

Analisa Dataset Penjualan Teh Menggunakan Algoritma Linear Regresi

Iis Riyana^{1*}, Nana Suarna², Gifthera Dwilestari³

Program Studi Teknik Informatika, STMIK IKMI Cirebon, Indonesia

Email: ^{1*}iisriyana5@gmail.com, ²st_nana@yahoo.com, ³gifthera.ikmi@gmail.com

Email Corresponding Author: iisriyana5@gmail.com

Abstrak-Dalam industri minuman, analisis penjualan menjadi salah satu faktor kunci untuk keberhasilan bisnis. Dalam konteks ini, analisis dataset penjualan teh dapat memberikan wawasan yang berharga bagi perusahaan teh dalam hal pengambilan keputusan strategis. Dalam penelitian ini menggunakan algoritma linear regresi untuk menganalisis dataset penjualan teh dan mengidentifikasi pola dan tren yang terkait dengan penjualan. **Masalah** yang dihadapi dalam analisis dataset penjualan teh adalah menemukan hubungan antara variabel penjualan dengan variabel-variabel yang mempengaruhinya, seperti harga, promosi, cuaca, dan sebagainya. Selain itu, juga perlu memahami apakah faktor-faktor ini secara signifikan berdampak pada penjualan teh dan sejauh mana dampak tersebut. **Tujuan** penelitian ini adalah untuk menganalisis dataset penjualan teh menggunakan algoritma linear regresi guna menemukan hubungan antara variabel penjualan dengan variabel-variabel yang mempengaruhinya. Selain itu, peneliti bertujuan untuk mengidentifikasi faktor-faktor yang signifikan dalam memprediksi penjualan teh dan memberikan wawasan yang berharga bagi perusahaan teh. **Metode** dalam penelitian ini menggunakan metode algoritma linear regresi. mengumpulkan dataset penjualan teh yang mencakup variabel-variabel seperti harga teh, promosi, cuaca, dan penjualan harian. Setelah melakukan analisis eksplorasi data dan pra-pemrosesan, kami menerapkan model linear regresi untuk menemukan hubungan antara variabel-variabel tersebut. peneliti juga melakukan evaluasi model untuk memeriksa keakuratan dan kecocokannya dengan data. **Hasil** analisis ini menunjukkan adanya hubungan signifikan antara variabel penjualan teh dengan faktor-faktor seperti harga teh dan promosi. Dengan adanya penurunan harga teh dan peningkatan promosi berkontribusi positif terhadap peningkatan penjualan. Namun, faktor cuaca tidak terbukti memiliki dampak yang signifikan pada penjualan teh dalam dataset Model linear regresi yang kami terapkan juga menunjukkan tingkat keakuratan yang baik dalam memprediksi penjualan teh berdasarkan variabel-variabel yang dianalisis.

Kata Kunci: Analisa, Penjualan, The Jawa, Linear Regression

Abstract- In the beverage industry, sales analysis is one of the key factors for business success. In this context, analysis of tea sales datasets can provide valuable insights for tea companies in terms of strategic decision making. In this study using a linear regression algorithm to analyze the tea sales dataset and identify patterns and trends related to sales. The problem encountered in the analysis of tea sales datasets is finding the relationship between sales variables and the variables that influence them, such as price, promotion, weather, and so on. In addition, it is also necessary to understand whether these factors significantly impact tea sales and the extent of this impact. The purpose of this study was to analyze the tea sales dataset using a linear regression algorithm to find the relationship between sales and the variables that influence it. In addition, the researcher aims to identify factors that are significant in predicting tea sales and provide valuable insights for tea companies. The method in this study uses a linear regression algorithm method. collects a tea sales dataset that includes variables such as tea price, promotions, weather, and daily sales. After conducting exploratory and pre-processing data analysis, we apply a linear regression model to find the relationship between the variables. the researcher also evaluates the model to check its accuracy and suitability with the data. The results of this analysis indicate that there is a significant relationship between tea sales and factors such as tea prices and promotions. With the reduction in tea prices and increased promotions contributed positively to increased sales. However, the weather factor is not proven to have a significant impact on tea sales in the dataset. The linear regression model that we apply also shows a good level of accuracy in predicting tea sales based on the variables analyzed.

Keywords: Analysis, Sales, Javanese Tea, Linear Regression

1. PENDAHULUAN

Strategi pemasaran mempunyai tujuan yang dapat dicapai melalui usaha mempertahankan dan meningkatkan tingkat keuntungan atau laba perusahaan dengan cara mempertahankan dan meningkatkan penjualannya. Penelitian ini dilakukan di PT. Sari Melati Sejahtera pada bulan Januari sampai dengan bulan April 2023. Data yang digunakan dalam penelitian ini adalah data primer dan sekunder. Data primer diperoleh dari pengamatan langsung dan wawancara kepada bidang pemasaran, pembimbing lapang serta karyawan di PT. Sari Melati Sejahtera sedangkan data sekunder diperoleh dengan cara mengutip dari buku, arsip, jurnal yang berkaitan dengan analisa penjualan the Jawa. Bagian pemasaran perusahaan melakukan strategi yang baik untuk dapat menggunakan kesempatan atau peluang yang ada dalam pemasaran. Penelitian ini bertujuan untuk mengetahui strategi produk, strategi harga, strategi distribusi, strategi promosi. Strategi pemasaran yang dilakukan dengan cara bersaing dengan para kompetitor. Strategi distribusi melalui pemasaran lokal meliputi wilayah sekitar Cirebon sedangkan pemasaran ekspor Asia Strategi promosi dilakukan dengan pengiklanan lewat media

elektronik dan mengikuti pameran dan menjalin kerjasama dengan Hotel, restoran dan cafe-cafe yang ada di Wilayah Cirebon dan sekitarnya.

