

Data Mining Untuk Prediksi Penjualan Menggunakan Metode Simple Linear Regression

Patrisius Ando Duran¹, Anik Vega Vitianingsih^{2*}, Moch. Syaiful Riza³,
Anastasia Lidya Maukar⁴, Seftin Fitri Ana Wati⁵

^{1,2,3} Program Studi Teknik Informatika, Universitas Dr. Soetomo, Surabaya, Jawa Timur

⁴ Program Studi Teknik Industri, President University, Bekasi, Jawa Barat

⁵ Program Studi Sistem Informasi, Universitas Pembangunan Nasional Veteran, Surabaya, Jawa Timur

Email: ¹ androduran80@gmail.com, ^{2*} vega@unitomo.ac.id, ³ syaiful.riza@unitomo.ac.id,

⁴ almaukar@president.ac.id, ⁵ seftin.fitri.si@upnjatim.ac.id

(Naskah masuk: 13 Nov 2023, direvisi: 10 Jan 2024, diterima: 11 Jan 2024)

Abstrak

Prediksi penjualan menjadi elemen penting dalam perencanaan perusahaan karena akan menentukan rencana anggaran penjualan, yang mempengaruhi banyak faktor dalam perusahaan. Produsen penjualan sering kali mengalami kesulitan untuk memprediksi kisaran jumlah produk yang terjual pada periode yang akan datang. Permasalahan tersebut mengakibatkan ketidakefektifan dalam pengelolaan stok dan jadwal produksi, sehingga stok produk digudang menumpuk yang mengakibatkan beberapa produk mengalami kerusakan karena disimpan terlalu lama. Tujuan dari penelitian ini adalah mengimplementasikan metode *Simple Linear Regression* untuk memprediksi penjualan produk agar dapat mengetahui tentang kisaran jumlah penjualan produk di periode yang akan datang, sehingga dapat menyesuaikan penyediaan stok yang lebih efektif berdasarkan prediksi yang diperoleh. Metode *Simple Linear Regression* digunakan untuk mengevaluasi dan memahami arah serta kekuatan hubungan antara dua variabel, yaitu variabel independen (X) dan variabel dependen (Y). Parameter yang digunakan adalah periode bulan dan jumlah penjualan. Periode bulan yang merujuk pada rentang waktu, bagaimana perubahan waktu dapat memengaruhi jumlah penjualan produk sedangkan jumlah penjualan merupakan *output* yang menjadi fokus utama, mengetahui sejauh mana variabilitas dalam periode bulan dapat dijelaskan oleh variasi dalam jumlah penjualan produk. *Dataset* yang digunakan untuk uji tingkat kesalahan terhadap hasil prediksi menggunakan data penjualan mulai periode Februari 2021 sampai dengan September 2023. Hasil uji menyatakan nilai rata-rata dari selisih absolut antara nilai prediksi atau MAD sebesar 3,778563, rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual atau MSE sebesar 21,661444 dan rata-rata persentase kesalahan absolut antara nilai prediksi dan nilai aktual atau MAPE sebesar 12%. Berdasarkan nilai MAPE yang diperoleh, dapat disimpulkan bahwa prediksi penjualan untuk penjualan produk ini dapat dikategorikan Baik. Hasil penelitian ini dapat dijadikan rekomendasi untuk memprediksi kisaran jumlah penjualan produk di periode yang akan datang agar menyesuaikan persediaan stok, anggaran dan jadwal produksi

Kata Kunci: *Data Mining*, Prediksi Penjualan, Prediksi Stok Penjualan, *Simple Linear Regression*

Data Mining for Clothing Sales Prediction Using Simple Linear Regression Method

Abstract

Sales forecasting is an important element in company planning because it will determine the sales budget plan, which affects many factors in the company. Sales producers often have difficulty predicting the range of products sold in the coming period. These problems result in ineffectiveness in managing stock and production schedules, so that product stock in the warehouse accumulates which results in some products being damaged because they are stored for too long. The purpose of this research is to implement the Simple Linear Regression method to predict product sales in order to find out about the range of product sales in the coming period, so that it can adjust more effective stock provision based on the predictions obtained. The Simple Linear Regression method is used to evaluate and understand the direction and strength of the relationship between two

variables, namely the independent variable (X) and the dependent variable (Y). The parameters used are month period and sales amount. The month period refers to the time span, how changes in time can affect the number of product sales while the number of sales is the output that is the main focus, knowing the extent to which variability in the month period can be explained by variations in the number of product sales. The dataset used to test the error rate of the prediction results uses sales data from February 2021 to September 2023. The test results state the average value of the absolute difference between predicted values or MAD of 3.778563, the average of the squared difference between predicted values and actual values or MSE of 21.661444 and the average percentage of absolute error between predicted values and actual values or MAPE of 12%. Based on the MAPE value obtained, it can be concluded that sales predictions for sales of this product can be categorized as Good. The results of this study can be used as a recommendation to predict the range of product sales in the coming period in order to adjust the stock inventory, budget and production schedule.

