PREDIKSI PENJUALAN PRODUK TERLARIS PADA CV. GORONTALO COMPUTER MENGGUNAKAN ALGORITMA REGRESI LINEAR

Muhammad Faldy Syahfa'ad, Wahyudin Hasyim, Hilmansyah Gani
Program Studi Sistem Informasi, Universitas Muhammadiyah Gorontalo
Jl. Prof. Dr. H. Mansoer Pateda, Desa Pentadio Timur, Kec. Telaga Biru, Kab. Gorontalo, Gorontalo 96181, Indonesia
fldysyahfaad@gmail.com

ABSTRAK

Prediksi penjualan merupakan aspek penting dalam mendukung pengambilan keputusan bisnis, khususnya dalam pengelolaan stok produk terlaris. CV. Gorontalo Computer masih menghadapi kendala dalam menentukan produk yang berpotensi laris karena proses analisis penjualan dilakukan secara manual dan belum berbasis data. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan algoritma regresi linear sederhana dalam memprediksi penjualan produk berdasarkan data historis perusahaan. Metode yang digunakan adalah pendekatan data mining terhadap data penjualan periode Januari 2022 hingga Mei 2025, dengan tahapan preprocessing, pemodelan, dan evaluasi. Proses analisis dilakukan menggunakan tiga *tools*, yaitu Microsoft Excel, RapidMiner, dan SPSS, untuk membandingkan konsistensi hasil dan kinerja model. Hasil penelitian menunjukkan bahwa variabel periode waktu tidak berpengaruh signifikan terhadap jumlah penjualan pada seluruh kategori produk, dengan nilai signifikansi lebih besar dari 0,05 dan nilai R-Square yang rendah. Evaluasi menggunakan *Root Mean Square Error* (RMSE) menunjukkan bahwa kategori laptop memiliki tingkat kesalahan terendah, sedangkan kategori aksesoris memiliki kesalahan tertinggi. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma regresi linear sederhana kurang optimal digunakan untuk prediksi penjualan pada CV. Gorontalo Computer.

Kata kunci: *Prediksi Penjualan, Regresi Linear, Data Mining, R-Square, RMSE, SPSS*

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah membawa dampak signifikan terhadap dunia bisnis, termasuk dalam sektor penjualan produk komputer [1]. Persaingan yang semakin ketat menuntut perusahaan untuk memiliki strategi yang tepat dalam mengelola persediaan, memahami kebutuhan pasar, serta memprediksi tren penjualan di masa depan. Salah satu langkah strategis yang dapat dilakukan adalah dengan memanfaatkan data historis penjualan untuk melakukan prediksi terhadap produk-produk yang berpotensi menjadi terlaris [2]. Pemanfaatan data historis ini menjadi penting karena dapat membantu perusahaan dalam mengambil keputusan yang lebih objektif dan berbasis data, bukan hanya berdasarkan pengalaman atau perkiraan semata.

Dengan adanya kemajuan teknologi di bidang data mining dan machine learning, perusahaan kini dapat memanfaatkan algoritma prediksi untuk mengolah data historis penjualan guna menghasilkan informasi yang lebih akurat dan objektif [3]. Salah satu metode yang dapat digunakan adalah algoritma regresi linear. Metode ini mampu memodelkan hubungan antara variabel independen, seperti jumlah penjualan di waktu sebelumnya, harga, promosi, dan musim dengan variabel dependen yaitu jumlah penjualan di masa mendatang [4]. Penggunaan algoritma regresi linear dinilai sesuai sebagai langkah awal dalam analisis prediksi karena memiliki konsep yang sederhana dan mudah diimplementasikan pada data penjualan.

CV. Gorontalo Computer sebagai salah satu perusahaan yang bergerak di bidang penjualan produk komputer, laptop, aksesoris, dan berbagai perangkat teknologi lainnya, telah melayani kebutuhan masyarakat Gorontalo baik individu maupun instansi. Dalam menjalankan usahanya, CV. Gorontalo Computer mencatat berbagai data transaksi penjualan setiap harinya dengan jumlah data yang cukup besar. Namun, proses pengambilan keputusan dalam menentukan produk yang berpotensi laris masih sering dilakukan secara manual dan berdasarkan intuisi. Kondisi ini menyebabkan data penjualan yang tersedia belum dimanfaatkan secara optimal untuk analisis prediktif, sehingga berpotensi menimbulkan ketidaktepatan dalam manajemen stok dan perencanaan strategi penjualan.

