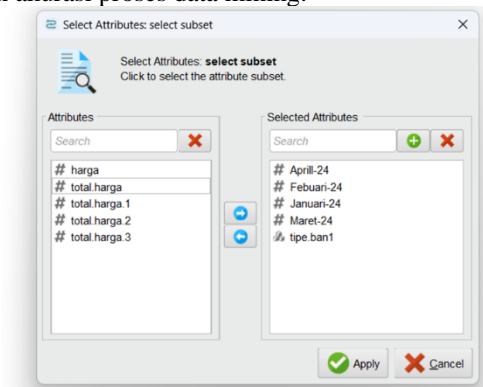
1. Data Selection



Gambar 1. 2 Tahapan Data Selection

Pada proses yang terlihat dalam gambar 1.2, operator Read Excel berfungsi sebagai langkah awal untuk membaca data dari file Excel yang menjadi sumber utama dataset. Operator ini mengimpor data mentah dari file dengan format seperti .xls atau .xlsx, di mana pengguna dapat menentukan pengaturan tambahan seperti memilih lembar kerja (sheet) tertentu, rentang sel yang relevan, atau memastikan

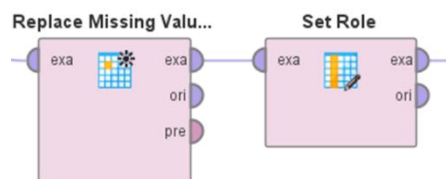
bahwa baris pertama berisi nama atribut (header). Setelah data berhasil dimuat, operator Select Attributes digunakan untuk menyaring dan memilih atribut (kolom) yang relevan sesuai kebutuhan analisis. Proses seleksi ini memungkinkan pengguna untuk mempertahankan atribut yang penting dan menghilangkan atribut yang tidak diperlukan. Seleksi atribut dapat dilakukan secara manual, otomatis berdasarkan tipe data, atau melalui aturan tertentu. Kombinasi kedua operator ini bertujuan untuk memastikan bahwa hanya data yang relevan dan berkualitas tinggi yang diteruskan ke tahapan analisis berikutnya, sehingga meningkatkan efisiensi dan akurasi proses data mining.



Gambar 1. 3 Tahapan Select Data

Gambar 3.3 menunjukkan proses seleksi data menggunakan operator Select Attributes. Dalam proses ini, memilih atribut yang relevan dari dataset menggunakan dua panel yang disediakan. Panel kiri menampilkan semua atribut yang tersedia dalam dataset, sementara panel kanan menunjukkan atribut yang dipilih untuk dianalisis lebih lanjut. Pada gambar ini, atribut yang dipilih meliputi April-24, Febuari-24, Januari-24, Maret-24, dan tipe.ban1, sedangkan atribut seperti harga dan total.harga tidak disertakan. Pengguna dapat memindahkan atribut dari panel kiri ke panel kanan menggunakan tombol panah dan menerapkan seleksi dengan mengklik tombol Apply. Proses ini penting untuk memfokuskan analisis hanya pada kolom-kolom yang relevan, mengurangi kompleksitas data, dan meningkatkan efisiensi dalam pemodelan atau transformasi data berikutnya.

2. Data Preprocessing

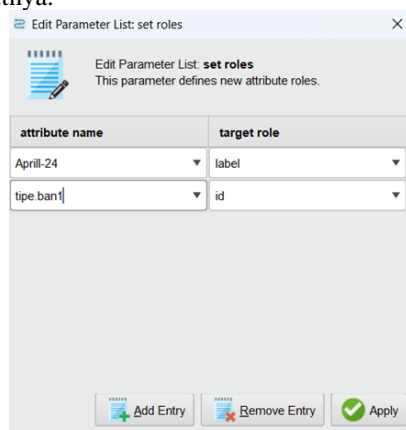


Gambar 1. 4 Data Preprocessing

Pada gambar 1.4 adalah Tahapan Replace Missing Values dan Set Role merupakan bagian dari proses pra-pemrosesan data yang krusial untuk

memastikan kualitas data sebelum digunakan dalam analisis atau pembuatan model. Replace Missing Values bertujuan untuk menangani nilai-nilai yang hilang (missing values) dalam dataset, yang jika tidak diatasi, dapat mengganggu hasil analisis atau menurunkan performa model. Operator ini menggantikan nilai yang hilang menggunakan metode tertentu, seperti rata-rata (mean), median, modus, atau nilai tetap, sesuai dengan jenis atribut. Dengan menyelesaikan masalah nilai yang hilang, dataset menjadi lebih lengkap dan siap untuk digunakan.