Keywords : Data Mining, Sales Prediction, Sales Stock Prediction, Simple Linear Regression

I. PENDAHULUAN

Pesatnya pertumbuhan industri di era globalisasi saat ini, membuat persaingan antara perusahaan di Indonesia semakin ketat [1]. Memenangkan persaingan penjualan menjadi tujuan utama pembuatan strategi penjualan dan strategi promosi. Merancang strategi perusahaan memerlukan kemampuan untuk memprediksi potensi peristiwa yang akan mempengaruhi perusahaan di masa mendatang, termasuk prediksi penjualan. Prediksi penjualan adalah kegiatan yang memperkirakan produk yang akan dijual dimasa depan dengan mempertimbangkan kondisi khusus melalui analisis data historis [2]. Prediksi penjualan menjadi elemen penting dalam perencanaan perusahaan karena akan menentukan rencana anggaran penjualan, yang memengaruhi pengembangan produk, perencanaan biaya operasional, proyeksi laba-rugi, dan neraca perusahaan. Volume penjualan merupakan jumlah total yang diperoleh dari hasil penjualan barang [3]. Penjualan bulanan merupakan angka-angka dari hasil jumlah produk yang terjual setiap bulan [4].

Toko 99 merupakan sebuah toko yang beralamat di Jl. Pumpungan 2 No.33 Surabaya. Toko 99 bergerak dibidang *fashion* pria dan wanita, menjual produk seperti baju kaos, jaket, celana, sepatu, topi dan lainnya. Terdapat permasalahan yang dialami oleh toko, salah satunya adalah ketidakmampuan untuk memprediksi kisaran jumlah produk yang terjual di masa yang akan datang, mengakibatkan ketidakefektifan dalam pengelolaan stok dan jadwal produksi, dengan ini stok produk di gudang menumpuk dan beberapa produk mengalami kerusakan karena disimpan terlalu lama. Selain itu, pengelolaan data penjualan pada toko masih dilakukan secara manual seperti pencatatan data transaksi yang masih menggunakan buku tulis, yang tentu pencatatan transaksi seperti ini memiliki banyak kekurangan seperti ketidakefisienan, potensi kesalahan, kesulitan dalam pencarian data, kurangnya keamanan, ketidakmampuan akses jarak jauh, serta kurangnya kemudahan dalam pelaporan. Oleh karena itu, diperlukan modernisasi dalam sistem pencatatan transaksi untuk meningkatkan efisiensi dan keamanan operasional toko.

Penelitian sebelumnya [5] menggunakan algoritma C4.5 untuk memprediksi jumlah penjualan rokok pada PT.

Indomarco Prismatama, dengan parameter nama produk, jenis, kategori, ukuran dan minat, didapatkan nilai tingkat akurasi 92,11%, nilai tingkat kesalahan 7,894% dan nilai AUC 87,8%. Tingkat akurasi sebesar 92,11% dan tingkat kesalahan 7,894% menunjukkan kinerja model yang baik. Algoritma C4.5 efektif untuk memprediksi penjualan rokok dengan mempertimbangkan parameter seperti nama produk, jenis, kategori, ukuran, dan minat, perbedaan dengan penelitian ini terletak pada metode dan parameter yang digunakan. Sementara penelitian [6] menggunakan metode *Simple Linear Regression* untuk memprediksi persediaan jenis obat, dengan parameter periode minggu dan tablet, didapatkan akurasi sebesar 98,505%. Akurasi menunjukkan bahwa model tersebut mampu memberikan prediksi persediaan obat dengan tingkat akurasi yang tinggi. Perbedaan dengan penelitian ini terletak pada fokus variabel prediktor dan objek prediksi. Sementara penelitian [7] menggunakan metode *Simple Linear Regression* untuk memprediksi penyebaran vaksin *covid-19* di kabupaten Cilacap, dengan parameter kasus positif dan vaksin dosis 1, didapatkan nilai MAPE 11,04% yang berarti tingkat akurasiannya sebesar 88,96%. Akurasi menunjukkan bahwa model ini secara efektif dapat memprediksi penyebaran vaksin dengan tingkat keakuratan yang baik, perbedaan dengan penelitian ini terletak pada studi kasus dan parameter yang digunakan. Sementara penelitian [8] menggunakan metode *Simple Linear Regression* untuk memprediksi jumlah penerimaan mahasiswa baru, dengan menggunakan parameter periode tahun dan jumlah mahasiswa, didapatkan nilai *error* 3,444% yang berarti tingkat akurasinya sebesar 96,556%. Akurasi yang dihasilkan menunjukkan model ini mampu memprediksi jumlah penerimaan mahasiswa baru dengan tingkat akurasi yang tinggi, perbedaan dengan penelitian ini terletak pada studi kasus dan parameter yang digunakan. Sedangkan penelitian [9] menggunakan metode *Simple Linear Regression* untuk prediksi penjualan *liquid vape* di toko Vapor Pandaan, dengan menggunakan parameter jumlah terjual dan jumlah stok, didapatkan nilai MAPE sebesar 2.332285115304 atau 2%. Akurasi yang dihasilkan menunjukkan bahwa model ini memiliki tingkat akurasi yang tinggi dalam memprediksi penjualan *liquid vape*. Perbedaan

dengan penelitian ini terletak pada studi kasus dan variabel yang digunakan. Sedangkan penelitian [10] menggunakan metode *least square* untuk prediksi jumlah pendapatan dengan menggunakan parameter unit waktu, curah hujan dan pendapatan, didapatkan nilai MAPE Pada bulan April sebesar 17,5 dan pada bulan Mei mendapatkan nilai MAPE sebesar 133,6. Akurasi yang dihasilkan menunjukkan algoritma *least square* tergolong baik untuk memprediksi jumlah pendapatan, perbedaan dengan penelitian ini terletak pada studi kasus dan parameter yang digunakan.