Menurut Annisa Vini Fitria dkk, Indonesia ialah negara agraris dengan bidang pertanian yang dapat mencukupi kebutuhan pangan masyarakat. Bidang pertanian mampu menghasilkan bahan untuk makanan dan minuman yang diperlukan untuk memenuhi kebutuhan hidup masyarakat. Salah satu produk bidang pertanian berupa bahan minuman dibutuhkan guna mengatur proses metabolisme tubuh, menghilangkan haus dan mengatur agar tubuh tetap sehat. Salah satu macam minuman yang sering dikonsumsi masyarakat ialah teh. Tumbuhan teh (*Camellia sinensis*) berasal dari daerah subtropika, mampu tumbuh baik dengan kondisi sejuk dan suhu berkisar 13-25 oC, memerlukan sinar matahari yang cukup, dan kelembaban relatif lebih dari 70% saat siang hari (Fitria et al., 2022)

Penelitian yang dilakukan oleh Yosia Tutus Angriawan Jaya Diningrat, menyimpulkan Strategi produk atau pemasaran adalah untuk meningkatkan nilai jual dari Teh Manggata dengan cara meningkatkan pelayanan, menonjolkan pelayanan konsumen yang dimiliki dengan cara menghubungi konsumen terlebih dahulu, menambah mesin baru. Strategi harga adalah menjual Teh Manggata dalam kemasan yang lebih kecil. Strategi tempat adalah memperbanyak sub-distributor dan memberikan pelatihan kepada sub-distributor baru. Strategi promosi adalah melakukan endorse di media sosial, melakukan pengenalan produk ke masyarakat melalui seminar kesehatan, membuat acara talkshow, meningkatkan anggaran promosi untuk mendukung promosi di media sosial.(Diningrat, 2018)

Penelitian terdahulu yang dilakukan oleh: Yaya Asohi¹, Andri², dalam jurnalnya yang berjudul : Impelementasi algoritma regresi linier berganda untuk prediksi penjualan menyimpulkan bahwa : Prediksi penjualan barang menggunakan algoritma regresi linier dengan proses pengolahan data menggunakan tahapan dari Knowledge Discovery In Database (KDD), serta penggunaan tool rapidminer sebagai alat bantu dalam proses prediksi barang menghasilkan: a. Hasil yang diperoleh dari prediksi penjualan barang untuk priode waktu tahun 2020 sebesar 169715 barang. b. Dari proses pediksi penjualan berdasarkan jenis barang menghasilkan nilai penjualan barang tertinggi yaitu berjenis barang FORTUNE PP 1 L/24, yang memiliki angka prediksi penjualan barang sebesar 419,494 & menghasilkan nilai penjualan barang terendah untuk jenis barang TEH BOTOL KOTAK SOSRO 200 ML dengan angka 23,872 barang. c. Pengukuran tingkat error, diperoleh nilai dari Root Mean Square Error (RMSE) dengan nilai 40,476.(Yaya Asohi, 2020) Berdasarkan table diatas dapat penulis deskripsikan mulai dari satuan, harga adalah 1. Teh jawa premium 9 gr isi karton 50 dengan harga 370.000.00 per karton, 2. Teh Jawa Premium 45gr dengan harga 330.000,00per karton, 3. TEH JAWA BIRU 9gr dengan harga 315.000.00 per karton, 4. TEH JAWA BIRU 40gr dengan harga 285.000.00 per karton, 5. TEH CELUP JASMINE TEA @25 Dus dengan harga 400.000.00 per karton. 6. TEH CELUP JASMINE TEA @25 (AMPLOP) dengan harga 3.368.000,00 per karton, 7. TEH CELUP BLACK TEA @25 dengan harga 240.000.00 per karton, dst.

Berdasarkan latar belakang di atas, permasalahannya adalah Bagaimana mencari pengaruh parameter feature selection M5 Prime, Greedy, T-Test, dan Iterative T-Test terhadap nilai root mean squared error serta bagaimana mencari nilai root mean squared error paling optimal.

Fokus masalah dari penelitian ini adalah Selama ini belum tahu mencari pengaruh parameter feature selection M5 Prime, Greedy, T-Test, dan Iterative T-Test terhadap nilai root mean squared error. Berdasarkan permasalahan data dan latarbelakang yang telah diutarakan maka penulis membuat penelitian dengan judul “Analisa Dataset Penjualan Teh Menggunakan Algoritma Linear Regresi” dengan alasan Ingin mencari pengaruh parameter feature selection M5 Prime, Greedy, T-Test, dan Iterative T-Test terhadap nilai root mean squared error dan ingin mencari tahu, nilai root mean squared error paling optimal.

2. KERANGKA TEORI

2.1 Data Mining

Data mining adalah suatu rangkaian proses, maka data mining dapat dibagi menjadi beberapa tahap yaitu : (M. Hadi Prayitno¹, 2018). 1). Pembersihan data (membuang data yang tidak konsisten dan noise), 2). Integrasi data (penggabungan data dari beberapa sumber), 3). Transformasi data (mengubah data menjadi bentuk lain), 4). Aplikasi teknik DM, 5). Evaluasi dan Presentasi pengetahuan (dengan teknik visualisasi). Penelitian yang dilakukan oleh Yudi Herdiana¹), Wildan Adhitya Geraldine, Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemuka keteraturan, pola atau hubungan dalam kumpulan data berukuran besar(Yudi Herdiana¹), 2022).

2.2 Penjualan

Dalam penelitian yang telah dilakukan oleh Jayaprana dan Rahayu menjelaskan bahwa untuk meningkatkan penjualan, maka sebuah perusahaan harus memiliki informasi penting dalam meningkatkan

penjualannya, salah satu caranya adalah mengolah data transaksi yang menumpuk menjadi sebuah informasi yang berguna. Penggunaan algoritma apriori membantu pebisnis dalam proses pengambilan keputusan terhadap segala hal yang berhubungan dengan persediaan barang. Penelitian lain menyebutkan bahwa dengan menerapkan Algoritma Apriori, dapat membantu dalam membentuk kandidat kombinasi item yang mungkin, kemudian dilakukan pengujian apakah kombinasi tersebut memenuhi parameter support dan confidence minimum yang merupakan nilai minimal yang diberikan oleh pengguna (Nurajizah, 2019)

2.3 Algoritma Linear Regression

Linear regression adalah salah satu algoritma yang digunakan data science dan tergolong pada algoritma supervised learning. Algoritma ini menggunakan prinsip regresi. Regresi membuat model prediksi untuk target variabel berdasarkan dari variabel bebasnya. (Hamdanah & Fitriana, 2021). Jenis algoritma ini sering digunakan untuk mencari hubungan antara variabel-variabel yang ada dan prediksinya. Sehingga pada linear regression, ditunjukkan untuk melakukan prediksi pada variabel terikat (y) berdasarkan variabel bebas yang diberikan (x). Hasil akhirnya berupa hubungan linear antara variabel input (x) dengan variabel output (y) (Komputer et al., 2022).