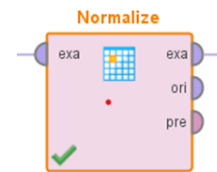
Setelah nilai hilang ditangani, tahapan Set Role dilakukan untuk menentukan peran (role) setiap atribut dalam dataset. Operator ini memungkinkan pengguna untuk menetapkan atribut sebagai target (label), atribut prediktor, atau atribut tambahan (seperti ID). Penentuan peran atribut sangat penting untuk memastikan bahwa algoritma data mining memahami fungsi setiap atribut dalam proses analisis. Kombinasi kedua tahapan ini memastikan bahwa dataset tidak hanya bersih tetapi juga terstruktur dengan baik, sehingga memaksimalkan efisiensi dan akurasi pada tahapan analisis berikutnya.



Gambar 1.5 Tahapan Set Role

Dalam gambar 1.5 melihat tahapan set role, dua atribut diatur dengan peran tertentu: atribut April-24 ditetapkan sebagai label, yang menunjukkan bahwa atribut ini akan menjadi target prediksi dalam proses machine learning, sedangkan atribut tipe banl ditetapkan sebagai id, yang biasanya digunakan sebagai pengidentifikasi unik untuk setiap entri data. Dengan menetapkan peran ini, RapidMiner dapat memahami fungsi spesifik dari setiap atribut dalam alur kerja, yang memudahkan proses pelatihan model atau evaluasi data. Tahap ini memastikan bahwa data siap untuk digunakan dalam analisis yang lebih kompleks dan meningkatkan akurasi hasil prediksi,

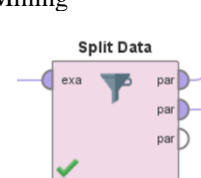
3. Data Transformation



Gambar 1.6 Tahapan Normalize

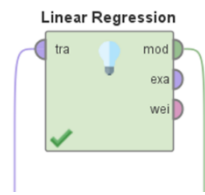
Dalam penelitian ini, tahap Normalize digunakan sebagai bagian dari proses Data Transformation untuk memastikan bahwa atribut numerik dalam dataset berada pada skala yang seragam. Normalisasi mengubah nilai atribut ke rentang tertentu, seperti [0, 1], sehingga mengurangi dampak perbedaan skala antaratribut yang dapat memengaruhi analisis dan hasil model. Langkah ini sangat penting karena model yang digunakan, seperti Linear Regression, sensitif terhadap perbedaan skala data, di mana atribut dengan nilai lebih besar dapat mendominasi proses pembelajaran. Dengan menerapkan normalisasi, semua atribut numerik diberikan bobot yang seimbang, sehingga hubungan antarvariabel dapat dianalisis secara lebih obyektif. Tahap ini dilakukan setelah proses pra-pemrosesan data, seperti menangani nilai hilang, dan sebelum pembagian dataset ke dalam data pelatihan dan pengujian, untuk memastikan bahwa transformasi diterapkan secara konsisten. Dengan adanya normalisasi, kualitas dataset meningkat, dan kinerja model menjadi lebih optimal dalam memberikan hasil analisis yang akurat.

4. Data Mining



Gambar 1.7 Tahapan Split Data

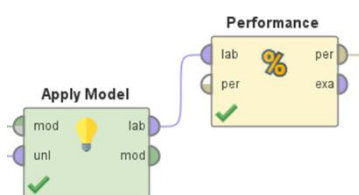
Gambar 3,7 adalah tahapan split data, proses pemisahan dataset dilakukan menggunakan operator Split Data pada RapidMiner dengan rasio 0.7:0.3. Rasio ini berarti bahwa 70% data akan digunakan sebagai data latih (training data) untuk membangun dan melatih model regresi linear, sedangkan 30% sisanya digunakan sebagai data uji (testing data) untuk mengevaluasi kinerja model yang dihasilkan. Operator Split Data pada RapidMiner memungkinkan pembagian dataset secara acak namun terkontrol, sehingga distribusi data tetap representatif. Proses ini penting untuk memastikan model yang dibangun mampu menangkap pola dari data latih dan menggeneralisasi dengan baik terhadap data uji, yang nantinya akan mencerminkan kinerja model di dunia nyata.