Berdasarkan studi literatur tersebut, perlunya melakukan penelitian dengan studi kasus dan parameter yang berbeda untuk memastikan keakuratan dan keandalan hasil prediksi yang dihasilkan.

Tujuan dari penelitian ini adalah menciptakan sebuah aplikasi yang mampu melakukan prediksi terhadap jumlah penjualan dengan menerapkan metode *Simple Linear Regression*, dengan menggunakan variabel bulan sebagai parameter independen dan jumlah penjualan sebagai variabel dependen. Parameter yang digunakan adalah periode bulan dan jumlah penjualan. Periode bulan yang merujuk pada rentang waktu, bagaimana perubahan waktu dapat memengaruhi jumlah penjualan produk sedangkan jumlah penjualan merupakan output yang menjadi fokus utama, mengetahui sejauh mana variabilitas dalam periode bulan dapat dijelaskan oleh variasi dalam jumlah penjualan produk Kemampuan dari aplikasi yang di bangun yaitu mampu memprediksi kisaran jumlah penjualan di periode yang akan datang. Penelitian ini memilih metode *Simple Linear Regression* sebagai alat prediksi karena keunggulannya dalam kemampuan menyusun model sederhana serta kemampuan untuk memodelkan data sekuensial. Selain itu, *Simple Linear Regression* memungkinkan analisis yang lebih akurat dengan melibatkan beberapa variabel independen (X), sehingga memungkinkan hasil prediksi yang lebih tepat [11].

II. STUDI PUSTAKA

A. Data Mining

Data Mining adalah tahap-tahap yang dilakukan untuk mengekstrak informasi berharga serta pola yang berguna dari sejumlah data yang besar [12]. Data mining pada prosesnya merupakan sebuah metode untuk menggali informasi berharga dengan cara menganalisis pola atau hubungan tertentu dalam data yang berskala besar. Dalam dunia bisnis, data mining merupakan teknik yang digunakan untuk mendukung beragam aplikasi kecerdasan bisnis, seperti profil pelanggan, pemasaran yang tepat sasaran, manajemen alur kerja, tata letak toko, dan deteksi penipuan [13]. Teknik data mining dapat memberikan jawaban terhadap pertanyaan krusial dalam bisnis, seperti "siapa pelanggan yang paling berpotensi memberikan keuntungan tertinggi?" dan "bagaimana proyeksi pendapatan perusahaan untuk tahun depan?" [13]

Data Mining dapat dikelompokkan menjadi beberapa kelompok, sesuai tugas yang dapat dilakukan yaitu [12]:

1. Deskripsi

Penerapan *data mining* bertujuan untuk menemukan metode yang mudah dipahami untuk mengidentifikasi pola serta tren yang ada dalam data. Menguraikan pola dan tren tersebut membuka peluang untuk memberikan penjelasan terhadap aspek-aspek tertentu dari pola dan tren tersebut.

2. Estimasi

Dalam proses estimasi, variabel target lebih cenderung mengarah pada nilai numerik dibandingkan kategori. Saat mengembangkan model, data lengkap yang mencakup nilai variabel target digunakan sebagai dasar prediksi. Kemudian, pada tahap peninjauan berikutnya, nilai estimasi untuk variabel target dihitung berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi mirip dengan klasifikasi dan estimasi, tetapi perbedaannya terletak pada fokusnya pada nilai hasil di masa depan.

4. Klasifikasi

Dalam pemrosesan klasifikasi variabel, tujuannya adalah untuk mengategorikan variabel ke dalam kelompok-kelompok tertentu yang bersifat kategoris. Sebagai contoh, pengelompokan persediaan ke dalam tiga kategori: persediaan tinggi, persediaan sedang, dan persediaan rendah

5. Pengklusteran

Tidak seperti klasifikasi, pengklusteran tidak melibatkan variabel target. Pengklusteran dilakukan tanpa melakukan usaha untuk mengklasifikasi, mengestimasi, atau meramalkan nilai dari variabel yang dijadikan tujuan. Pengklusteran bertujuan untuk memecah data keseluruhan menjadi kelompok-kelompok yang memiliki tingkat kesamaan atau homogenitas yang tinggi, di mana rekor-rekor dalam satu kelompok memiliki kesamaan maksimum di antara mereka, sementara kesamaan dengan rekor-rekor dalam kelompok lain adalah minimal.

6. Asosiasi

Asosiasi dalam data mining memiliki peran dalam mengidentifikasi atribut-atribut yang muncul bersamaan dalam suatu waktu.