3. METODE PENELITIAN

3.1 Metode penelitian

Metode penelitian yang digunakan yaitu menggunakan algoritma linear regression yaitu : Sebelum kita bicara jauh tentang linear regression & Machine Learning, mari kita pahami terlebih dahulu apa itu “regression” (regresi)? regresi merupakan salah satu metode dalam ilmu statistik yang sering digunakan dalam pengkajian ilmu ekonomi, keuangan, bisnis, dan disiplin ilmu lainnya. Tujuan dari regresi yaitu untuk menentukan seberapa kuat hubungan antara dependen variabel (Y-variable) dan independen variabel (X-variable). (Sukirman, 2022)

Terdapat dua cabang utama dari metode *regression*, yaitu: *simple/univariate linear regression* dan *multiple/multivariate linear regression*. Pada *simple linear regression* hanya terdapat satu X-variabel yang digunakan untuk memprediksi satu Y-variabel, contohnya, memprediksi hasil penjualan the jawa (Y-variable) berdasarkan tinggi rendahnya temperatur suhu udara (X-variable).

Sedangkan, pada *multiple linear regression*, dibutuhkan lebih dari satu X-variabel yang digunakan untuk mengkaji atau memprediksi sesuatu hal (Y-variable). Contohnya, memprediksi harga the karton (Y-variabel) berdasarkan beberapa fitur tertentu misalnya berdasarkan ukuran, kategori, arsitektur satuan, dan lainnya (multiple Y-variabel).

Dalam ilmu statistik, kita dapat menggunakan rumus di bawah ini untuk menghitung Model Regresi berdasarkan tipenya:

- Simple/univariate linear regression: $Y = A + BX + \epsilon$
- Multiple/multivariate linear regression: $Y = A + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_tX_t + \epsilon$

Keterangan :

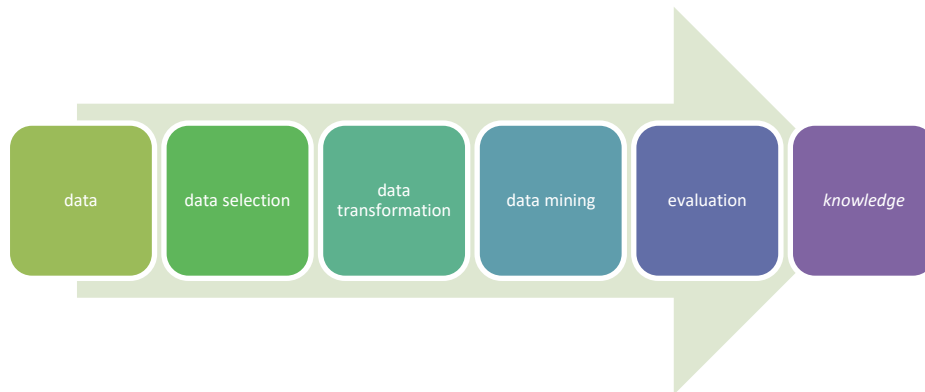
- Y = variabel yang ingin diprediksi (dependent variable).
- X = variabel yang digunakan untuk memprediksi Y (independent variable).
- A = intercept. Yaitu nilai rata-rata pada variabel Y apabila nilai pada variabel X bernilai 0.
- B = Slope. Yaitu nilai seberapa besar kontribusi yang diberikan oleh variabel X terhadap variabel Y.
- ϵ = error / residuals. Yaitu selisih antara nilai dua (Predictive value dinotasikan sebagai $\hat{Y}$) dengan nilai pengamatan sebenarnya (Y). Intinya error adalah semua hal yang mungkin mempengaruhi variabel Y tapi tidak diamati oleh peneliti (Ayuningsih et al., 2022)

3.2 Operasionalisasi Variabel

Operasional variabel yang akan dilakukan dalam penelitian ini dibagi menjadi 2 (dua) variabel yaitu terdiri dari variabel (X) : Analisa Dataset Penjualan The (Y) adalah Menggunakan Algoritma Linear Regresi.

3.3 Teknik Analisis Data

Adapun untuk menganalisis dataset rumah tidak layak huni dalam penerapan data mining ini menggunakan proses tahapan *Knowledge Discovery in Databases (KDD)* yang terdiri dari Data, Data Cleaning, Data transformation, Data mining, *Pattern evolution, knowledge*, seperti terlihat gambar 1:



Gambar 1. Teknik Analisis Data

4. HASIL

4.1 Hasil Penelitian

Hasil penelitian yang dilakukan dalam pembahasan ini yaitu akan menguraikan proses bagaimana menganalisa penjualan the Jawa menggunakan algoritma regresi linear. analisa ini dilakukan dengan proses pengujian menggunakan machine learning yaitu *Software RapidMiner Studio 9.12*

4.2 Data

Data diperoleh dari hasil transaksi penjualan teh di Pt Sari Melati Sejahtera dari bulan Januari 2023 sampai dengan bulan April 2023.

Dataset terdiri dari 7 (tujuh) atribut terdiri dari **No.**, **Nama Barang**, **Isi**, **Inner**, **Harga**, **Harga Baru**, dan **Bulan**. Untuk deskripsi dari masing-masing atribut sebagai berikut :

Tabel 1. Atribut

No.	Atribut	Deskripsi
1.	No.	Nomor urut dan menunjukkan banyaknya record pada dataset
2.	Nama Barang	Jenis teh yang ada dan dijual belikan
3.	Isi	Banyaknya jumlah teh yang ada dalam setiap paket
4.	Inner	Banyaknya jumlah teh yang ada dalam setiap karton atau bal
5.	Harga	Harga yang diperoleh dari pihak produsen
6.	Harga Baru	Harga yang dijual kepada konsumen
7.	Bulan	Nama bulan transaksi penjualan terjadi

Jumlah record atau data pada dataset sebanyak 128 record.