Penggunaan algoritma regresi linear dalam memprediksi penjualan produk terlaris diharapkan dapat membantu CV. Gorontalo Computer dalam membuat keputusan yang lebih efektif dalam hal pengelolaan persediaan, promosi, dan penyusunan strategi pemasaran. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan algoritma regresi linear sederhana dalam menganalisis dan memprediksi penjualan produk berdasarkan data historis perusahaan. Rencana penyelesaian masalah dalam penelitian ini dilakukan melalui tahapan pengumpulan data penjualan, preprocessing data, pemodelan menggunakan algoritma regresi linear, serta evaluasi kinerja model menggunakan beberapa tools analisis. Dengan pendekatan tersebut, diharapkan penelitian ini dapat memberikan gambaran mengenai kemampuan dan keterbatasan regresi linear dalam prediksi penjualan produk pada CV. Gorontalo Computer.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Menurut [5], Data mining adalah proses analisis untuk menemukan pola, hubungan, atau informasi tersembunyi dari kumpulan data besar menggunakan teknik statistik, pembelajaran mesin, dan kecerdasan buatan. Tujuannya adalah mengubah data mentah menjadi informasi yang berguna untuk pengambilan keputusan bisnis yang lebih tepat dan efisien. data mining mencakup tahapan seperti pengumpulan data, pra-pemrosesan, pemodelan, dan evaluasi hasil.

Penelitian lain yang dilakukan oleh [6] untuk menentukan penjualan barang IT pada PT. Javas Karya Tungga Jakarta Selatan menggunakan metode apriori, menunjukkan bahwa data mining sangat efektif dalam mengidentifikasi asosiasi antar produk berdasarkan data transaksi historis. Dalam penelitian tersebut, algoritma Apriori digunakan untuk menemukan hubungan antara produk-produk yang sering dibeli secara bersamaan, sehingga dapat dimanfaatkan untuk strategi pemasaran seperti bundling produk dan promosi silang (*cross-selling*).

Dari referensi di atas bisa diambil kesimpulan bahwa data mining menjadi alat penting dalam dunia bisnis modern karena kemampuannya untuk menggali informasi berharga dari data yang besar dan kompleks. Dengan mengikuti tahapan yang sistematis serta menerapkan algoritma yang sesuai, perusahaan dapat mengidentifikasi pola dan tren yang berguna dalam pengambilan keputusan, termasuk dalam meningkatkan akurasi sistem prediksi penjualan.

2.2. Prediksi

Prediksi adalah untuk memperkirakan hasil masa depan berdasarkan data historis. Dalam bisnis, sistem ini digunakan untuk memperkirakan penjualan guna membantu pengambilan keputusan seperti pengelolaan stok dan perencanaan produksi. Salah satu metode yang banyak digunakan adalah regresi linear, karena kemampuanya dalam memodelkan hubungan antara variable-variabel penentu penjualan (Pradita, 2024).

Selain itu, menurut Asrul (2024) dalam penelitiannya pada kebutuhan alat kesehatan, sistem prediksi dapat membantu perencanaan distribusi dan pengadaan barang secara lebih efisien. Dalam studi tersebut, metode Least Square digunakan untuk meramalkan kebutuhan alat kesehatan dengan hasil prediksi yang cukup akurat, sehingga bisa menjadi dasar kuat bagi manajemen logistik perusahaan dalam mengambil keputusan.

Dari landasan teori di atas, dapat disimpulkan bahwa sistem prediksi memiliki peran vital dalam mendukung pengambilan keputusan strategis perusahaan. Melalui pemanfaatan metode statistik seperti regresi linear atau Least Square, sistem ini mampu memperkirakan nilai penjualan di masa depan dengan mengidentifikasi pola dari data historis. Dengan demikian, perusahaan dapat mengantisipasi kebutuhan pasar secara lebih akurat dan efisien.

2.3. Penjualan

Penjualan merupakan proses pertukaran barang atau jasa antara penjual dan pembeli untuk memperoleh keuntungan serta memenuhi kebutuhan konsumen. Aktivitas ini tidak hanya melibatkan transaksi, tetapi juga strategi pemasaran yang efektif agar mampu meningkatkan minat beli dan loyalitas pelanggan. [7] menekankan pentingnya strategi pemasaran dalam meningkatkan penjualan produk UMKM, seperti yang terlihat pada usaha keripik pisang di Tanjung Jabung Timur.

Sejalan dengan itu, [8] menunjukkan bahwa strategi pemasaran dan kualitas produk sangat berpengaruh terhadap keputusan pembelian konsumen. [9] juga menambahkan bahwa media digital, seperti media sosial, kini menjadi sarana yang efektif untuk menjangkau pasar yang lebih luas dan meningkatkan volume penjualan, khususnya di sektor bisnis kreatif seperti florist.

Bedasarkan pernyataan diatas, penjualan tidak hanya bergantung pada transaksi langsung, tetapi juga dipengaruhi oleh strategi pemasaran, kualitas produk, dan pemanfaatan teknologi digital. Optimalisasi ketiga aspek tersebut sangat penting untuk meningkatkan penjualan dan daya saing bisnis secara berkelanjutan.

2.4. Produk Terlaris

Produk terlaris adalah produk yang memiliki volume penjualan tertinggi dalam suatu periode tertentu dibandingkan produk lain di dalam suatu perusahaan atau pasar. Identifikasi produk terlaris sangat penting bagi perusahaan karena dapat membantu dalam perencanaan stok, strategi pemasaran, dan pengambilan keputusan bisnis secara umum. Menurut [10], analisis produk terlaris membantu perusahaan dalam mengoptimalkan manajemen stok, merencanakan strategi pemasaran, dan memprediksi permintaan pasar secara lebih akurat. Dalam konteks prediksi penjualan, mengetahui produk terlaris memungkinkan perusahaan untuk fokus pada produk yang memiliki potensi pasar terbesar sehingga dapat meningkatkan efisiensi dan efektivitas operasional.