Data mining mempunyai serangkaian tahapan dalam proses menggali informasi, yaitu [14]

1. Data cleaning

Proses membersihkan data yang kotor atau tidak terpakai yang menyebabkan eror dalam perhitungan

2. Data Integration

Proses penggabungan data dari sumber yang berbeda menjadi satu kesatuan.

3. Data Selection

Proses penyeleksian data yang sesuai dengan perhitungan,

4. Data transformation

Data yang dipilih mengalami proses transformasi, diubah menjadi prosedur *data mining* melalui agregasi dan konsolidasi sehingga data memiliki distribusi yang sesuai dengan yang diharapkan.

5. Data Mining

Proses di mana beragam teknik diterapkan untuk mengekstraksi berbagai pola yang berpotensi, sehingga

kita dapat menggali data yang memiliki nilai dan relevansi yang signifikan.

6. *Pattern evolution*

Proses di mana informasi yang signifikan dipahami dengan mengidentifikasinya berdasarkan ukuran (*measure*) yang telah ditentukan sebelumnya.

7. *Knowledge Presentation*

Proses terakhir dalam *data mining* di mana dilakukan penyajian visual dari hasil penelitian untuk mempermudah pemahaman pengguna terhadap informasi yang ditemukan dari eksplorasi data.

B. Prediksi

Prediksi merupakan kegiatan yang dilakukan seseorang untuk melakukan perkiraan terhadap data baik perkiraan untuk saat ini ataupun untuk masa yang akan datang [14]. Proses prediksi erat kaitannya dengan perhitungan matematis pada *dataset* yang besar. Prediksi dapat dilakukan melalui berbagai macam metode yang terkait, kadang-kadang menggabungkan dua atau lebih metode untuk memperoleh hasil prediksi yang akurat. Prediksi dapat membantu perusahaan dalam hal pengambilan keputusan perusahaan seperti menentukan tingkat persediaan barang yang optimal serta membantu dalam manajemen stok dengan harapan dapat mengurangi risiko kesalahan perencanaan seminimal mungkin. Penggunaan prediksi umumnya bertujuan untuk mengekstrak wawasan dari data set yang luas, dan oleh karena itu, data mining menjadi suatu kebutuhan esensial [15]

C. Regresi Linier

Metode regresi linier merupakan metode yang digunakan untuk menilai dampak variabel bebas, yang dapat disebut sebagai variabel independen atau variabel X (faktor pemicu), terhadap variabel tergantung, juga dikenal sebagai variabel dependen atau variabel Y (hasil) [16]. Salah satu metode regresi linier adalah metode *Simple Linear Regression*. *Simple Linear Regression* merupakan salah satu metode statistik yang umum digunakan dalam upaya untuk memprediksi atau meramalkan karakteristik kualitatif dan kuantitatif [17]. Metode *Simple Linear Regression* dalam melakukan peramalan menggunakan Persamaan (1)

$$Y=a+bX \tag{1}$$

Perhitungan nilai a akan dicari menggunakan Persamaan (2)

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

Perhitungan nilai b akan dicari menggunakan Persamaan (3)

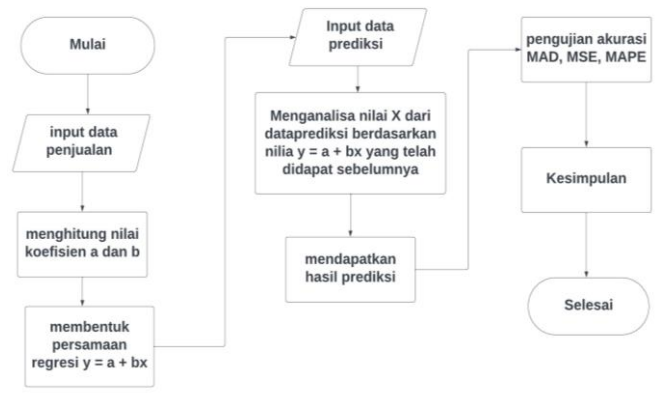
$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \tag{3}$$

Dengan variabel y mewakili jumlah penjualan, x sebagai periode atau bulan penjualan, a merupakan konstanta yang

menunjukkan nilai y ketika x = 0, dan b menggambarkan perubahan besaran nilai y.

III. METODE

Pembuatan sistem prediksi penjualan pakaian menggunakan metode *Simple Linear Regression* melibatkan beberapa proses yang dapat divisualisasikan dalam bentuk *flowchart*. *Flowchart* ini memberikan gambaran umum tentang bagaimana melakukan prediksi penjualan produk menggunakan metode *Simple Linear Regression*, Keseluruhan *flowchart* dapat membantu memvisualisasikan proses secara menyeluruh dan membantu memahami bagaimana setiap langkah berhubungan satu sama lain dalam mencapai tujuan akhir penjualan. *Flowchart* adalah representasi grafis yang mengilustrasikan urutan logis dari operasi dan langkah-langkah dalam sebuah program atau prosedur sistem [18]. *Flowchart* juga berperan sebagai alat komunikasi yang efektif antara berbagai pemangku kepentingan, memudahkan pemecahan masalah, perencanaan proyek, dokumentasi sistem, pelatihan, dan evaluasi kinerja. *Flowchart* algoritma sistem prediksi penjualan pakaian menggunakan metode *Simple Linear Regression* akan ditampilkan pada Gambar 1.