Gambar 1. 8 Tahapan Linear Regression

Gambar 1.8 adalah tahapan Linear Regression dilakukan dengan menggunakan operator Linear Regression di RapidMiner. Operator ini bertugas untuk membangun model prediksi yang mempelajari hubungan linear antara atribut-atribut independen (prediktor) dan atribut dependen (target). Setelah data melalui tahapan pra-pemrosesan seperti seleksi atribut, penanganan nilai hilang, dan normalisasi, operator Linear Regression diterapkan pada data pelatihan. Operator ini menghitung koefisien regresi untuk setiap atribut prediktor dengan meminimalkan mean squared error (MSE), sehingga model dapat memprediksi nilai target secara akurat. Proses ini menghasilkan model matematis yang merepresentasikan hubungan antara variabel-variabel dalam dataset. Hasil model kemudian diterapkan pada data pengujian untuk mengukur kinerja dan validitasnya melalui operator evaluasi seperti Performance. Dengan menggunakan operator ini, tahapan linear regression menjadi terstruktur, efisien, dan mudah diintegrasikan dalam alur analisis yang lebih besar. Tahap ini membantu penelitian dalam menghasilkan model yang dapat digunakan untuk memahami pola dan memberikan prediksi berdasarkan data yang tersedia.

5. Interpretation/Evaluation



Gambar 1. 9 Tahapan Test Performance

tahapan Apply Model dan Performance digunakan untuk menguji dan mengevaluasi model yang telah dibangun. Operator Apply Model bertugas untuk menerapkan model hasil pelatihan (dalam hal ini, model dari operator Linear Regression) pada data pengujian. Proses ini memungkinkan prediksi dilakukan berdasarkan pola yang telah dipelajari dari data pelatihan. Data pengujian yang digunakan dalam tahapan ini harus melalui proses pra-pemrosesan yang sama seperti data pelatihan untuk memastikan konsistensi hasil. Setelah model diaplikasikan, hasil prediksi dibandingkan dengan nilai aktual pada data pengujian.

Operator Performance digunakan untuk mengevaluasi kualitas dan keakuratan model. Operator ini menghitung metrik evaluasi seperti Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), atau koefisien determinasi (R^2), yang memberikan gambaran tentang seberapa baik model dapat memprediksi data. Hasil evaluasi ini membantu menentukan apakah model sudah sesuai dengan tujuan analisis atau perlu dilakukan penyesuaian lebih lanjut. Dengan kombinasi kedua operator ini, penelitian dapat memastikan bahwa model yang dihasilkan tidak hanya sesuai secara teoritis, tetapi juga mampu memberikan hasil yang andal dalam konteks data nyata.

4. HASIL DAN PEMBAHASAN

Hasil analisis regresi linear pada gambar 1.10 menunjukkan bahwa kedua variabel prediktor, yaitu Januari-24 dan Februari-24, memiliki pengaruh signifikan terhadap variabel target. Koefisien regresi untuk variabel Januari-24 adalah sebesar 4.249, yang menunjukkan bahwa setiap peningkatan satu unit pada Januari-24 akan meningkatkan nilai target sebesar 4.249 unit, dengan nilai p-value 0.000 yang signifikan pada tingkat kepercayaan 95%. Untuk variabel Februari-24, koefisiennya sebesar 2.979, yang berarti peningkatan satu unit pada Februari-24 akan meningkatkan nilai target sebesar 2.979 unit, dengan p-value 0.004, juga signifikan. Nilai intercept adalah 21.169, yang menunjukkan prediksi awal ketika semua variabel independen bernilai nol. Nilai t-statistic menunjukkan tingkat signifikansi masing-masing koefisien, di mana kedua variabel prediktor memiliki nilai t yang tinggi (masing-masing 4.516 untuk Januari-24 dan 3.145 untuk Februari-24), menguatkan validitas pengaruhnya terhadap model. Secara keseluruhan, model regresi ini dapat digunakan untuk memprediksi variabel target dengan baik, mengingat tingkat signifikansi variabelnya yang memenuhi syarat.