Salah satu penelitian yang relevan adalah oleh [11] untuk prediksi penjualan produk terlaris menggunakan metode K Nearest Neighbor (KNN). Dalam penelitian ini, digunakan data penjualan dari Rakuni Bakery, dan algoritma KNN diterapkan untuk memprediksi produk-produk terlaris. Hasil pengujian menunjukkan akurasi model sebesar 85%, dan *cross-validation* menghasilkan akurasi 87%, yang menunjukkan bahwa metode klasifikasi mampu memprediksi produk laris dengan baik

Berdasarkan pernyataan diatas Pemahaman mengenai produk terlaris sangat penting dalam strategi bisnis karena produk ini memiliki kontribusi besar terhadap pendapatan perusahaan. Dengan mengidentifikasi produk yang paling diminati

konsumen berdasarkan data penjualan, perusahaan dapat mengoptimalkan stok, memfokuskan promosi, serta meningkatkan efisiensi distribusi dan pemasaran. Dukungan metode analisis data seperti K-Means maupun regresi linear dalam proses ini membantu perusahaan dalam pengambilan keputusan yang lebih tepat dan berbasis data historis yang valid.

2.5. CV. Gorontalo Computer

CV Gorontalo Computer adalah Perusahaan yang berlokasi di Kota Gorontalo, Provinsi Gorontalo tepatnya di Jl. Prof. Dr. H.B. Jassin, Dulalowo, Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128, perusahaan yang bergerak di bidang penjualan perangkat keras dan perangkat lunak komputer, serta menyediakan layanan servis dan perawatan perangkat teknologi informasi. Perusahaan ini melayani berbagai kebutuhan teknologi bagi individu, institusi pendidikan, instansi pemerintah, dan pelaku usaha di wilayah Gorontalo dan sekitarnya.

2.6. Algoritma Regresi Linear

Menurut [12], penerapan regresi linear berganda dalam prediksi penjualan pada marketplace menunjukan bahwa variabel jumlah produk terjual dan potongan biaya pengiriman berpengaruh signifikan terhadap jumlah total penjualan. Model yang dibangun memiliki nilai koefisien determinasi (R2) sebesar 91,9% dan *Mean Absolute Percentage Error* (MAPE) sebesar 8,99% yang menandakan bahwa model tersebut dapat diandalkan untuk mendukung pengambilan keputusan dalam meningkatkan penjualan di marketplace.

Secara matematis, regresi linear sederhana dirumuskan pada Persamaan (1).

$$y = a + bX \quad (1)$$

Dimana:
Y adalah variabel dependen (hasil yang diprediksi)
X adalah variabel independent
a adalah intercept (titik potong sumbu-Y)
b adalah slope (kemiringan garis regresi atau koefisien regresi).

2.7. Penelitian Terdahulu

Penelitian terdahulu yang relevan menunjukkan bahwa algoritma regresi linear sederhana banyak digunakan dalam prediksi penjualan berbasis data historis. Putri, Ratih Dewi (2022) menerapkan regresi linear sederhana untuk memprediksi produk elektronik terlaris pada CV [13]. Istana Komputer Palembang dan menunjukkan bahwa metode ini cukup efektif dalam memodelkan tren penjualan serta membantu pengambilan keputusan terkait stok dan pemasaran. Penelitian lain oleh Amansyah et al. (2024) menggunakan regresi linear sederhana untuk memprediksi penjualan kendaraan Toyota di Indonesia dengan variabel waktu sebagai variabel independen, dan hasil evaluasi menggunakan RMSE serta MAPE menunjukkan tingkat akurasi yang baik, sehingga metode ini dinilai dapat diandalkan dalam mendukung keputusan strategis di sektor otomotif. Selain itu, Aqsho Ramadhan et al. (2023) juga membuktikan bahwa regresi linear sederhana mampu memberikan estimasi penjualan handphone yang mendekati nilai aktual dengan tingkat error yang masih dapat diterima, sehingga dapat digunakan sebagai alat bantu awal dalam perencanaan stok dan strategi promosi pada sektor ritel elektronik. Secara umum, penelitian-penelitian tersebut menegaskan bahwa regresi linear sederhana memiliki potensi sebagai metode prediksi penjualan, meskipun tingkat akurasinya sangat dipengaruhi oleh karakteristik data dan variabel yang digunakan.

3. METODE PENELITIAN

3.1 Objek Penelitian

Objek dalam penelitian ini adalah data historis penjualan pada CV Gorontalo Computer. Penelitian ini bertujuan untuk memprediksi produk terlaris berdasarkan data penjualan yang telah dikumpulkan selama periode tertentu. Fokus dari penelitian ini adalah menerapkan Algoritma Regresi Linear untuk mengetahui pola dan tren penjualan yang dapat digunakan sebagai dasar dalam pengambilan keputusan bisnis seperti pengelolaan stok dan perencanaan strategi pemasaran.