- Pada atribut Nama Barang, terdiri dari 32 jenis nama barang sebagai berikut :

Tabel 2. Data Set Teh Jawa

No.	Nama Barang
1	PAKET GIFTPACK AMPLOP (ISI 2)
2	PAKET GIFTPACK JASMINE CLP (ISI 2)
3	PAKET GIFTPACK JASMINE CLP (ISI 3)
4	PAKET GIFTPACK MIX (ISI 3)
5	TEH CELUP BLACK TEA @25 (AMPLOP)
6	TEH CELUP BLACK TEA @25 (Dus)
7	TEH CELUP BLACK TEA @5 (SASHET)
8	TEH CELUP BLACK TEA 100 (AMPLOP)
9	TEH CELUP GREEN TEA @25 (DUS)
10	TEH CELUP GREENTEA @25 (AMPLOP)
11	TEH CELUP JASMINE TEA @25 (AMPLOP)
12	TEH CELUP JASMINE TEA @25 (Dus)
13	TEH CELUP JASMINE TEA @5 (SASHET)

No.	Nama Barang
14	TEH CELUP VANILLA TEA @25 (DUS)
15	TEH JAWA BIRU 40gr
16	TEH JAWA BIRU 80gr
17	TEH JAWA BIRU 9gr (Dusunan)
18	TEH JAWA BIRU SPARASI 40gr
19	TEH JAWA GOLD 45gr
20	TEH JAWA GOLD 90gr
21	TEH JAWA GREENTEA CUP 180ml
22	TEH JAWA ORANGE 40gr
23	TEH JAWA ORANGE 80gr
24	TEH JAWA PREMIUM 45gr
25	TEH JAWA PREMIUM 90gr
26	TEH JAWA PREMIUM 9gr (Dusunan)
27	Teh Jawa Signature Fine Loose BLACKTEA 100gr
28	Teh Jawa Signature Fine Loose GREENTEA 100gr
29	Teh Jawa Signature Fine Loose JASMINE TEA 100gr
30	TEH MEDALI HIJAU 50 gr
31	TEH SEPATU MERAH / SPM 50 gr
32	TEH SEPATU TERBANG / OST 50 gr

- Pada atribut Isi, terdiri dari 7 jenis isi sebagai berikut :

Tabel 3. Atribut

No.	Isi
1	10
2	12
3	24
4	40
5	50
6	100
7	200

- Pada attribut Inner, terdiri dari 9 jenis inner sebagai berikut :

Tabel 4. Ettribut Inner

No.	Inner
1	1 BAL = 10 Pak / 1 Pak = 10 Pcs
2	1 BAL = 10 Pak / 1 Pak = 5 Pcs
3	1 Karton = 10 Paket
4	1 Karton = 12 Pcs
5	1 Karton = 20 Pcs
6	1 Karton = 200 Pcs
7	1 Karton = 24 Pcs
8	1 Karton = 40 Pcs
9	1 Karton = 50 Pcs

- Pada atribut Harga , harga dipengaruhi dari nama barang sebagai berikut :

Tabel 5. Atribut Harga

No.	Nama Barang	Harga
1	PAKET GIFTPACK AMPLOP (ISI 2)	21700
2	PAKET GIFTPACK JASMINE CLP (ISI 2)	16600
3	PAKET GIFTPACK JASMINE CLP (ISI 3)	23900
4	PAKET GIFTPACK MIX (ISI 3)	24100
5	TEH CELUP BLACK TEA @25 (AMPLOP)	6200
6	TEH CELUP BLACK TEA @25 (Dus)	4800
7	TEH CELUP BLACK TEA @5 (SASHET)	800
8	TEH CELUP BLACK TEA 100 (AMPLOP)	18000
9	TEH CELUP GREEN TEA @25 (DUS)	10000
10	TEH CELUP GREENTEA @25 (AMPLOP)	13000
11	TEH CELUP JASMINE TEA @25 (AMPLOP)	8700
12	TEH CELUP JASMINE TEA @25 (Dus)	7300
13	TEH CELUP JASMINE TEA @5 (SASHET)	1500
14	TEH CELUP VANILLA TEA @25 (DUS)	5000
15	TEH JAWA BIRU 40gr	2700
16	TEH JAWA BIRU 80gr	5400
17	TEH JAWA BIRU 9gr (Dusunan)	6000
18	TEH JAWA BIRU SPARASI 40gr	2200
19	TEH JAWA GOLD 45gr	3700
20	TEH JAWA GOLD 90gr	7400
21	TEH JAWA GREENTEA CUP 180ml	19000
22	TEH JAWA ORANGE 40gr	2300
23	TEH JAWA ORANGE 80gr	4600
24	TEH JAWA PREMIUM 45gr	3100
25	TEH JAWA PREMIUM 90gr	6200
26	TEH JAWA PREMIUM 9gr (Dusunan)	7000
27	Teh Jawa Signature Fine Loose BLACKTEA 100gr	63000
28	Teh Jawa Signature Fine Loose GREENTEA 100gr	65000
29	Teh Jawa Signature Fine Loose JASMINE TEA 100gr	67000
30	TEH MEDALI HIJAU 50 gr	1750
31	TEH SEPATU MERAH / SPM 50 gr	1950
32	TEH SEPATU TERBANG / OST 50 gr	2270

- Sedangkan pada atribut Harga Baru, merupakan atribut bebas yang terpengaruhi dari atribut lainnya.
- Pada atribut Bulan, terdiri dari 4 jenis bulan yaitu Januari, Februari, Maret dan April.

Pada Rapidminer, untuk membaca file excel dapat menggunakan operator **Read Excel** seperti tampak pada gambar dibawah ini.