Varibel	Coefficient	Std Error	Std Coefficient	Tolerance	t-Stat	p-Value	Cats
Januari-24	4.249	0.941	0.385	0.103	4.516	0.000	---
Februari-24	2.979	0.947	0.427	0.103	3.145	0.004	---
(Intercept)	21.169	0.348	?	?	61.105	0	---

Gambar 1. 10 Hasil Regresi Linear

Pada gambar 1.11 menunjukkan analisis hasil algoritma regresi linear menunjukkan bahwa nilai prediksi untuk variabel April-24 pada masing-masing baris mendekati nilai aktualnya. Misalnya, untuk baris pertama dengan tipe ban "700-16 GRAD," nilai aktual adalah 21, sementara prediksi adalah 21.942, menunjukkan kesalahan prediksi yang kecil. Secara umum, prediksi model cukup baik untuk sebagian besar data, seperti pada baris kelima dan keenam, di mana nilai aktualnya adalah 24 dan 29, dengan prediksi masing-masing 24.686 dan 25.913. Variabel independen seperti Januari-24, Februari-24, dan Maret-24 berkontribusi pada model dengan nilai-nilai yang terstandarisasi, membantu

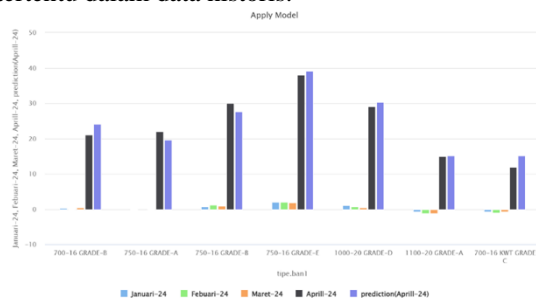
menghasilkan prediksi yang relevan. Hal ini menunjukkan bahwa model regresi linear yang dibangun mampu menangkap hubungan antara variabel independen dan variabel target dengan akurasi yang memadai.

Row No.	tipe.ban1	April-24	prediction(A...	Januari-24	Februari-24	Maret-24
1	700-16 GRAD...	21	21 942	0.388	-0.294	0.373
2	700-16 GRAD...	16	16 300	-0.756	-0.557	-0.276
3	1000-20 GRA...	20	20 175	-0.121	-0.162	0.373
4	700-16 KWT...	20	19 391	-0.121	-0.425	-0.536
5	750-16 KWT...	24	24 686	0.388	0.628	0.763
6	750-16 KWT...	29	25 913	0.769	0.496	1.022
7	750-16 KWT...	19	18 016	-0.629	-0.162	-0.406
8	1000-20 KWT...	29	26 401	0.515	1.022	0.763
9	1000-20 KWT...	23	20 959	-0.121	0.101	-0.147
10	1100-20 KWT...	20	18 311	-0.375	-0.425	-0.276
11	1100-20 KWT...	26	24 442	0.515	0.364	0.503
12	1100-20 KWT...	15	18 067	-0.248	-0.688	-0.276

ExampleSet (12 examples, 3 special attributes, 3 regular attributes)

Gambar 1. 11 Hasil Predksi Apply Model

Pada gambar 1.12 menunjukan Grafik hasil prediksi menggunakan algoritma regresi linear menunjukkan bahwa prediksi nilai penjualan untuk bulan April 2024 (digambarkan dalam kolom ungu) secara umum berada dalam rentang yang mendekati nilai aktual (kolom hitam). Model regresi linear mampu memprediksi dengan tingkat akurasi yang baik untuk sebagian besar jenis produk, seperti pada tipe 750-16 GRADE-B dan 1000-20 GRADE-D, di mana nilai prediksi hampir identik dengan nilai aktual. Namun, terdapat perbedaan yang cukup signifikan pada beberapa produk, seperti 700-16 GRADE-B dan 1100-20 GRADE-A, di mana prediksi sedikit lebih tinggi dibandingkan nilai aktual. Hal ini mengindikasikan adanya potensi bias pada model atau kurangnya sensitivitas terhadap fluktuasi tertentu dalam data historis.