Dataset yang digunakan dalam penelitian ini merupakan data internal perusahaan, yang terdiri dari nama produk, jumlah penjualan, dan waktu transaksi, dengan menggunakan data tersebut, penelitian ini bertujuan membangun model prediksi unntuk mengetahui produk-produk yang cenderung memiliki angka penjualan tertinggi dimasa mendatang.

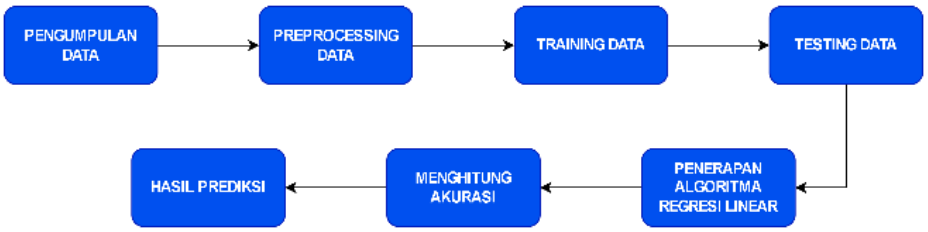
3.2 Metode Penelitian

Penelitian ini dilakukan dengan menerapkan pendekatan data mining untuk memanfaatkan data historis atau data numerik dari hasil penjualan produk di CV. Gorontalo Computer. Proses penelitian ini mengikuti tahapan umum *Knowledge Discovery* in Database, yaitu pengumpulan data, preprocessing, pemodelan, evaluasi dan interpretasi [14]. Data tersebut kemudian diolah menggunakan teknik data mining, khususnya menggunakan Algoritma Regresi Linear, guna membangun model prediksi terhadap penjualan produk terlaris.

Metode penelitian ini bertujuan untuk menemukan hubungan antara variabel-variabel yang mempengaruhi jumlah penjualan, seperti waktu penjualan dan jenis produk. Melalui pendekatan ini, maka peneliti dapat memperoleh gambaran prediktif yang membantu dalam pengambilan keputusan strategis perusahaan, seperti perencaan stok dan strategi promosi.

3.3 Tahapan Penelitian

Adapun tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

a. Pengumpulan Data
Data dikumpulkan melalui wawancara langsung dengan CV Gorontalo Computer untuk memperoleh informasi penjualan yang akan dianalisis.
b. Preprocessing Data
Data yang dikumpulkan kemudian dibersihkan dari duplikasi dan data kosong. Data juga ditransformasikan kemudian dipisahkan menjadi data latih dan data uji.
c. Pembagian dan Pelatihan Data
Training data digunakan untuk melatih model agar dapat mengenali pola dan hubungan antar variabel dalam dataset. Proses ini penting supaya model mampu membuat prediksi yang relevan berdasarkan data historis. Testing data digunakan untuk mengukur kinerja dan akurasi model terhadap data baru yang belum pernah dilihat oleh model sebelumnya. Hasil pengujian ini akan menentukan seberapa baik model dapat digunakan untuk prediksi di dunia nyata.
d. Penerapan Algoritma Regresi Linear
Pada tahap ini, algoritma regresi linear diterapkan untuk membangun model prediktif. Model ini bekerja dengan mencari hubungan linier antara variabel independen (X) dan variabel dependen (Y).
e. Hasil Prediksi
Pada tahap ini, peneliti menyimpulkan hasil dari model regresi linear dan prediksi yang telah dilakukan, serta mengevaluasi sejauh mana tujuan penelitian berhasil dicapai. Selain itu, peneliti juga memberikan rekomendasi yang berkaitan dengan pengembangan model di masa mendatang atau penerapan hasil penelitian dalam konteks nyata, seperti perencanaan stok barang atau prediksi penjualan untuk kebutuhan operasional toko.

3.4 Dataset

Dataset yang digunakan dalam penelitian ini berasal dari data penjualan CV Gorontalo Computer. Data diatas ini dikumpulkan dalam format excel (.xlsx) dan berisi catatan transaksi penjualan produk dari januari 2022 sampai dengan mei 2025. Tujuan penggunaan dataset ini adalah untuk memprediksi produk terlaris dengan menggunakan algoritma regresi linear. Dataset terdiri dari 11 atribut utama, yaitu:

1. No Dok.- Nomor dokumen transaksi penjualan
2. Tgl- Tanggal transaksi penjualan
3. Kode Departemen- Kode kategori barang.
4. Deskripsi Departemen- Nama kategori barang, seperti “KABEL” atau “KONEKTOR”.
5. Kode Barang- Kode unik untuk setiap barang
6. Nama Barang- Nama lengkap barang yang dijual
7. Qty- Jumlah unit barang yang terjual
8. Satuan- jenis satuan barang (contoh: unit, meter)
9. Qty*- salinan dari kolom Qty
10. Harga- harga satuan untuk setiap barang
11. Total (IDR)- total nilai transaksi per baris data (Qty x Harga)

Dataset ini memiliki lebih dari 29.409 baris, mencerminkan volume transaksi harian CV. Gorontalo Computer. Dengan data tersebut, penelitian ini dapat menganalisis pola pembelian konsumen serta mengidentifikasi barang dengan tingkat penjualan tertinggi. Untuk keperluan analisis, atribut-atribut utama yang digunakan adalah Tanggal, Nama Barang, Qty, Harga, dan Total (IDR) yang ditunjukkan pada Tabel 1.

