

ALGORITMA REGRESI LINIER UNTUK MEMPREDIKSI PENGARUH PROMOSI TERHADAP PENJUALAN PRODUK K31 S-2 DENGAN *FRAMEWORKS* CRISP-DM

Ipang Sasono¹⁾, Ria Hartati²⁾, Gusli Chidir³⁾, Windu Gata⁴⁾

^{1,2,3}Dosen Tetap Universitas Insan Pembangunan Indonesia

⁴Universitas Nusa Mandiri

*Penulis Korespondensi: ipangsasono@gmail.com, ria.hartati@gmail.com

ABSTRAK

Tujuan dari penelitian ini untuk mengetahui penggunaan budget promosi yang terukur dan menghasilkan penjualan pada produk K31 S-2 yang optimal dan apakah pengeluaran budget saat ini sudah sesuai atau belum. Penelitian ini menggunakan perusahaan dibidang manufaktur dengan tujuan untuk mendapatkan informasi yang akurat mengenai prediksi pengaruh promosi terhadap penjualan dengan menggunakan *frameworks* CRISP-DM. Hasil penelitian didapatkan nilai coefficients sebesar 9.61652512 dan intercept 123757.37057197 dengan nilai R² adalah 0.93 yang sehingga model tersebut fit dengan data yang ada yang menghasilkan budget promosi lebih besar dari yang seharusnya sehingga omset yang didapatkan tidak sesuai dengan budget yang dikeluarkan, sehingga budget yang dikeluarkan lebih besar 17% dari yang seharusnya.

Keywords: Promosi, Penjualan Produk K31 S-2, Algoritma, CRISP-DM

PENDAHULUAN

Dalam perkembangan bisnis yang terus berkembang pesat saat ini, peran pengambilan keputusan berbasis data telah menjadi sangat penting dalam mencapai hasil yang optimal. Hal ini terutama berlaku dalam bidang pemasaran dan penjualan, di mana efektifitas dari kegiatan promosi secara langsung mempengaruhi penjualan produk. Namun, banyak organisasi masih mengandalkan metode intuitif dalam alokasi anggaran promosi dan kurang memiliki pendekatan yang terstruktur dalam mencapai tujuan penjualan. Dalam perkembangan bisnis saat ini IT menjadi sangat penting untuk memanfaatkan kekuatan data dan analisis lanjutan guna membuat keputusan yang didasarkan pada data dan menghasilkan hasil yang terukur.

Tujuan utama dari proyek ini adalah untuk mengoptimalkan alokasi anggaran promosi guna mencapai penjualan optimal untuk Produk K31 S-2. Pendekatan yang ada saat ini, yang mengandalkan intuisi dan target penjualan yang luas, tidak cukup memberikan hubungan yang jelas antara kegiatan promosi dan hasil penjualan. Dengan menggunakan teknik berbasis data, khususnya menggunakan algoritma Regresi Linear, tujuan kami adalah

untuk membentuk hubungan yang dapat diukur antara upaya promosi dan penjualan, yang akan menghasilkan strategi alokasi anggaran yang efektif dan berdasarkan data.

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkan menjadi informasi – informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field dari sebuah relasional database yang besar. Data mining sebuah alat untuk mencari hubungan dari suatu data ke data lainnya untuk menghasilkan data yang baru yang mempunyai arti dan bermanfaat bagi orang yang melakukan pengolahan data.

Dalam penelitian ini menggunakan *frameworks* CRISP-DM salah satu pendekatan yang dibutuhkan oleh industry. Metode yang tepat perlu digunakan untuk melakukan tahapan penelitian ini. CRISP-DM menyediakan standar proses baku untuk data mining yang dapat diterapkan ke dalam strategi pemecahan masalah umum pada bisnis atau pada unit penelitian. CRISP-DM membandingkan metodologi data mining lain

lebih lengkap dan terdokumentasi dengan baik. Setiap fase terstruktur dan terdefinisi dengan jelas sehingga mudah diaplikasikan bahkan bagi pemula sekalipun.