Gambar 1. *Flowchart* Algoritma

Pada Gambar 1 proses perhitungan metode *Simple Linear Regression*. Admin memasukkan data penjualan yang berasal dari transaksi yang telah dibersihkan. Setelah itu, sistem akan menentukan nilai koefisien a dan b untuk membentuk persamaan regresi. Selanjutnya, sistem akan menggunakan nilai a dan b tersebut untuk menentukan $y = a + bx$, yang digunakan dalam perhitungan. Setelah mendapatkan nilai $y = a + bx$, admin dapat langsung melakukan proses prediksi. Proses ini dimulai dengan memasukkan data prediksi ke dalam sistem. Sistem kemudian menganalisis nilai X berdasarkan $y = a + bx$ yang telah dihasilkan sebelumnya, memberikan hasil prediksi yang dapat digunakan untuk mengevaluasi atau merencanakan kebijakan berdasarkan tren penjualan yang diprediksi. Setelah memperoleh hasil prediksi. Selanjutnya melakukan pengujian akurasi dengan mencari nilai MAD, MSE dan MAPE. Setelah nilai akurasi didapat, selanjutnya admin langsung dapat menyimpulkan dan

mengambil keputusan yang lebih terinformasi mengenai strategi penjualan dan stok produk.

Kebutuhan data yang digunakan pada penelitian ini adalah sebagai berikut:

1. Transaksi: Data transaksi merupakan data yang menginformasikan transaksi pada toko. Sumber data ini berasal dari toko 99, yang nantinya digunakan dalam proses perhitungan untuk memprediksi jumlah penjualan. Sebelum di gunakan, data ini di bersihkan terlebih dahulu untuk menghilangkan variabel yang tidak terpakai ataupun nilai yang hilang.
2. Penjualan: Data penjualan merupakan data yang siap digunakan dalam proses perhitungan untuk memprediksi penjualan pakaian. Data ini berasal dari data transaksi yang telah dibersihkan dan siap dimasukkan dalam proses perhitungan. Data penjualan dapat dilihat pada Tabel 1.

Tabel 1. Parameter Independen dan dependen

Parameter	Keterangan
X	Periode bulan
Y	Jumlah penjualan

IV. HASIL DAN PEMBAHASAN

Hasil dan pembahasan berikut akan membahas langkah-langkah metode *Simple Linear Regression* dalam melakukan prediksi penjualan produk. Jenis produk yang tersedia pada Toko 99 yaitu *dress* cewek, celana kolor, *flanel*, kaos band, kemeja, celana jeans, kemeja cewek, celana kargo pendek, *sweater*, jaket *bombers*, *polo*, *huddie* polos, *hotpans*, jaket bahan jeans, celana pendek *cinos*. Penelitian ini mengambil produk yang pertama yaitu produk *dress* yang akan lakukan proses perhitungan, untuk produk selanjutnya dapat menerapkan formula yang sama. Tahapan – tahapan pada penelitian ini, yaitu:

1. Pengumpulan data

Tahap ini merupakan tahap mengumpulkan data yang di butuhkan selama proses perhitungan, kemudian disusun secara kronologis untuk mempermudah analisis dan penggunaan dalam metode *Simple Linear Regression*. Data penjualan *dress* bulan Februari 2021 hingga bulan September 2023 akan dilihat pada Tabel 2.

Tabel 2. Data Penjualan Dress

Keterangan	Bulan	Jml Penjualan
Februari-2021	2	31
Maret-2021	3	33
April-2021	4	34
Mey-2021	5	41
Juni-2021	6	32
Juli-2021	7	32
Agustus-2021	8	35
September- 2021	9	32
Oktober-2021	10	39

Keterangan	Bulan	Jml Penjualan
November-2021	11	41
Desember-2021	12	30
Januari-2022	13	30
Februari-2022	14	31
Maret-2022	15	28
April-2022	16	35
Mey-2022	17	30
Juni-2022	18	32
Juli-2022	19	30
Agustus-2022	20	28
September- 2022	21	30
Oktober-2022	22	30
November-2022	23	31
Desember-2022	24	32
Januari-2023	25	30
Februari-2023	26	28
Maret-2023	27	27
April-2023	28	31
Mey-2023	29	34
Juni-2023	30	31
Juli-2023	31	33
Agustus-2023	32	30
September- 2023	33	32

2. Pre-processing

Setelah data dikumpul dan urutkan, selanjutnya melakukan *pre-processing*, tetapi pada Tabel 2 dapat dilihat bahwa tidak ditemukannya data *outlier*, sehingga memungkinkan untuk bisa langsung ke tahap selanjutnya.

3. Perhitungan nilai X, Y, XY dan X²

Pada tahap ini melakukan perhitungan nilai X, Y, XY, dan X² berdasarkan Tabel 2, yang berisi data penjualan *dress* di toko 99. Hasil perhitungan X, Y, XY dan X² dapat ditemukan dalam Tabel 3.