Tabel 1. Contoh data set penjualan produk CV Gorontalo Computer

No	Tanggal Transaksi	Nama Barang	Kategori	Jumlah Terjual (Qty)	Harga Satuan (Rp)	Total Penjualan (Rp)
1	2022-01-03	Audio Headphone & Mic Jack Splitter	Konektor	1	20.000	20.000
2	2022-01-03	Audio Splitter 3.5mm To Mic & Headphone	Konektor	1	30.000	30.000
3	2022-01-03	Kabel Data Harddisk Sata	Kabel	1	15.000	15.000
…	…	…	…	…	…	…
…	…	…	…	…	…	…
29404	2025-05-16	Epson Ink 003-Black	Tinta	1	100.000	100.000
29405	2025-05-16	Hdmi Splitter 4 Port	Kabel	1	165.000	165.000

No	Tanggal Transaksi	Nama Barang	Kategori	Jumlah Terjual (Qty)	Harga Satuan (Rp)	Total Penjualan (Rp)
29406	2025-05-16	Canon Printer G3730 Green	Printer	1	2.410.000	2.410.000

3.5 Preprocessing

Preprocessing adalah tahap penting dalam proses data mining yang bertujuan untuk mempersiapkan data mentah agar layak digunakan dalam pemodelan. Data yang tidak lengkap, tidak konsisten, atau memiliki format yang tidak sesuai dapat menyebabkan hasil prediksi yang kurang akurat. Oleh karena itu, pada penelitian ini dilakukan beberapa langkah preprocessing terhadap dataset penjualan dari CV Gorontalo Computer, sebagai berikut:

1. Pembersihan Data (Data Cleaning)
   Langkah pertama adalah menghapus data yang duplikat dan memeriksa nilai kosong (missing values). Beberapa transaksi yang tidak memiliki informasi lengkap, seperti jumlah terjual (Qty) atau Harga (Harga), dihapus agar tidak mengganggu proses perhitungan regresi.
2. Transformasi Data
   Kolom Tgl diubah ke format datetime agar dapat dilakukan pengelompokan berdasarkan waktu seperti harian, mingguan, bulanan. Ini penting dalam analisis tren penjualan produk.
3. Seleksi Atribut
   Dari seluruh atribut yang tersedia, hanya beberapa yang digunakan untuk proses pemodelan, yaitu
   a. Nama Barang sebagai variabel penentu produk
   b. Qty sebagai variabel pendukung dalam analisis hubungan antar fitur.
4. Normalisasi (jika diperlukan)
   Jika data memiliki skala yang sangat berbeda (misalnya harga jutaan dan jumlah hanya satuan), maka akan dilakukan normalisasi agar setiap fitur berada dalam rentang nilai yang sama, seperti menggunakan metode Min-Max Scaling. Namun, karena regresi linear dapat tetap bekerja meskipun tanpa normalisasi, tahap ini bersifat opsional tergantung hasil eksplorasi data awal.
5. Pembuatan Dataset Akhir
   Setelah semua tahapan diatas dilakukan, data disimpan kembali dalam bentuk CSV atau Excel dengan kolom-kolom yang sudah disesuaikan. Dataset ini kemudian digunakan dalam tahap pelatihan dan pengujian model

3.6 Training dan Testing

Setelah data selesai melalui proses preprocessing, langkah selanjutnya adalah membagi dataset menjadi dua bagian utama, yaitu data training dan data testing. Pembagian ini bertujuan untuk melatih model menggunakan sebagian data dan menguji performanya menggunakan data yang belum pernah dilihat oleh model.

1. Pembagian Dataset
   Total Dataset berjumlah 29.408, kemudian dataset dibagi dengan proposi umum yaitu: 80% data training sebanyak 23.526,4 data yang digunakan untuk membentuk model prediksi berdasarkan pola yang ditemukan dalam data. Kemudian 20% data testing sebanyak 5.881,6 yang digunakan untuk mengevaluasi sejauh mana model mampu memprediksi data baru yang belum pernah dianalisis sebelumnya
2. Implementasi Regresi Linear
   Model regresi linear dibangun menggunakan data training, di mana hubungan antara fitur-fitur seperti harga dan qty (atau total) dianalisis. Tujuannya adalah untuk mengetahui bagaimana fitur fitur tersebut memengaruhi jumlah penjualan suatu produk.
3. Pengujian Model (testing)
   Setelah model dilatih, pengujian dilakukan dengan membandingkan hasil prediksi dari model terhadap nilai aktual dari data testing. Performa model kemudian diukur menggunakan metrik evaluasi seperti nilai *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *R-squared* (R^2).
4. Tools
   Penelitian ini menggunakan dua perangkat lunak utama dalam proses pengolahan dan analisis data, yaitu:
   a. Microsoft Excel, digunakan pada tahap awal untuk melakukan pengecekan struktur data, menghapus data duplikat, serta menyelaraskan format dataset sebelum diimpor ke RapidMiner. Excel berperan sebagai alat bantu dalam pembersihan awal data (*initial data cleaning*).
   b. Rapid Minners, digunakan sebagai alat utama untuk proses preprocessing, training, testing, dan evaluasi model regresi linear. Rapid Miner dipilih karena menyediakan antarmuka visual yang memudahkan dalam membangun alur kerja data mining secara sistematis dan efisien, tanpa memerlukan pemrograman.
   c. SPSS
      Digunakan untuk melakukan analisis statistik regresi melalui output Model Summary, ANOVA, dan Coefficients, guna memvalidasi signifikansi variabel dan kekuatan model.