Penggunaan CRISP-DM dalam penelitian ini untuk mempermudah dalam menentukan alur kerja, melihat dari sisi bisnis sampai kedalam data dan menghasilkan penelitian yang akurat karena dapat dilihat dari berbagai aspek. Dengan metode CRISP-DM ini diharapkan dapat memberikan gambaran yang jelas untuk mendapatkan bisnis goal dalam mengetahui business understanding dan data understanding.

METODOLOGI PENELITIAN

Metodologi penelitian ini menggunakan metode *Cross Industry Standart Process Model for Data Mining*. Berikut gambaran alur proses yang terjadi pada CRISP-DM



Gambar 1. Framework CRISP-DM

Kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) memberikan pendekatan yang terstruktur dalam analisis data, memastikan bahwa prosesnya bersifat sistematis, dapat diulang, dan sejalan dengan tujuan bisnis. Kerangka kerja ini terdiri dari langkah-langkah berikut:

Pemahaman Bisnis

Langkah pertama dari kerangka kerja CRISP-DM melibatkan pemahaman yang mendalam terhadap masalah bisnis dan tujuan yang ingin dicapai. Ini meliputi identifikasi stakeholder utama, klarifikasi tujuan bisnis, dan definisi masalah yang perlu dipecahkan. Dalam kasus kami, tujuannya adalah untuk mengoptimalkan alokasi anggaran promosi

untuk Produk K31 S-2 guna mencapai penjualan yang optimal.

Pemahaman Data

Dalam fase ini, kami mengumpulkan dan mengeksplorasi sumber data yang relevan, termasuk data promosi dan data penjualan selama tahun terakhir. Data ini akan sangat penting dalam mengembangkan model prediktif. Dengan menganalisis data, kami bertujuan untuk mengidentifikasi pola, tren, dan tantangan potensial yang dapat memengaruhi akurasi model.

Persiapan Data

Persiapan data melibatkan pembersihan, transformasi, dan strukturisasi data agar sesuai untuk analisis. Langkah ini termasuk penanganan nilai yang hilang, pencilan, dan memastikan bahwa data berada dalam format yang kompatibel dengan algoritma yang dipilih (Regresi Linear, dalam kasus kami). Kualitas data secara langsung mempengaruhi akurasi model.

Pemodelan

Fase pemodelan melibatkan pemilihan dan penerapan algoritma yang sesuai pada data yang sudah dipersiapkan. Dalam proyek kami, kami akan menggunakan algoritma Regresi Linear untuk membentuk hubungan antara upaya promosi dan penjualan. Hal ini akan memungkinkan kami untuk membuat prediksi tentang penjualan di masa depan berdasarkan anggaran promosi yang diberikan. Dan juga untuk mengukur apakah langkah promosi yang dilakukan sudah optimal atau belum optimal.

Evaluasi

Setelah model dikembangkan, perlu dievaluasi untuk mengukur akurasi dan efektivitasnya. Mean Absolute Error (MAE), Mean Squared Error (MSE), dan skor R-squared (R2) adalah metrik umum yang digunakan untuk mengevaluasi model regresi. Metrik-metrik ini memberikan wawasan tentang seberapa baik model cocok dengan data dan membuat prediksi yang akurat.

Implementasi

Setelah model dievaluasi, langkah selanjutnya adalah mengimplementasikannya untuk penggunaan praktis dalam organisasi. Ini bisa melibatkan integrasi model ke dalam sistem

yang ada atau membuat antarmuka yang mudah digunakan bagi para pengambil keputusan untuk memasukkan anggaran promosi dan menerima prediksi penjualan.