Tabel 3. Perhitungan X,Y, XY, X²

Keterangan	X	Y	X ²	XY
Februari-2021	2	31	4	62
Maret-2021	3	33	9	99
April-2021	4	34	16	136
Mey-2021	5	41	25	205
Juni-2021	6	32	36	192
Juli-2021	7	32	49	224
Agustus-2021	8	35	64	280
September- 2021	9	32	81	288
Oktober-2021	10	39	100	390
November-2021	11	41	121	451
Desember-2021	12	30	144	360
Januari-2022	13	30	169	390
Februari-2022	14	31	196	434
Maret-2022	15	28	225	420
April-2022	16	35	256	560

Keterangan	X	Y	X ²	XY
Mey-2022	17	30	289	510
Juni-2022	18	32	324	576
Juli-2022	19	30	361	570
Agustus-2022	20	28	400	560
September- 2022	21	30	441	630
Oktober-2022	22	30	484	660
November-2022	23	31	529	713
Desember-2022	24	32	576	768
Januari-2023	25	30	625	750
Februari-2023	26	28	676	728
Maret-2023	27	27	729	729
April-2023	28	31	784	868
Mey-2023	29	34	841	986
Juni-2023	30	31	900	930
Juli-2023	31	33	961	1023
Agustus-2023	32	30	1024	960
September- 2023	33	32	1089	1056
Total	560	1039	12528	17713

4. Mencari nilai a dan b

Pada tahap ini, dilakukan perhitungan untuk mencari nilai koefisien a dan b berdasarkan hasil perhitungan yang dilakukan pada Tabel 2.

Nilai a :

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$$

$$a = \frac{((1039 * 12528) - (560 * 17713))}{((32 * 12528) - 313600)}$$

$$a = 35,48$$

Nilai b:

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$b = \frac{((32 * 17713) - (560 * 1039))}{((32 * 12528) - 313600)}$$

$$b = -0,17$$

Dari hasil perhitungan di atas, nilai koefisien a mendapatkan nilai sebesar 35,48 dan b sebesar -0,17. Nilai-nilai ini digunakan untuk membentuk persamaan regresi linier dengan bentuk $Y = a + bX$, dengan ini dapat disimpulkan bahwa bentuk persamaan regresi yang diperoleh adalah $Y = 35,48 + (-0,17)X$.

5. Melakukan Prediksi

Setelah persamaan regresinya didapat, selanjutnya melakukan proses prediksi. Penelitian ini akan memprediksi kisaran jumlah penjualan produk *dress* di lima bulan yang akan datang, mulai dari Oktober 2023 sampai Februari 2024,

seperti pada Tabel 4.

Tabel 4. Data Prediksi

Keterangan	X	Y
Okt-23	34	?
Sep-23	35	?
Des-23	36	?
Jan-24	37	?
Feb-24	38	?

Selanjutnya, menerapkan persamaan regresi $Y=35,48+(-0,17)X$ ke dalam proses perhitungan dengan menggantikan nilai variabel independen (X), seperti pada Tabel 5.

Tabel 5. Implementasi Persamaan Regresi

Keterangan	X	$Y=35,48+(-0,17)X$
Okt-23	34	$Y=35,48+(-0,17)34$
Sep-23	35	$Y=35,48+(-0,17)35$
Des-23	36	$Y=35,48+(-0,17)36$
Jan-24	37	$Y=35,48+(-0,17)37$
Feb-24	38	$Y=35,48+(-0,17)38$

Berdasarkan hasil perhitungan pada Tabel 5, maka diperoleh prediksi kisaran jumlah penjualan *dress* pada toko 99 dari Oktober 2023 sampai Februari 2024 akan disajikan pada Tabel 6 .

Tabel 6. Data Hasil Prediksi

Keterangan	X	Y
Okt-23	34	30
Sep-23	35	30
Des-23	36	29
Jan-24	37	29
Feb-24	38	29

6. Pengujian Akurasi

Evaluasi metode dilakukan untuk menguji seberapa akuratnya metode *Simple Linear Regression* dalam memprediksi penjualan pakaian pada toko 99, dengan ini perlu melakukan perbandingan hasil prediksi yang dihasilkan oleh metode dengan data aktualnya. Evaluasi ini mengacu pada tiga metrik utama, yaitu :

a) *Mean Absolute Deviation* (MAD)

MAD berfungsi untuk mengukur sejauh mana ketepatan prediksi dengan cara menghitung rata-rata dari nilai kesalahan absolut pada setiap prediksi [19]. Nilai MAD akan dicari menggunakan Persamaan (3).

$$MAD = \frac{1}{n} \sum Y_t - Y'_t \tag{3}$$

b) *Mean Squared Error* (MSE)

MSE berfungsi untuk mengukur seberapa besar rata-rata kesalahan kuadrat antara data aktual dan hasil prediksi [19]. Nilai MSE akan dicari menggunakan Persamaan (4)

$$MSE = \frac{1}{n} \sum (Y_t - Y'_t)^2 \tag{4}$$

c) Mean Absolute Percentage Error (MAPE)

MAPE berfungsi untuk mengukur perbandingan antara data asli dengan hasil prediksi, diikuti oleh perhitungan untuk menentukan selisihnya [20]. Nilai MSE akan dicari menggunakan Persamaan (5)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - Y'_t|}{Y_t} \times 100\% \tag{5}$$

Perhitungan nilai MAPE mempunyai range penilaian seperti pada Tabel 7 [21].