3.7 Algoritma Regresi Linear

Persamaan umum dari Regresi Linear Sederhana adalah sebagai berikut:

1. Pengumpulan Data
   Data diperoleh dari file transaksi penjualan harian pada CV Gorontalo Computer, yang berisi informasi mengenai tanggal transaksi, nama barang, kategori barang, jumlah barang, jumlah terjual, harga satuan, dan total penjualan.
2. Pemilihan Variabel

Variabel independen yang digunakan adalah jumlah terjual (Qty) dan harga satuan (Rp). Sedangkan variabel dependen adalah total penjualan (Rp).

3. Pra-pemrosesan Data
   Data dibersihkan dari duplikasi dan entri kosong. Kategori barang dapat dikonversi menjadi variabel dummy jika dianggap memengaruhi total penjualan.
4. Pembentukan Model Regresi Linear
   Menggunakan metode Least Squares, model regresi linear dibentuk dengan menghitung koefisiensi regresi (b) dan konstanta (a) yang meminimalkan selisih antara nilai prediksi dan nilai aktual
5. Pelatihan dan Pengujian Model
   Dataset dibagi menjadi dua bagian: data latih dan data uji. Model dilatih menggunakan data latih, dan kemudian diuji akurasinya menggunakan data uji.
6. Evaluasi Model
   Kinerja model di evaluasi menggunakan metrik seperti RMSE, MAE, dan R^2 untuk menilai seberapa baik model melakukan prediksi.

3.8 Akurasi

Dalam penelitian ini, akurasi model regresi linear diukur menggunakan metode RMSE. RMSE ini merupakan salah satu metrik yang paling umum digunakan untuk mengukur seberapa dekat nilai yang diprediksi oleh model terhadap nilai aktualnya. Semakin kecil nilai RMSE, semakin baik model dalam memprediksi. Rumus RMSE ditunjukkan pada Persamaan (2).

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

Dimana:
y_i = nilai aktual total penjual
$\hat{y}_i$ = nilai prediksi total penjualan
n = jumlah data

3.9 Hasil Prediksi Terbaik

Setelah dilakukan pelatihan model regresi linear dan evaluasi akurasi menggunakan metode RMSE, diperoleh model regresi yang ditunjukkan pada Persamaan (3). Kemudian contoh hasil prediksi terbaik dapat dilihat pada Tabel 2.

$$y = 4700 + 1300 \cdot x_1 + 0.85 \cdot x_2 \quad (3)$$

Dimana:
y = total penjualan (Rp)
x_1 = jumlah terjual (Qty)
x_2 = harga satuan (Rp)

Tabel 2 . Contoh Hasil Prediksi

No	Jumlah terjual (Qty)	Harga satuan (Rp)	Penjualan aktual (Rp)	Penjualan prediksi (Rp)	Selisih (Error)
1	3	120.000	360.000	363.700	3.700
2	1	50.000	50.000	52.250	2.250
3	2	85.000	170.000	172.200	2.200

Dari Tabel 2 dapat dilihat bahwa hasil prediksi sangat mendekati nilai aktual. Perbedaan nilai yang kecil menunjukkan bahwa model regresi linear ini cukup efektif dalam memperkirakan nilai total penjualan berdasarkan jumlah terjual dan harga satuan.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Dalam Penelitian ini data yang diperoleh adalah dari arsip transaksi penjualan CV. Gorontalo Computer, yang diambil dari Laporan detail penjualan periode Januari 2022 hingga Mei 2025. Data tersebut merupakan data primer yang bersumber langsung dari sistem pencatatan penjualan atau transaksi perusahaan, yang meliputi informasi mengenai tanggal transaksi, kode produk, nama produk, jumlah terjual, serta harga satuan.

Secara umum, data transaksi mentah yang diperoleh terdiri dari 29.408 baris transaksi dengan 4 kategori pokok produk yang berbeda. Produk yang dijual mencakup 4 kategori pokok yaitu, Laptop, Printer, Sparepart, dan Aksesoris. Untuk memudahkan analisis, data transaksi penjualan diolah menjadi data agregat bulanan berdasarkan 4 kategori pokok produk. Ringkasan ini menampilkan jumlah total produk yang terjual dalam periode penelitian, ditunjukkan pada Tabel 3.