HASIL DAN PEMBAHASAN

Tabel 1. Data Promosi dan Penjualan

K31	Jan			Feb			Mar			Apr			Mei			Jun			Jul			Agt			Sep			Okt		
	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val	Promo	Qty	Val
Banten	1,500	156	2,421,900	1,500	492	7,638,300	1,500	288	4,782,240	2,000	372	6,177,060	3,000	276	3,861,300	5,053,455	2,000	324	5,269,320	4,000	455	7,378,155	4,000	-	624	10,228,680	4,000	432	7,106,940	
Jabar	1,500	2,335	36,575,181	1,500	2,310	36,274,170	3,000	1,560	26,257,005	2,500	1,647	28,197,372	2,000	2,020	33,960,319	26,924,930	2,000	1,542	25,901,035	1,500	2,099	35,156,476	1,500	-	2,249	37,753,219	3,000	1,333	22,444,896	
Jabar	2,800	1,912	33,476,040	2,800	1,764	31,581,300	3,000	2,653	49,099,326	3,500	2,944	56,192,989	3,500	3,665	69,965,072	32,213,569	2,500	1,736	29,615,582	2,500	1,785	30,429,581	2,500	-	1,838	31,317,030	3,500	1,561	26,645,490	
Jabar	1,200	226	3,839,010	1,200	214	3,396,180	2,000	341	6,078,900	3,000	553	10,070,019	2,500	297	5,374,096	7,511,240	1,500	299	5,075,236	1,500	408	6,942,366	1,500	-	265	4,532,658	3,000	318	5,421,248	
Jabar	2,000	561	9,014,333	2,000	822	13,039,620	3,000	1,170	19,853,580	3,500	1,577	26,763,570	3,500	1,476	25,049,196	9,044,190	2,500	1,231	20,889,542	2,500	1,136	19,279,504	2,500	-	1,231	20,889,080	3,500	650	11,027,183	
Jabar	1,500	280	4,347,000	1,500	444	6,893,100	2,000	356	5,716,980	3,000	348	5,778,540	2,000	360	5,977,800	1,594,080	1,500	156	2,590,380	1,500	480	7,344,723	1,500	-	384	5,826,362	2,500	228	3,785,940	
Jabar	1,500	261	4,280,243	1,500	192	3,146,400	2,000	96	1,682,640	2,500	195	3,420,630	1,500	279	4,892,940	4,627,260	2,000	78	1,372,680	1,500	86	1,509,210	1,500	-	207	3,630,960	3,000	201	3,523,028	
Jabar	1,500	2,109	34,561,238	1,500	1,457	23,876,588	2,000	803	14,073,723	3,000	1,608	28,184,220	2,000	1,319	23,116,773	34,701,743	1,500	809	14,074,583	1,500	1,127	19,753,493	826	1,508	1,348	23,623,380	3,000	880	15,424,200	
Jabar	1,500	287	4,455,675	1,500	576	8,942,400	1,500	600	9,963,000	3,500	600	9,963,000	3,500	600	9,963,000	14,712,960	2,000	420	6,974,100	1,500	600	9,963,000	1,500	-	1,483	24,625,215	3,500	-	-	
Jateng	1,500	7,550	117,213,750	1,500	1,689	26,221,725	3,500	6,480	107,600,400	3,500	7,260	120,552,300	3,500	12,020	199,592,100	94,316,400	2,000	5,880	97,637,400	1,500	5,818	96,607,890	1,500	-	3,132	52,006,860	3,500	6,480	107,600,400	
Jateng	1,500	960	15,235,200	1,500	3,360	53,323,200	1,500	6	101,844	2,000	1,302	22,100,148	2,000	486	8,249,364	20,368,800	2,500	960	16,295,040	1,500	1,198	20,334,852	826	-	1,200	20,368,800	3,500	720	12,221,280	
Jateng	1,500	1,404	22,281,480	1,500	1,384	21,329,280	2,000	1,632	27,701,568	2,000	1,104	18,739,296	1,500	3,580	26,479,440	26,479,440	1,500	1,200	20,368,800	1,500	1,427	24,221,898	826	-	1,844	31,260,056	3,000	1,152	18,554,048	
Jateng	1,500	1,130	20,156,625	1,500	1,188	19,668,900	2,000	3,112	59,959,236	3,500	2,460	43,117,650	3,500	3,284	22,505,110	47,524,580	2,500	990	17,352,225	1,500	930	16,300,375	1,500	-	720	12,618,800	3,500	2,213	38,809,653	
Jateng	1,500	3,480	55,317,645	1,500	2,934	46,638,497	2,000	4,380	74,467,337	3,000	4,188	71,203,015	2,000	4,998	84,974,372	98,133,668	1,500	5,202	88,442,713	1,500	3,882	66,000,502	1,500	-	5,790	98,439,698	3,000	5,811	98,794,652	