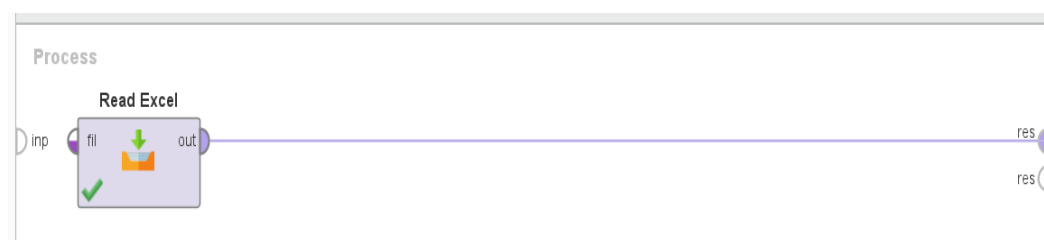
Gambar 2. Read excel

Parameter yang digunakan pada operator **Read Excel** di Rapidminer versi 10.1 adalah default. Hasil dari penggunaan operator Read Excel diperoleh informasi statistik sebagai berikut.

Tabel 6. Tabel operator Read Excel

No.	Atribut	Type	Missing	Jenis Atribut
1.	No.	Integer	0	Reguler
2.	Nama Barang	Nominal	0	Reguler
3.	Isi	Integer	0	Reguler
4.	Inner	Nominal	0	Reguler
5.	Harga	Integer	0	Reguler
6.	Harga Baru	Integer	0	Reguler
7.	Bulan	Nominal	0	Reguler

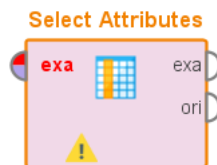
Untuk model proses sampai dengan proses Data tampak pada gambar di bawah ini.



Gambar 3. Operator read excel

1. Selection

Untuk memilih atribut yang akan digunakan, dapat menggunakan operator **Select Attribute** seperti tampak pada gambar di bawah ini.



Gambar 4. Select Attribute

Parameter yang digunakan pada operator **Select Attribute** di Rapidminer versi 10.1 sebagai berikut.

Tabel 7. Tabel Hasil Select Attribute

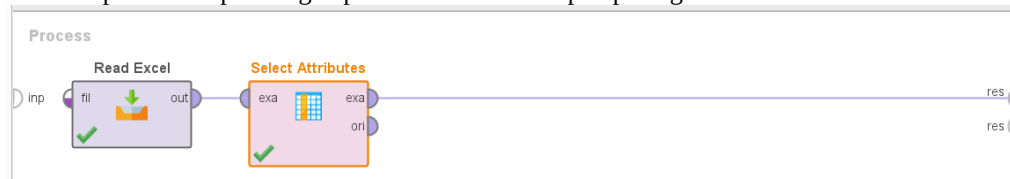
No.	Parameter	Value
1.	type	include attributes
2.	attribute filter type	a subset
3.	select subset	Select attribute:
		Bulan
		Harga
		Harga Baru
		Inner
		Isi
		Nama Barang

Hasil dari penggunaan operator **Select Attribute** diperoleh informasi statistic sebagai berikut.

Tabel 8. Hasil Statistic

No.	Atribut	Type	Missing	Jenis Atribut
1.	Nama Barang	Nominal	0	Reguler
2.	Isi	Integer	0	Reguler
3.	Inner	Nominal	0	Reguler
4.	Harga	Integer	0	Reguler
5.	Harga Baru	Integer	0	Reguler
6.	Bulan	Nominal	0	Reguler

Untuk model proses sampai dengan proses **Selection** tampak pada gambar di bawah ini.



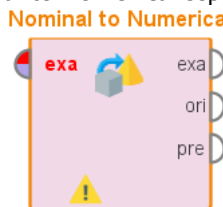
Gambar 5. Proses selection

2. Preprocessing

Sesuai dengan informasi dari hasil proses sebelumnya bahwa tidak ada nilai **missing**, maka proses **Preprocessing** tidak dilakukan.

3. Transformation

Pada proses Transformation dilakukan dengan mengubah type data Nominal menjadi Numeric dengan menggunakan operator **Nominal to Numerical** seperti tampak pada gambar di bawah ini.



Gambar 6. Nominal to Numerical

Parameter yang digunakan pada operator Nominal to Numerical di Rapidminer versi 10.1 tampak sebagai berikut.

Tabel 9. Hasil Nominal to Numerical

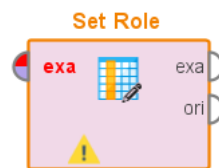
No.	Parameter	Value
1.	attribute filter type	subset
2.	attributes	Select attribute:
		Bulan
		Inner
		Harga
		Harga Baru
3.	coding type	unique_integers

Hasil dari penggunaan operator **Nominal to Numerical** diperoleh informasi statistik sebagai berikut.

Tabel 10. Hasil Nominal to Numerical

No.	Atribut	Type	Missing	Jenis Atribut
1.	Nama Barang	Polynomial	0	Reguler
2.	Isi	Integer	0	Reguler
3.	Inner	Numeric	0	Reguler
4.	Harga	Integer	0	Reguler
5.	Harga Baru	Integer	0	Reguler
6.	Bulan	Numeric	0	Reguler

Untuk menentukan id dan label digunakan operator **Set Role** seperti tampak pada gambar di bawah ini.



Gambar 7. Operator Set Role

Parameter yang digunakan pada operator **Set Role** di Rapidminer versi 10.1 tampak sebagai berikut.