Tabel 3. Ringkasan Data Penjualan Produk

No	Tanggal Transaksi	Nama Barang	Kategori	Jumlah Terjual (Qty)	Harga Satuan (Rp)	Total Penjualan (Rp)
1	2022-01-03	Audio Headphone & Mic Jack Splitter	Konektor	1	20.000	20.000
2	2022-01-03	Audio Splitter 3.5mm To Mic & Headphone	Konektor	1	30.000	30.000
3	2022-01-03	Kabel Data Harddisk Sata	Kabel	1	15.000	15.000
…	…	…	…	…	…	…
…	…	…	…	…	…	…

No	Tanggal Transaksi	Nama Barang	Kategori	Jumlah Terjual (Qty)	Harga Satuan (Rp)	Total Penjualan (Rp)
29404	2025-05-16	Epson Ink 003-Black	Tinta	1	100.000	100.000
29405	2025-05-16	Hdmi Splitter 4 Port	Kabel	1	165.000	165.000
29406	2025-05-16	Canon Printer G3730 Green	Printer	1	2.410.000	2.410.000

4.1.1 Transformasi Data

Tahap transformasi dilakukan untuk mengubah data transaksi harian menjadi data agregat bulanan berdasarkan kategori produk. Tujuan utama dari proses ini adalah untuk mempermudah analisis pola penjualan serta menyesuaikan bentuk data dengan kebutuhan algoritma Regresi Linear. Transformasi dilakukan dengan mengelompokkan seluruh transaksi berdasarkan bulan dan kategori produk, kemudian menghitung jumlah total penjualan (Qty) pada masing-masing kategori. Selain itu, ditambahkan kolom *Total Qty* untuk menunjukkan jumlah keseluruhan produk yang terjual dalam satu bulan. Contoh dataset setelah transformasi dapat dilihat pada Tabel 4.

Tabel 4. Contoh dataset setelah transformasi (Ringkasan Kategori Produk)

Date YY/MM	Laptop	Printer	Sparepart	Aksesoris	Total Qty
202201	2	3	84	95	184
202202	1	1	73	85	160
202203	2	2	80	139	223
202204	4	0	78	94	176
202205	1	3	81	111	196
…	…	…	…	…	…

4.2 Analisis Hasil Penjualan Produk

Berdasarkan hasil pengujian algoritma regresi linear sederhana menggunakan aplikasi SPSS terhadap data penjualan 4 kategori produk, dapat disimpulkan bahwa variabel periode sebagai representasi waktu tidak memiliki pengaruh yang signifikan terhadap jumlah penjualan pada keempat kategori tersebut. Hal ini terlihat dari nilai signifikansi seluruh model yang lebih besar dari 0,05, sehingga secara statistik hubungan antara waktu dan jumlah dinyatakan tidak signifikan. Secara keseluruhan, hasil pengujian menunjukkan bahwa penggunaan algoritma regresi linear pada ke 4 data kategori produk ini hanya memberi gambaran umum mengenai tren data, namun tidak cukup kuat untuk melakukan prediksi penjualan jangka panjang.

Transformasi data memberikan manfaat yang sangat penting dalam penelitian ini. Data mentah yang awalnya kompleks berhasil disederhanakan menjadi data agregat bulanan yang lebih terstruktur dan mudah dianalisis. Dengan bentuk data seperti ini, pola tren penjualan dapat diamati dengan lebih jelas sehingga proses prediksi menggunakan metode Regresi Linear dapat dilakukan secara lebih efektif.

Berdasarkan hasil pengujian model regresi linear sederhana terhadap data penjualan pada CV. Gorontalo Computer, diperoleh nilai koefisien determinasi (R^2) yang relatif rendah pada seluruh kategori produk. Nilai R^2 berkisar antara 0,00 hingga 0,046, yang menunjukkan bahwa variabel waktu sebagai variabel independen tidak mampu menjelaskan variasi jumlah penjualan secara signifikan.

Selain itu, hasil evaluasi menggunakan Root Mean Square Error (RMSE) menunjukkan bahwa tingkat kesalahan prediksi bervariasi pada setiap kategori produk. Kategori laptop memiliki nilai RMSE terendah, yaitu berkisar antara 1,07 hingga 1,209, yang menunjukkan bahwa model relatif lebih stabil dalam memprediksi penjualan laptop dibandingkan kategori lainnya. Sebaliknya, kategori aksesoris memiliki nilai RMSE tertinggi, yaitu mencapai 38,363, yang menunjukkan bahwa model kurang mampu menangkap pola penjualan pada kategori tersebut.