Data promosi dan data penjualan untuk Produk K31 S-2 akan dikumpulkan selama setahun terakhir. Ini akan memberikan dataset yang komprehensif untuk analisis dan pemodelan.

Dari data tersebut dilakukan verifikasi dan kualitas data, dalam mengolah data tersebut dilakukan beberapa tahap diantaranya:

Convert Data to Colum

Data yang diperoleh masih berupa data baris kekanan, agar dapat diload diconvert ke data baris

Aliasing Data

Merubah bulan, area menjadi kode angka karena akan digunakan sebagai variable dalam perhitungan

Type Data

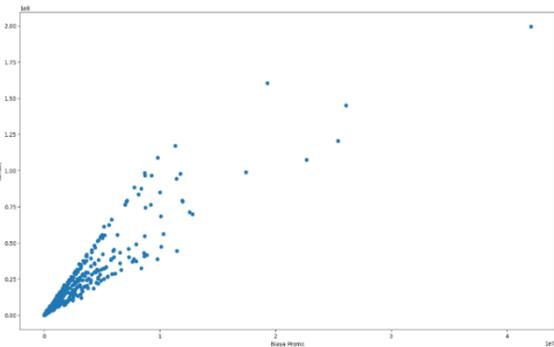
Dalam data ini terdapat 3 jenis type data yaitu *Object (String)*, *integer* dan *Float*

Data Error

Tidak ditemukan *data error(n/a)* yang berarti data sudah bersih

Persiapan data

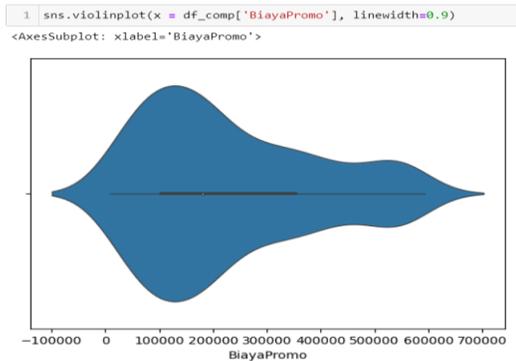
Persiapan data meliputi penanganan nilai yang hilang, pencilan, dan mentransformasi data menjadi format yang sesuai untuk Regresi Linear. Langkah ini penting untuk memastikan akurasi dan keandalan model.



Gambar 2. Sebaran Data Asli

Dalam diagram scatter plot dapat dilihat sebaran data yang menghasilkan bahwa data tersebut sebarannya terlalu luas dapat disimpulkan terdapat beberapa data yang menyimpang, sehingga harus dibersihkan data agar prediksi menjadi tepat.

	Omset	BiayaPromo
341	6403652	540000
342	6310054	540000
343	6194218	558000
345	6175500	537000
349	5994202	594000