Tabel 11. Hasil Operator Set Role

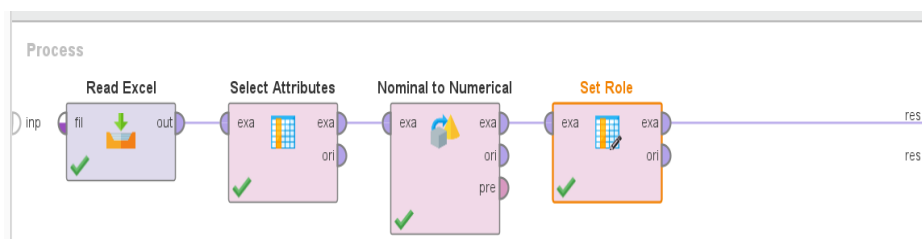
No.	Parameter	Value	
1.	set roles	Edit List	
		<i>Attribute name</i>	<i>Target role</i>
		NAMA BARANG	id
		HARGA BARU	Label

Hasil dari penggunaan operator **Set Role** diperoleh informasi statistik sebagai berikut.

Tabel 12. Hasil Operator Set Role

No.	Atribut	Type	Missing	Jenis Atribut
1.	Nama Barang	Nominal	0	Special (id)
2.	Isi	Integer	0	Reguler
3.	Inner	Real	0	Reguler
4.	Harga	Integer	0	Reguler
5.	Harga Baru	Integer	0	Special (label)
6.	Bulan	Real	0	Reguler

Untuk model proses sampai dengan proses **Tranformation** tampak pada gambar di bawah ini.



Gambar 8. Proses Tranformation

5. Data mining

Pada proses Datamining menggunakan algoritma Linear Regresi, perlu dilakukan pembagian dataset dulu untuk menjadi data training dan data test. Untuk melakukan tersebut menggunakan operator **Split Data** seperti tampak pada gambar di bawah ini.



Gambar 9. Operator Split Data

Parameter yang digunakan pada operator **Split Data** di Rapidminer versi 10.1 tampak sebagai berikut.

Tabel 13. Parameter Split Data

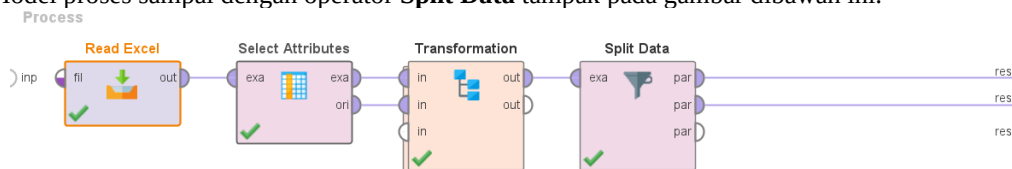
No.	Parameter	Value
1.	Partitions	Edit Enumeration
		<i>Ratio</i> : 0,8 dan 0,2
2.	sampling type	automatic

Hasil dari penggunaan operator **Split Data** diperoleh informasi sebagai berikut.

Tabel 14. Hasil Split Data

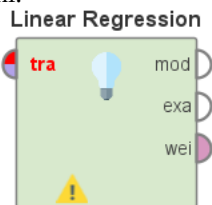
No.	Ratio	Jumlah Record	Special Atribut	Reguler Atribut
1.	0,8	102	2	4
2.	0,2	26	2	4

Model proses sampai dengan operator **Split Data** tampak pada gambar dibawah ini.



Gambar 10. Operator Split Data

Setelah menggunakan operator **Split Data**, dilanjutkan menggunakan operator **Linear Regression** seperti tampak pada gambar di bawah ini.



Gambar 11. Linear Regression

Parameter yang digunakan pada operator **Linear Regression** di Rapidminer versi 10.1 tampak sebagai berikut.

Tabel 15. Parametr Linear Regression

No.	Parameter	Value
1.	feature selection	none
		M5 prime
		greedy
		T-Test
		Iterative T-Test

Hasil dari penggunaan operator **Linear Regression** dengan parameter feature selection tersebut di atas diperoleh informasi sebagai berikut.

a. Linear Regression.weight

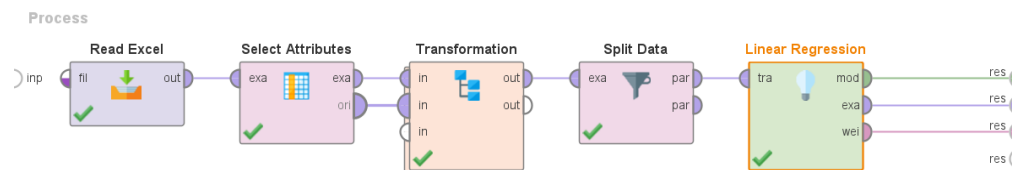
Attribute	Weight
INNER	-255.103
BULAN	-472.072
ISI	-11.199
HARGA	0.992

b. Linear Regression.model

Tabel 16. Linear Regression.model

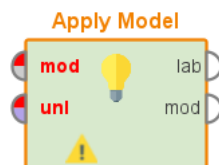
Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value
INNER	-255.103	36.423	-0.042	0.976	-7.004	3.278E-10
BULAN	-472.072	95.594	-0.028	1.000	-4.938	3.282E-6
ISI	-11.199	2.619	-0.031	0.667	-4.277	4.444E-5
HARGA	0.992	0.007	0.986	0.608	137.424	0.0
(Intercept)	4054.370	307.194	NaN	NaN	13.198	0.0

Model proses sampai dengan operator **Linear Regression** tampak pada gambar dibawah ini.



Gambar 12. Proses Linear Regression

Model yang dihasilkan oleh operator **Linear Regression** akan diimplementasikan dengan menggunakan operator **Apply Model** seperti tampak pada gambar di bawah ini.



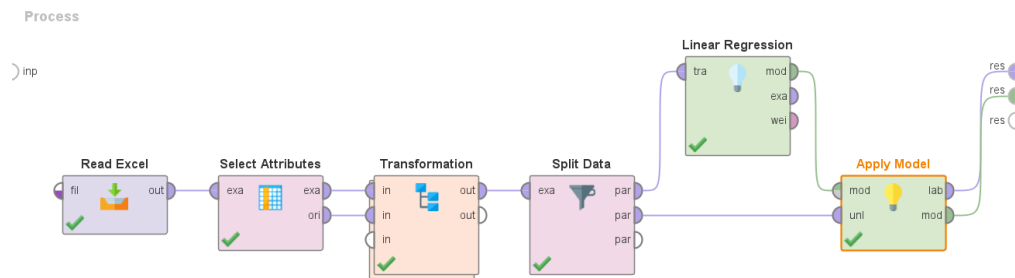
Gambar 13. Apply Model

Parameter yang digunakan secara default pada operator **Apply Model** di Rapidminer versi 10.1. Hasil dari penggunaan operator **Apply model** diperoleh informasi sebagai berikut.