Tabel 7. Nilai Range MAPE

Range MAPE	Keterangan
MAPE ≤ 10%	Tinggi
10% < MAPE ≤ 20%	Baik
20% < MAPE ≤ 50%	Reasonable
MAPE > 50%	Rendah
MAPE ≤ 10%	Tinggi

Pada perhitungan MAD, MSE dan MAPE, langkah awal yang dilakukan adalah mencari nilai prediksi jumlah penjualan yang dihasilkan dari metode untuk membuat perbandingan dengan data aktualnya. Data prediksi ini dicari menggunakan persamaan regresi yaitu $Y = 35,48 + (-0,17)X$ yang di hasilkan pada perhitungan sebelumnya. Perhitungan untuk mencari nilai MAD, MSE, dan MAPE akan disajikan pada Tabel 8.

Tabel 8. Proses Perhitungan Nilai MAD, MSE, MAPE

X(t)	Y(t)	Y't=35,48 + (-,17)X	Yt - Y't	(Yt-Y't) ²	((Yt-Y't)/Yt)
2	31	35,137	4,137	17,11	0,13
3	33	34,965	1,965	3,86	0,06
4	34	34,793	0,793	0,63	0,02
5	41	34,621	6,379	40,69	0,16
6	32	34,449	2,449	6,00	0,08
7	32	34,277	2,277	5,18	0,07
8	35	34,105	0,895	0,80	0,03
9	32	33,933	1,933	3,74	0,06
10	39	33,761	5,239	27,45	0,13
11	41	33,589	7,411	54,92	0,18
12	30	33,417	3,417	11,68	0,11
13	37	33,245	3,755	14,10	0,10
14	32	33,073	1,073	1,15	0,03
15	31	32,901	1,901	3,61	0,06
16	27	32,729	5,729	32,82	0,21
17	30	32,557	2,557	6,54	0,09
18	26	32,385	6,385	40,77	0,25
19	40	32,213	7,787	60,64	0,19
20	31	32,041	1,041	1,08	0,03
21	36	31,869	4,131	17,07	0,11
22	40	31,697	8,303	68,94	0,21
23	31	31,525	0,525	0,28	0,02

X(t)	Y(t)	Y't=35,48 + (-,17)X	Yt - Y't	(Yt-Y't) ²	((Yt-Y't)/Yt)
24	25	31,353	6,353	40,36	0,25
25	30	31,181	1,181	1,39	0,04
26	42	31,009	10,991	120,80	0,26
27	27	30,837	3,837	14,72	0,14
28	23	30,665	7,665	58,75	0,33
29	31	30,493	0,507	0,26	0,02
30	30	30,321	0,321	0,10	0,01
31	28	30,149	2,149	4,62	0,08
32	35	29,977	5,023	25,23	0,14
33	27	32,557	2,805	7,87	0,10
Total		120,914		693,17	3,72

Berdasarkan perhitungan pada Tabel 8, selanjutnya nilai MAD dihitung menggunakan Persamaan (3), nilai MSE menggunakan Persamaan (4), dan nilai MAPE menggunakan Persamaan (5).

- a) $MAD = \frac{1}{n} \sum Y_t - Y'_t$
 $= \frac{1}{32} 120,914$
 $= 3,778563$
- b) $MSE = \frac{1}{n} \sum (Y_t - Y'_t)^2$
 $= \frac{1}{32} \times 693,17$
 $= 21,661444$
- c) $MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - Y'_t|}{Y_t} \times 100\%$
 $= \left(\frac{1}{32} \times 3,72 \right) \times 100\%$
 $= 12\%$

Hasil dari prediksi untuk penjualan *dress* pada Toko 99 menggunakan Metode *Simple Linear Regression* memperoleh nilai MAD 3,778563, MSE 21,661444 dan MAPE 12%. Berdasarkan nilai MAPE yang diperoleh, dapat disimpulkan bahwa tingkat akurasi model prediksi berada dalam kategori Baik. MAPE sebesar 12% tersebut menunjukkan bahwa kesalahan relatif antara prediksi dan nilai aktual cukup rendah, interpretasi bahwa model memiliki kinerja yang baik dalam memprediksi kisaran jumlah penjualan produk di periode yang akan datang.

V. KESIMPULAN

Penelitian ini menghasilkan aplikasi data mining untuk memprediksi kisaran jumlah penjualan produk di bulan yang akan datang dengan metode *Simple Linear Regression* digunakan untuk mengevaluasi dan memahami arah serta kekuatan hubungan antara dua variabel, yaitu variabel independen (X) dan variabel dependen (Y). Analisis yang mendalam terhadap data historis, pihak toko dapat mengidentifikasi pola, tren, dan hubungan antar variabel yang mendasari perilaku konsumen dan dinamika penjualan. Aplikasi yang dibuat dapat mempermudah pihak toko dalam

mengambil keputusan terkait pengelolaan persediaan stok, jadwal produksi, dan strategi pemasaran, secara langsung membantu mengurangi risiko kerugian seperti kelebihan atau kekurangan stok. Akurasi yang dihasilkan untuk prediksi penjualan salah satu produk, yaitu *dress*, menunjukkan nilai MAD sebesar 3,778563, MSE sebesar 21,661444 dan MAPE sebesar 12%. Berdasarkan nilai MAPE yang diperoleh, dapat disimpulkan bahwa prediksi penjualan untuk produk *dress* ini dapat dikategorikan sebagai baik. Prediksi produk berikutnya, dapat menggunakan metode atau formula yang sama dengan harapan hasil yang baik dan akurat, seperti yang telah dicapai untuk produk *dress* sebelumnya.