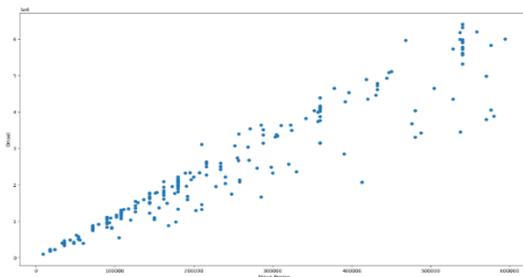


Gambar 3. Violin Plot

Dari data diatas dapat dilakukan sebaran data dengan menggunakan violin, agar dapat dilihat kesesuaian data apakah data merata atau tidak. Ternyata dari data tersebut terlalu lebar dan sebarannya terlalu luas, ketika data terlalu luas maka menyebabkan hasil prediksi menjadi kurang akurat. Maka perlu direduksi atau perlu dibatasi sebarannya.

```
1 #menghapus outlier dan mengambil data yang diperlukan
2 df_comp=df[(df['Omset']<100000000) & (df['BiayaPromo']< 600000) & (df['BiayaPromo']>1)]
3 df_company=df_comp[['Omset','BiayaPromo']].copy()
```

Dari data tersebut dilakukan pembatasan nilai omset yang dibawah Rp. 100.000.000,- dan biaya dibawah Rp. 600.000,- maka hasil pembatasan data tersebut sebaran menjadi lebih merata.



Gambar.4 Sebaran Data setelah Direduksi

Setelah dilakukan reduksi, maka sebaran data menjadi siap untuk dilakukan pemodelan setelah ini dilakukan uji linier regresi . Analisis Data Eksploratori (EDA) akan dilakukan untuk mengidentifikasi tren, musiman, dan potensi pencilan dalam data. Analisis ini akan membantu memahami hubungan antara upaya promosi dan penjualan.

Pemodelan dan Evaluasi

Dalam pemodelan ini dilakukan pembagian data menjadi komposisi 80:20 yaitu data training dan data testing, dengan tujuan 80 untuk membuat modeling dan 20 dilakukan membuat testing yang sudah dibuat.

```
1 #data company
2 msk_company=np.random.rand(len(df_comp))<0.8
3 train_company=df_company[msk_company]#data 80%
4 test_company=df_company[~msk_company]#sisa dari msk sebanyak 20%
```

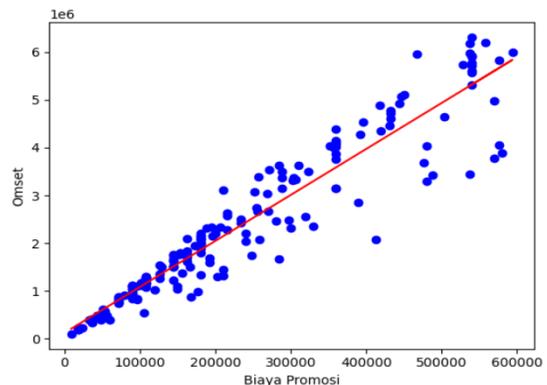
Gambar 5. Pembagian Model

Mencari Nilai Coefficient dan Intercept

Dari rumus Coefficient ini untuk mengetahui tren yang seharusnya, sesuai dengan gambar garis linier.

```
1 #modelling
2 from sklearn import linear_model
3 regr=linear_model.LinearRegression()
4 train_x =np.asarray(train_company[['BiayaPromo']])
5 train_y=np.asarray(train_company[['Omset']])
6 regr.fit(train_x,train_y)
7 #the coefficeints
8 print("Coefficients: ",regr.coef_)
9 print('Intercept: ', regr.intercept_)
```

Coefficients: [[9.61652512]]
Intercept: [123757.37057197]



Gambar 6. Trend Biaya Promosi dan Omset

Membuat garis plot linear regresi untuk melihat trend seharusnya. Dari data terlihat tren naik dapat dilihat dari grafik diatas semakin besar biaya promosi yang dikeluarkan memang seharusnya omset harus naik karena menjadi trend yang seharusnya naik. Artinya jika plot dibawah garis linier artinya promosi terlalu besar dibandingkan omset yang didapatkan.