Tabel 17. Hasil Apply model

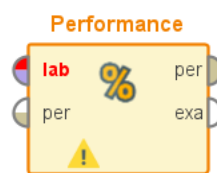
No.	Nama Barang	Harga Barang	Prediction	Inner	Bulan	Isi	Harga
1	PAKET GIFTPACK AMPLOP (ISI 2)	30275	25465.23	0	0	10	21700
2	PAKET GIFTPACK JASMINE CLP (ISI 2)	21150	20406.86	0	0	10	16600
3	PAKET GIFTPACK JASMINE CLP (ISI 2)	17600	19462.72	0	2	10	16600
4	PAKET GIFTPACK JASMINE CLP (ISI 3)	25400	26231.05	0	3	10	23900
5	PAKET GIFTPACK MIX (ISI 3)	30800	27845.64	0	0	10	24100
6	TEH CELUP BLACK TEA @25 (Dus)	6190	7745.03	2	0	50	4800
7	TEH CELUP BLACK TEA @25 (Dus)	6025	7272.95	2	1	50	4800
8	TEH CELUP BLACK TEA 100 (AMPLOP)	19800	19202.02	4	3	24	18000
9	TEH CELUP GREEN TEA @25 (DUS)	12520	11665.19	5	1	50	10000
10	TEH CELUP GREENTEA @25 (AMPLOP)	14300	15301.03	1	2	40	13000
11	TEH CELUP JASMINE TEA @25 (AMPLOP)	11170	11980.28	1	0	40	8700
12	TEH CELUP JASMINE TEA @25 (AMPLOP)	9570	11036.14	1	2	40	8700
13	TEH CELUP VANILLA TEA @25 (DUS)	6425	7943.39	2	0	50	5000
14	TEH JAWA BIRU 40gr	2970	2410.48	7	3	100	2700
15	TEH JAWA BIRU 9gr (Dusunan)	7570	7404.61	8	0	50	6000
16	TEH JAWA BIRU 9gr (Dusunan)	6600	6460.47	8	2	50	6000
17	TEH JAWA BIRU 9gr (Dusunan)	6600	5988.39	8	3	50	6000
18	TEH JAWA GOLD 45gr	4565	4818.54	7	0	100	3700
19	TEH JAWA GOLD 45gr	4070	3874.39	7	2	100	3700
20	TEH JAWA GREENTEA CUP 180ml	20000	20665.93	4	2	24	19000
21	TEH JAWA GREENTEA CUP 180ml	20000	20193.85	4	3	24	19000
22	TEH JAWA ORANGE 80gr	5060	4599.82	8	3	50	4600
23	TEH JAWA PREMIUM 45gr	3410	3279.29	7	2	100	3100
24	TEH JAWA PREMIUM 9gr (Dusunan)	7700	6980.23	8	3	50	7000
25	Teh Jawa Signature Fine Loose BLACKTEA 100gr	63000	62693.53	9	3	12	63000
26	Teh Jawa Signature Fine Loose GREENTEA 100gr	65000	66093.42	9	0	12	65000

Model proses sampai dengan operator **Apply Model** tampak pada gambar dibawah ini.



Gambar 14. Proses Apply Model

Untuk mengukur besarnya performan dari model yang dibuat maka menggunakan operator Performance (Regression) seperti tampak pada gambar di bawah ini.



Gambar 2Performance (Regression)

Parameter yang digunakan pada operator **Performance (Regression)** di Rapidminer versi 10.1 tampak sebagai berikut.

Tabel 18. Parameter Performance (Regression)

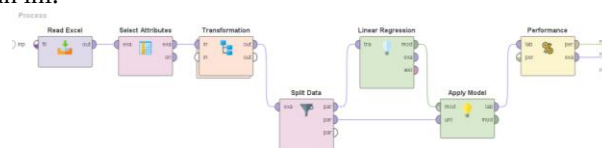
No.	Parameter	Value
1.	main criterion	root_mean_squared_error

Hasil dari penggunaan operator **Performance (Regression)** diperoleh informasi sebagai berikut.

root_mean_squared_error

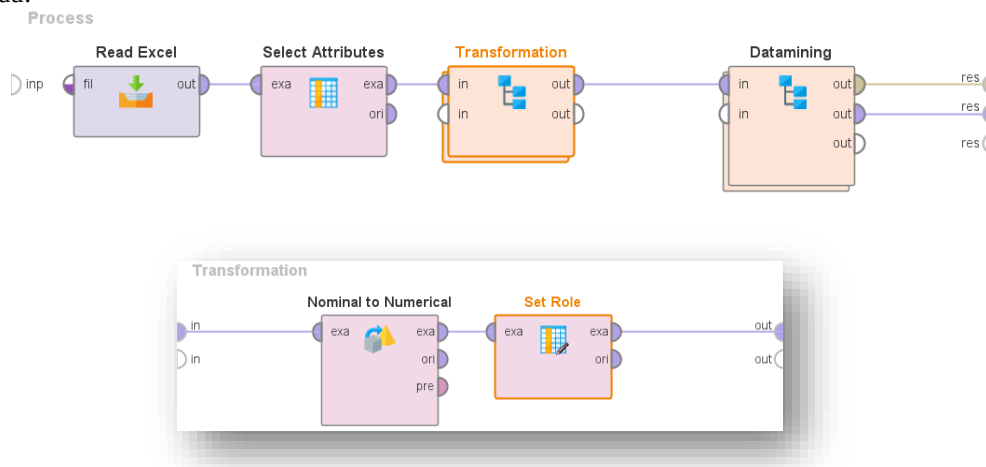
root_mean_squared_error: 1402.007 +/- 0.000

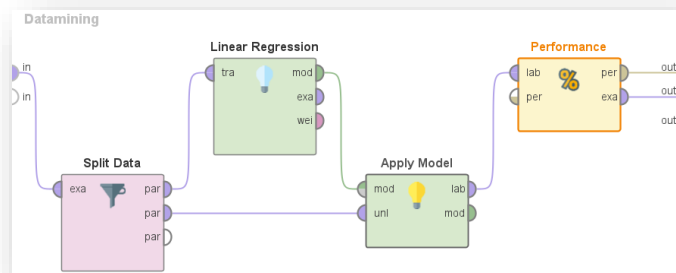
Model proses sampai dengan operator **Performance (Regression)** atau proses Datamining tampak pada gambar dibawah ini.