REFERENSI

- [1] F. D. Ramadhani and M. Ardhiyansyah, *Sistem Prediksi Penjualan Dengan Metode Single Exponential Smoothing Dan Trend Parabolik*. Tangerang Selatan: Pascal Books, PT.Mediatama Digital Cendekia, 2021.
- [2] S. Manalu and S. J. Poluan, *Cara Akurat Menyusun Penganggaran Perusahaan Manufaktur*. Malang: Seribu Bintang, 2019.
- [3] F. Rangkuti, *Strategi Promosi Yang Kreatif Dan Analisis Kasus Integrated Marketing Communication*. Jakarta: PT. Gramedia Pustaka Utama, 2009.
- [4] I. D. Sudirman, *Data-Driven Entrepreneur: Bisnis Berdaya Saing Dengan Data Science Dan RapidMiner*. Jakarta Selatan: Salemba Infotek, 2023.
- [5] M. Leonardi, R. Emilda, I. Katrin, and A. Yulianto, "Prediksi Penjualan Produk Rokok Pada pt. Indomarco Prisma menggunakan algoritma c4.5," *Paradigma - Jurnal Komputer dan Informatika*, vol. 23, no. 2, 2021. doi:10.31294/p.v23i2.11151
- [6] Harsiti, Z. Muttaqin, and E. Srihartini, "Penerapan metode Regresi Linier Sederhana Untuk prediksi Persediaan obat jenis tablet," *JSiI (Jurnal Sistem Informasi)*, vol. 9, no. 1, pp. 12–16, 2022. doi:10.30656/jsii.v9i1.4426
- [7] R. Kurniawan, "Penerapan metode Regresi Linear Sederhana Untuk prediksi penyebaran vaksin covid 19 di Kabupaten Cilacap," *Journal ICTEE*, vol. 4, no. 1, p. 43, 2023. doi:10.33365/jictee.v4i1.2696
- [8] A. A. Muhartini, O. Sahroni, S. D. Rahmawati, T. Febrianti, and I. Mahuda, "Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana," *Jurnal Bayesian : Jurnal Ilmiah Statistika dan Ekonometrika*, vol. 1, no. 1, 2021. doi:10.46306/bay.v1i1.2
- [9] O. J. Ababil, S. A. Wibowo, and H. Zulfia Zahro', "Penerapan metode regresi Linier Dalam Prediksi Penjualan liquid vape di toko vapor pandaan berbasis website," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 1, pp. 186–195, 2022. doi:10.36040/jati.v6i1.4537
- [10] R. A. Laksono, S. Achmadi, and A. P. Sasmito, "Implementasi Data Mining Menggunakan Metode Least Square Untuk Memprediksi Jumlah Pendapatan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, 2023.
- [11] G. N. Ayuni and D. Fitrihanah, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *Jurnal Telematika*, vol. 14, no. 2, 2019. doi:10.61769/jurtel.v14i2.321
- [12] Amna et al., *Data Mining*. Padang: PT. Global Eksekutif Teknologi, 2023.
- [13] G. A. Pradnyana and K. Agustini, *Konsep Dasar Data Mining*. Tangerang Selatan: Universitas Terbuka, 2022.
- [14] Mustika et al., *DATA MINING DAN APLIKASINYA*. Bandung: Widina Bhakti Persada, 2021.
- [15] A. Alfani W.P.R., F. Rozi, and F. Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 1, pp. 155–160, 2021. doi:10.29100/jipi.v6i1.1910
- [16] S. H. Sahir, *Metodologi Penelitian*. Jogjakarta: KBM Indonesia, 2021.
- [17] Z. Matondang and H. F. Nasution, *Praktik Analisis Data: Pengolahan Ekonometrika Dengan Eviews & SPSS*. Medan: Merdeka Kreasi, 2021.
- [18] Kusriani and A. Koniyo, *Tuntunan Praktis Membangun Sistem Informasi Akuntansi Dengan Visual Basic Dan Microsoft SQL Server*. Yogyakarta: CV Andi Offset, 2007.
- [19] R. Ambarwati and Supardi, *Manajemen Operasional Dan Implementasi Dalam Industri*. Sidoarjo: Pustaka Rumah Cinta, 2021.
- [20] G. P. N. Hakim, D. Septiyana, A. Firdausi, F. R. I. Mariati, and M. Kika, *Sistem Fuzzy: Panduan Lengkap Aplikatif*. Yogyakarta: Andi, 2021.
- [21] A. K. Wardhani et al., *Teknik Peramalan Pada Teknologi Informasi*. Padang: PT. Global Eksekutif Teknologi, 2022.