Rendahnya nilai R^2 dan tingginya nilai error pada beberapa kategori menunjukkan bahwa hubungan antara variabel waktu dan jumlah penjualan tidak bersifat linear. Hal ini mengindikasikan bahwa terdapat faktor lain yang mempengaruhi penjualan, seperti promosi, tren pasar, musim, serta perilaku konsumen, yang tidak dimasukkan dalam model.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi dari ketiga tools menunjukkan konsistensi dalam kinerja dan keterbatasan model. Untuk kategori laptop yang dapat dinilai paling stabil, nilai RMSE berkisar antara 1,07 (Excel) hingga 1,209 (SPSS), dan nilai R^2 menunjukkan pengaruh variabel periode yang sangat rendah, yaitu sekitar 0,00 (Excel dan Rapidminer) hingga 0,046 (SPSS). Sementara itu, kategori aksesoris memliki tingkat kesalahan prediksi tertinggi, dengan nilai RMSE mencapai 38,363 (SPSS) dan

37,145 (Rapidminer). Nilai R^2 yang rendah pada seluruh kategori membuktikan bahwa variabel periode tidak memiliki hubungan signifikan dengan jumlah penjualan, sehingga perubahan penjualan lebih dipengaruhi oleh faktor lain diluar model. Dengan demikian, dapat disimpulkan bahwa metode regresi linear kurang optimal digunakan sebagai metode prediksi penjualan pada kasus ini, namun tetap memberikan informasi penting mengenai kestabilan kategori penjualan serta arah trend penjualan perusahaan. Untuk penelitian selanjutnya, disarankan menggunakan metode prediksi lain yang lebih mampu menangani data non-linear dan fluktuatif seperti Random Forest, Support Vector Machine (SVM), atau metode time series seperti ARIMA agar hasil prediksi penjualan menjadi lebih akurat dan optimal.

DAFTAR PUSTAKA

[1] Miftahuljannah, Aswan Supriyadi Sunge, and Ahmad Turmudi Zy, "Analisis Prediksi Penjualan Dengan Metode Regresi Linear Di Pt. Eagle Industry Indonesia," *J. Inform. Teknol. dan Sains*, vol. 5, no. 3, pp. 398–403, 2023, doi: 10.51401/jinteks.v5i3.3325.

[2] D. A. Pohan, M. H. Dar, and Irmayanti, "Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana," *J. Nas. Inform. dan Teknol. Jar.*, vol. 2, no. 1, pp. 2–6, 2022.

[3] M. R. Athallah and A. F. Rozi, "Implementasi Data Mining Untuk Prediksi Peramalan Penjualan Produk Hj Karpet Menggunakan Metode Linear Regression," *J. Sains dan Teknol.*, vol. 2, no. 3, pp. 180–187, 2023, doi: 10.47233/jsit.v2i3.550.

[4] F. Ramdhani and K. Setiawan, "Penerapan Data Mining untuk Prediksi Pelanggan di PT. XYZ Menggunakan Algoritma Linear Regression," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 490–497, 2024, doi: 10.57152/malcom.v4i2.1217.

[5] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[6] S. Lestari And A. Damaiyanti, "Penerapan Data Mining Menggunakan Metode Algoritma Apriori Untuk Menentukan Penjualan Barang It Pada Pt. Javas Karya Tungga Jakarta Selatan," *J. Sains Dan Teknol.*, Vol. 5, No. 1, Pp. 237–242, 2023.

[7] R. Adolph, "Strategi Pemasaran Untuk Meningkatkan Penjualan Produk Keripik Pisang 'Merasa'dikabupaten Tanjung Jabung Timur," Vol. 12, No. 1, Pp. 1–23, 2025.

[8] F. E. . Siregar, R. D., Ulya, B., & Srimulat, "Pengaruh Strategi Pemasaran, Kualitas Produk Terhadap Minat Beli Konsumen Pada Toko OKEY Elektronik, Kota Pinang," Sultra Journal Of Economic And Business, Vol. 6, No. April, Pp. 395–402, 2025.

[9] R. D. Sartika, "Strategi Pemasaran Melalui Media Digital Untuk Meningkatkan Penjualan Pada Produk Kerajinan Tangan," Vol. 6, No. 2, Pp. 1289–1292, 2023.

[10] D. P. Tri Wahyudi, Naini Sa'adah, "Penerapan Metode K-Means Clustering Pada Data Penjualan Optik Chantika," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 7, No. 2, Pp. 1303–1307, 2023, Doi: 10.36040/Jati.V7i2.6562.

[11] M. A. A. Sandy Aldy Pradana, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Rakuni Bakery, Pastry And Cake)," *Oktal J. Ilmu Komput. Dan Sci.*, Vol. 3, No. 6, Pp. 1596–1610, 2024.

[12] M. Putri, "Prediksi Penjualan Marketplace Menggunakan Regresi Linear Berganda Berdasarkan Produk Terjual Dan Potongan Biaya Pengiriman (Studi Kasus : Aufar Grosir Jeans)," Vol. 13, No. 3, Pp. 609–624, 2025.

[13] D. R. Putri, "Prediksi Penjualan Produk Elektronik Yang Terlaris Pada Cv. Istana Komputer Palembang Menggunakan Algoritma Regresi Linear Sederhana Skripsi," 2022.

[14] Y. A. Ramadhan, A. Faqih, And G. Dwilestari, "Prediksi Penjualan Handphone di Toko X Menggunakan ALgoritma Regresi Linear," Jurnal Informatika Terpadu, Vol. 9, No. 1, Pp. 40–44, 2023.