Gambar 1. 12 Grafik Hasil Prediksi

Pada gambar 1.13 menunjukan hasil evaluasi model regresi linear menggunakan operator Performance menunjukkan bahwa model memiliki performa yang cukup baik. Nilai root mean squared error (RMSE) sebesar 1.778 menunjukkan rata-rata kesalahan prediksi model terhadap nilai aktual dalam skala data. Nilai absolute error sebesar 1.478 dengan deviasi ± 0.989 mengindikasikan bahwa rata-rata kesalahan prediksi absolut berada dalam rentang kecil, yang menunjukkan model cukup akurat. Selain itu, relative error sebesar $6.81\% \pm 5.09\%$ mengindikasikan bahwa kesalahan prediksi model

relatif terhadap nilai target berada dalam batas toleransi rendah. Secara keseluruhan, model ini dapat dikategorikan sebagai model yang efektif untuk memprediksi variabel target dengan tingkat kesalahan yang terkontrol dan dapat diterima untuk tujuan analisis.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 1.778 +/- 0.000
absolute_error: 1.478 +/- 0.989
relative_error: 6.81% +/- 5.09%
```

Gambar 1. 13 Hasil Perhitungan MSE, RMSE dan MAE

5. KESIMPULAN

Berikut adalah kesimpulan dari penelitian ini menunjukkan bahwa metode regresi linear dapat menjadi alat yang efektif untuk memprediksi penjualan. Dengan tahapan implementasi yang sistematis dan hasil evaluasi yang memuaskan, model ini berhasil menangkap pola hubungan antara variabel independen dan variabel target. Meskipun tingkat kesalahan relatif rendah, terdapat peluang untuk pengembangan lebih lanjut guna meningkatkan akurasi prediksi, terutama untuk menghadapi kasus-kasus yang lebih kompleks. Penelitian ini memberikan dasar yang kuat bagi penerapan regresi linear dalam analisis data penjualan dan membuka peluang untuk eksplorasi metode prediksi lainnya.

1. Metode regresi linear telah berhasil diimplementasikan dalam penelitian ini untuk memprediksi penjualan. Proses implementasi melibatkan tahapan pra-pemrosesan data seperti pemilihan atribut, penanganan nilai yang hilang, normalisasi, dan pembagian data menjadi data latih dan data uji. Selanjutnya, model regresi linear dilatih menggunakan data latih dan diaplikasikan pada data uji untuk menghasilkan prediksi. Hasil implementasi menunjukkan bahwa regresi linear dapat mengidentifikasi hubungan antara variabel-variabel independen, seperti data penjualan pada bulan sebelumnya, untuk memprediksi penjualan di bulan berikutnya secara efektif. Model ini mampu menghasilkan prediksi yang mendekati nilai aktual, membuktikan bahwa metode ini cocok untuk diterapkan pada kasus serupa.

2. Model prediksi penjualan menggunakan regresi linear terbukti cukup efektif berdasarkan hasil evaluasi yang ditunjukkan. Dengan root mean squared error (RMSE) sebesar 1.778, absolute error sebesar 1.478, dan relative error sebesar 6.81% , model ini mampu memberikan prediksi yang mendekati nilai aktual dengan tingkat kesalahan yang relatif rendah. Hal ini menunjukkan bahwa regresi linear dapat menangkap pola hubungan

antara variabel independen dan variabel target dengan baik. Meskipun model ini efektif untuk digunakan, ada ruang untuk perbaikan lebih lanjut, seperti dengan mengoptimalkan fitur atau mencoba algoritma lain untuk meningkatkan akurasi, terutama jika digunakan untuk skenario penjualan yang lebih kompleks.

UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) di Program Studi Teknik Informatika STMIK IKMI Cirebon. Penulis menyadari sepenuhnya bahwa skripsi ini masih memiliki berbagai kekurangan, yang disebabkan oleh keterbatasan kemampuan dan pengetahuan yang dimiliki. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bantuan, dukungan, dan bimbingan selama proses penulisan skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada Bapak Assoc. Prof. Dr. Dadang Sudrajat, S.Si., M.Kom, selaku Ketua STMIK IKMI Cirebon; Bapak Dian Ade Kurnia, M.Kom, selaku Wakil Ketua I Bidang Akademik, Kerjasama, Riset, dan Inovasi; Ibu Dra. Nining R., M.Si., selaku Wakil Ketua II Bidang Keuangan; Ibu Fatihanursari Dikananda, S.Tr.I.Kom., M.Kom, selaku Wakil Ketua III Bidang Kemahasiswaan dan Alumni; Bapak H. Eka Jayawangsa, BBA., selaku Wakil Ketua IV Bidang Sarana dan Prasarana; Ibu Gifthera Dwilestari, S.I.Kom., M.Kom, selaku Ketua Program Studi Teknik Informatika; Bapak Martanto, M.Kom, sebagai Dosen Pembimbing Utama; Bapak Arif Rinaldi Dikananda, M.Kom, sebagai Dosen Pembimbing Kedua; teman-teman, serta semua pihak yang telah membantu. Tak lupa, penulis juga mengucapkan terima kasih yang tak terhingga kepada orang tua dan keluarga tercinta atas doa dan dukungan yang terus menyertai selama perjalanan kuliah. Semoga semua kebaikan yang diberikan mendapatkan balasan yang berlimpah dari Allah SWT.

DAFTAR PUSTAKA

- [1] A. A.-F. N. Wahyudin, "Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu," *Techno.COM*, vol. 19, no. 4, pp. 364–374, 2020.
- [2] N. A. S. Nur Mahar Aji1, Vihi Atina2, "Pemodelan prediksi kelulusan mahasiswa dengan metode naïve bayes di uniba," *MISI (Jurnal Manaj. Inform. Sist. Informasi)*, vol. 6, pp. 148–158, 2023.
- [3] S. M. Lilis Harianti Hasibuan, "Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang," *J. Sci. Technol.*, vol. 2, no. 1, pp. 85–95, 2022.
- [4] S. W. Ade Izyuddin1, "APLIKASI PREDIKSI PENJUALAN AC MENGGUNAKAN DECISION TREE DENGAN ALGORITMA C4.5 Ade," *MISI (Jurnal Manaj. Inform. Sist. Informasi)*, vol. 3, no. 2, 2020.
- [5] A. Anggrawan and N. Azmi, "Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear Sales Prediction of Unilever Products using the Linear Regression Method," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 123–132, 2022, doi: 10.30812/bite.v4i2.2416.
- [6] A. Yaya Asohi, "Impelementasi algoritma regresi linier berganda untuk prediksi penjualan," *J. Nas. Ilmu Komput.*, vol. 1, no. 3, pp. 149–158, 2020.
- [7] G. N. Ayuni and D. Fitriana, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2020, doi: 10.61769/telematika.v14i2.321.
- [8] B. A. Wisudaningsi, "Pengaruh Kualitas Pelayanan Dan Kualitas Produk Terhadap Kepuasan Konsumen Dengan Menggunakan Metode Analisis Regresi Linier Berganda," *Pengaruh Kualitas Pelayanan Dan Kualitas Prod. Terhadap Kepuasan Konsum. Dengan Menggunakan Metod. Anal. Regresi Linear Berganda*, vol. 1, no. 1, pp. 103–117, 2019.
- [9] E. R. Tri Indarwati, Tri Irawati, "Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone," *J. Teknol. Inf. dan Komun.*, vol. 6, no. 2, pp. 2–7, 2019, doi: 10.30646/tikomsin.v6i2.369.
- [10] S. Adiguno, "Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 275, 2022, doi: 10.53513/jursi.v1i4.5331.
- [11] K. Puteri and A. Silvanie, "Machine Learning untuk Model Prediksi Harga Sembako," *J. Nas. Inform.*, vol. 1, no. 2, pp. 82–94, 2020.
- [12] E. P. Ariesanto Akhmad, "Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran," *J. Apl. Pelayaran dan Kepelabuhanan*, vol. 10, no. 2, p. 120, 2020, doi: 10.30649/japk.v10i2.83.
- [13] D. H. Taufik Hidayat, Rahmi Darnis, "ALGORITMA REGRESI LINIER BERGANDA UNTUK ANALISIS EFISIENSI STOK PRODUK DI PT. MADU PRAMUKA BATANG Taufik," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 10, no. 3, pp. 239–247, 2024.