

Analisa Dataset Penjualan Teh Menggunakan Algoritma Linear Regresi

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

Program Studi Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer IKMI Cirebon,
Indonesia

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

Keywords: *analisa, penjualan, the jawa, 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 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.

2. KERANGKA TEORI

2.1 Sub Title 1

Berisi penjelasan tentang tahapan penelitian yang menggambarkan urutan logis untuk mendapatkan hasil penelitian sesuai dengan harapan dan gambaran sistem. Jika ada gambar dan tabel, itu harus disajikan dengan nama tabel dan gambar yang disertai dengan nomor urut.

2.2. Sub Title 2

Dalam naskah, nomor kutipan secara berurutan dalam tanda kurung siku [3], juga tabel angka dan angka secara berurutan seperti yang ditunjukkan pada Tabel 1 dan Gambar 1.

Tabel 1. Jenis jenis database

Nama	Nomor	Field
MySQL	10	100
Oracle	15	130
Access	20	400

(Judul tabel center, tidak ada garis vertical)

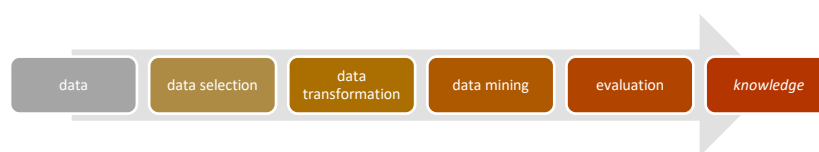
Gunakan kualitas yang tinggi untuk gambar yang dipakai. Rumus menggunakan Editor Rumus yang terdapat pada Microsoft Word. Penomoran rumus di buat berurut berdasarkan urutan rumus yang terdapat pada artikel, dan penulisannya seperti (1).

$$\int_0^{\infty} \exp(-\lambda |z_j - z_i|) \quad (1)$$

$$\int_0^{\infty} \exp(-\lambda |z_j - z_i|) \lambda^{-1} J_1(\lambda r_2) J_0(\lambda r_i) d\lambda \quad (2)$$

3. METODE PENELITIAN

Adapun untuk menganalisis data 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. *Knowledge Discovery in Databases (KDD)*

1. Database
Data capaian nilai ujian nasional bersumber dari repository Pusat Penilaian Pendidikan Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Data yang digunakan pada penelitian ini ialah sebanyak 111 dataset.
2. Data Selection & Cleaning
Pada umumnya, data yang diperoleh, baik dari database maupun survey, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performansi dari system data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.
3. Data transformation
Data Transformation merupakan langkah mengubah dataset menjadi satu format yang sama untuk mempermudah proses data. Tahap ini adalah proses transformasi pada data yang telah dipilih sehingga data tersebut sesuai untuk proses data mining.
4. Data mining
Tahap ini adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses knowledge discovery in databases (KDD) secara keseluruhan yang dilakukan untuk menganalisis data yang telah dibersihkan.
5. Interpretasi & evaluation
Evaluation ini menghasilkan pola-pola dari proses evaluasi guna memperkirakan tujuan yang diinginkan. Tahap ini juga menggunakan tools RapidMiner untuk menentukan nilai hasil evaluasi dari algoritma k-means. Hasil dari evaluasi adalah:
 - a. Melakukan import data, data yang digunakan dataset nilai format excel
 - b. Memanggil operator set role, yaitu untuk menentukan atribut name dan target role
 - c. Melakukan select atribut, yaitu atribut apa saja yang akan digunakan
 - d. Memanggil operator nominal to Numerical yaitu merubah variable menjadi numeric
 - e. Melakukan clustering yaitu untuk menentukan cluster (k) dan max
 - f. Performance untuk menampilkan hasilnya dari data mining secara keseluruhan performance vector dan Davies Bouldin indeks.

4. HASIL

4.1 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 2. 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

4.2 Preprocessing

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

4.3 Transformation

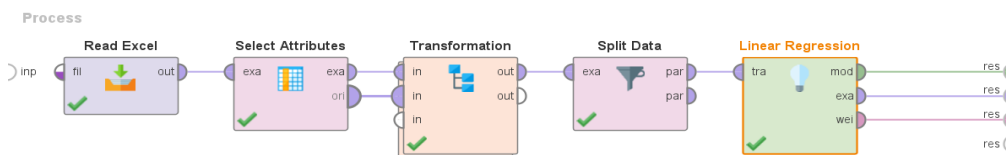
Pada proses Transformation dilakukan dengan mengubah type data Nominal menjadi Numeric dengan menggunakan operator Nominal to Numerical seperti tampak pada tabel di bawah ini.

Tabel 3. Hasil Nominal to Numerical

No.	Atribut	Type	Missing	Jenis Atribut
1.	Nama Barang	Polynominal	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

4.4 Data mining

Pada proses Datamining menggunakan algoritma Linear Regresi, perlu dilakukan pembagian dataset dulu untuk menjadi data training dan data test. Model proses sampai dengan operator **Linear Regression** tampak pada gambar dibawah ini.



Gambar 2. Proses Linear Regression

Model yang dihasilkan oleh operator Linear Regression akan diimplementasikan dengan menggunakan operator Apply Model sehingga kita mendapatkan hasil apply model pada tabel dibawah ini:

Tabel 4. Tabel 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

4.5 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 regresi, maka hasilnya dapat disimpulkan sebagai berikut :

1. 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.
2. 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 Journal*, 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 Prayitno, 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*, 4, 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 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.