Gambar 36. Operator Performance (Regression)

Atau:





Gambar 17. Operator Performance (Regression)

6. Evaluation

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola pola yang khas maupun model dievaluasi untuk menilai apakah memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai maka ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki data mining lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat. Alur yang dihasilkan dari data mining adalah alur informasi yang dihasilkan dari proses dan perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Proses ini merupakan proses KDD yang mencakup pemeriksaan, apakah alur atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada pada evaluasi sebelumnya.

5. KESIMPULAN

Berdasarkan hasil penelitian pengelompokkan Dataset penjualan the Jawa menggunakan algoritma Linear regression, maka hasilnya dapat disimpulkan sebagai berikut : Dapat mengetahui pengaruh parameter feature selection M5 Prime, Greedy, T-Test, dan Iterative T-Test terhadap nilai root mean squared error dengan menerapkan data mining menggunakan algoritma linear regression, Dapat mengetahui nilai root mean squared error yang paling optimal penjualan the Jawa terbaik berdasarkan Nilai Davies Bouldin Index yang dihasilkan dari algoritma linear regression ini sebesar 2,079. Dengan menghitung jarak antara rata-rata cluster diperoleh bahwa cluster 0 merupakan kelompok terbaik dengan jarak terdekat sebesar 9,728.

DAFTAR PUSTAKA

- Ayuningsih, A. P., Setiawan, N. Y., & Wijoyo, S. H. (2022). Analisis Prediksi Penjualan Obat Hewan menggunakan Metode Regresi Linier melalui Visualisasi Dashboard (Studi Kasus PT . Satwa Jawa Jaya). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(4), 1568–1575.
- Diningrat, Y. T. A. J. (2018). Strategi pemasaran teh manggata pada cv dari teman sejati mojokerto. *AGORA*, 6(2).
- Fitria, A. V., Setiyani, R., & Lestari, R. D. (2022). Analisis Tipe Perilaku Konsumen dalam Membeli Teh Celup. *Agricultural Socio-Economic Empowerment and Agribusiness Journa*, 1, 49–58.
- Hamdanah, F. H., & Fitriana, D. (2021). ANALISIS PERFORMANSI ALGORITMA LINEAR REGRESSION DENGAN GENERALIZED LINEAR MODEL UNTUK PREDIKSI PENJUALAN PADA USAHA MIKRO KECIL DAN MENENGAH. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, 10, 23–32.
- Haryawan, T. (2020). Analisa Pola Pendistribusian Produk Dengan Algoritma C5 . 0 (Bidang Studi : Teh Botol Sosro). *Pelita Informatika : Informasi Dan Informatika*, 9, 61–66.
- Komputer, J. S., Syakir, Y., Hermanto, T. I., Ramadhan, Y. R., Informatika, S. T., Tinggi, S., & Wastukencana, T. (2022). Analisis Marketplace Shopee Untuk Memprediksi Penjualan dengan Algoritma Regresi Linier. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(September), 904–915.
- Lestari, S. (2023). Analisis Algoritma Regresi Linear Sederhana dalam Memprediksi Tingkat Penjualan Album KPOP. *INSOLOGI: Jurnal Sains Dan Teknologi*, 2(1), 199–209. <https://doi.org/10.55123/insologi.v2i1.1692>
- M. Hadi Prayitno1, R. (2018). Analisis Penjualan Produk Retail Dengan Metode Data Mining Asosiasi. *Jurnal Kajian Ilmiah*, 18(3), 231–237.
- Nurajizah, S. (2019). Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori. *Jurnal Invotek Polbeng*, d, 35–44.
- Riyanto, A., & Susanti, M. (2020). Analisa Data Mining Terhadap Penjualan Food Dengan Metode Apriori Pada Kopsyahira. *Bianglala Informatika*, 8(1), 44–48.
- Sukirman, S. (2022). ANALISIS CLOUD COMPUTING UNTUK PENYIMPANAN DOKUMEN TERHADAP PROSES PEMBELAJARAN MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA (STUDI KASUS : SMA

Jurnal Teknologi Ilmu Komputer

Vol 2, No 1, Desember 2023, Hal. 66-79

ISSN 2985-3486 (Media Online)

DOI 10.56854/jtik.v2i2.81

<https://ejurnal.bangunharapanbangsa.id/index.php/jtik>

CHANDRA KUSUMA JAKARTA UTARA). *Jurnal Ilmiah Informatika (Scientific Informatics Journal)* with CC BY NC Licence, 7(1), 1–12.

Tida, A. M., & Zubair, A. (2022). ANALISA TRANSAKSI PENJUALAN MENGGUNAKAN METODE ALGORITMA APRIORI (STUDI KASUS TOKO BUSANA SAKATO ,GAJAH MADA PLAZA B-19 MALANG). *Jurnal Teknik Informatika Kaputama (JTIK)* Vol.6,, 6(1).

Yaya Asohi, A. (2020). Impelementasi algoritma regresi linier berganda untuk prediksi penjualan. *Jurnal Nasional Ilmu Komputer*, 1(3), 149–158.

Yudi Herdiana 1), W. A. G. (2022). PENERAPAN MACHINE LEARNING DENGAN MODEL LINEAR REGRESSION TERHADAP ANALISIS KUALITAS HASIL PETIK THE DI PT . PERKEBUNAN. *Jurnal Informatika – COMPUTING*, 09, 1–9.