Evaluasi Model

Dalam tahapan berikutnya untuk menghitung Akurasi dan efektivitas model akan dievaluasi menggunakan metrik seperti MAE, MSE, dan skor R2. MAE sebesar 275066.17, MSE sebesar 20519337364.86, dan skor R2 sebesar 0.93, menunjukkan bahwa bahwa sebagian besar variasi dalam variabel dependen dapat dijelaskan oleh model.

```

1 #evaluasi
2 #semakin tinggi R2 score(0-1) semakin bagus
3 from sklearn.metrics import r2_score
4 test_x = np.asarray(test_company[['BiayaPromo']])
5 test_y = np.asarray(test_company[['Omset']])
6 test_y_hat = regr.predict(test_x)
7 #kesalahan pada saat menjalankan perhitungan sebanyak
8 print("Mean Absolute error:", f"{np.mean(np.absolute(test_y_hat - test_y)):.2f}")#
9 #kemungkinan error dari prediksi sejauh mana
10 print("Residula sum of squares (MSE):", f"{np.mean((test_y_hat - test_y)**2):.2f}")
11 print("R2-score: %.2f" % r2_score(test_y_hat, test_y))

```

Mean Absolute error: 275,066.17
 Residula sum of squares (MSE): 20,519,337,364.86
 R2-score: 0.93

Gambar 7. Testing Model

Rumus yang dihasilkan dalam perhitungan untuk mendapatkan berapa banyak omset prediksi yang didapatkan. Berikut dibawah ini rumus yang didapatkan.

Tabel 2. Data Hasil Penelitian

Promo	Qty	Omset	Budget	Prediksi Omset	Selisih	Presentase Budget	Presentase Res
3000	1072	11,905,554.38	3,216,000.00	31,050,502	(19,144,947.79)	0.103573204	0.270126018
3500	3276	44,606,343.59	11,466,000.00	110,386,834	(65,780,490.85)	0.103871083	0.257048641
3500	2394	32,596,943.40	8,379,000.00	80,700,621	(48,103,677.98)	0.103828197	0.257048641
3500	2790	38,849,722.18	9,765,000.00	94,029,125	(55,179,403.02)	0.103850801	0.251353149
4000	455	7,378,155.00	1,820,000.00	17,625,833	(10,247,678.10)	0.103257531	0.246674135
4000	624	10,228,680.00	2,496,000.00	24,126,604	(13,897,924.08)	0.103454261	0.244019756
4000	432	7,106,940.00	1,728,000.00	16,741,113	(9,634,172.78)	0.103218945	0.243142618
3500	1221	18,472,581.80	4,273,500.00	41,219,977	(22,747,395.68)	0.103675457	0.231342865

$$=9.61652512347988 * [\text{Budget}] + 123757.37057197$$

Rumus yang diatas dimasukan dalam prediksi omset, hasil dari perbandingan diatas yang didapat dengan prediksi omset. Dari data diatas terlihat ternyata budget promosi tidak sebanding dari realisasi omset atau dapat dikatakan perusahaan mengeluarkan budget yang berlebih.

KESIMPULAN

Hasil penelitian ini menghasilkan dua perspektif yaitu yang pertama perhitungan prediksi dan yang kedua perbandingan hasil prediksi dengan realisasinya. Bahwa ternyata metode yang selama ini digunakan masih jauh

dari prediksi yang dilakukan dalam menggunakan metode regresi ini.

DAFTAR PUSTAKA

CRISP-DM Consortium. (2015). CRISP-DM 1.0 Step-by-step data mining guide. CRISP-DM Consortium

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate Data Analysis. Pearson.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis. Wiley.

Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python. John Wiley & Sons.

W Gata, H B Novitasari , R Nurfalah , R Hernawati , M J Shidiq (2019). Analysis of Regression Algorithm to Predict Administration, Production, and Delivery to Accuracy of Delivery of Products in Cosmetic Industry. IOP Conf. Series

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Wiga, Baihaqi. Et all. 2019. Regresi Linier Sederhana untuk Memprediksi Kunjungan Pasien di Rumah Sakit Berdasarkan Jenis Layanan dan Umur Pasien. Jurnal SIMETRIS