

PENERAPAN APLIKASI RAPIDMINER PADA PREDIKSI DATA PENJUALAN ALBUM K-POP MENGGUNAKAN METODE REGRESI LINEAR

Windi Inawati¹, Rudi Kurniawan², Yudhistira Arie Wijaya³

Program Studi Teknik Informatika

STMIK IKMI Cirebon

Jalan Perjuangan No. 10B Majasem Kec. Kesambi Kota Cirebon

windiinawati30@gmail.com¹, rudi226ikmi@gmail.com², yudhistira010471@gmail.com³

Abstrak

Perkembangan teknologi telah secara signifikan mempengaruhi berbagai aspek kehidupan manusia, terutama dalam dunia bisnis yang menjadi elemen kunci kesuksesan perusahaan dan pengusaha dalam meraih keuntungan yang lebih besar. Fenomena ini tidak terkecuali dalam dunia musik, khususnya genre musik K-Pop yang berasal dari Korea Selatan. Penjualan yang seringkali tidak menentu membuat perusahaan memproduksi album melebihi permintaan pembeli, sehingga terjadi penumpukan album yang tidak terjual dan membuat perusahaan rugi. Maka dibutuhkan prediksi disetiap penjualan album agar tidak terjadi penumpukan dan kerugian. Metode yang digunakan dalam penelitian ini adalah Regresi Linear, sebuah pendekatan statistik yang efektif dalam memodelkan hubungan antara variabel dependen dan variabel independen. Penelitian ini bertujuan untuk mengimplementasikan *Data Mining* menggunakan Algoritma Regresi Linear dengan mengidentifikasi faktor-faktor yang berpengaruh signifikan terhadap penjualan album K-Pop yang dilakukan oleh agensi dan idol K-Pop. Hasil penelitian dengan metode Regresi Linear dalam menggunakan RapidMiner mendapatkan hasil nilai *Root Mean Square Error* (RMSE) sebesar 276051.973, dan untuk nilai *Absolute Error* sebesar 97012.624. Diantara nilai tersebut, nilai Absolute Error relatif lebih kecil dibandingkan dengan nilai RMSE. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam memperkirakan tingkat penjualan, serta membantu perusahaan dalam mengatur produksi album untuk memenuhi permintaan yang terus berkembang di berbagai negara.

Kata Kunci : Data Mining, RapidMiner, Regresi Linear, Penjualan

Abstract

The development of technology has significantly affected various aspects of human life, especially in the business world, which is a key element of the success of companies and entrepreneurs in achieving greater profits. This phenomenon is no exception in the world of music, especially the K-Pop music genre originating from South Korea. Sales that are often erratic make companies produce albums that exceed buyer demand, resulting in a buildup of unsold albums and making the company lose money. So predictions are needed for each album sale so that there is no buildup and loss. The method used in this research is linear regression, a statistical approach that is effective in modeling the relationship between dependent variables and independent variables. This research aims to implement data mining using a linear regression algorithm by identifying factors that have a significant effect on K-Pop album sales made by agencies and K-Pop idols. The results of research with the linear regression method using RapidMiner obtained a root mean square error (RMSE) value of 276051.973 and an absolute error value of 97012.624. Among these values, the absolute error value is relatively smaller than the RMSE value. The results of this study are expected to contribute to forecasting sales levels as well as assisting companies in organizing album production to meet the growing demand in various countries.

Keywords: Data Mining, RapidMiner, Linear Regression, Sales

PENDAHULUAN

Perkembangan teknologi yang terjadi dari masa ke masa pada dasarnya mempengaruhi hampir semua aspek kehidupan manusia (Habibi, 2020) terutama dalam bidang bisnis yang menjadi salah satu faktor pendukung keberhasilan perusahaan dan pengusaha untuk memperoleh keuntungan yang lebih luas (Khoziyah & Lubis, 2021). Salah satu yang terkena dampak perkembangan teknologi yaitu pada dunia musik, terutama K-Pop. K-pop sendiri merupakan jenis musik populer

yang berasal dari Korea Selatan, dengan adaptasi beberapa genre dan style musik dari berbagai dunia seperti R&B, Jazz, Rock, Hip hop, ataupun musik tradisional Korea sendiri (Haida et al., 2022).

Permasalahan pada penelitian ini adalah Penjualan yang seringkali tidak menentu membuat perusahaan memproduksi album melebihi permintaan pembeli, sehingga terjadi penumpukan album yang tidak terjual dan membuat perusahaan rugi. Maka dari itu, dibutuhkan prediksi disetiap penjualan album agar tidak terjadi penumpukan dan kerugian. Salah satu metode yang dapat digunakan untuk memprediksi suatu nilai dengan baik dan akurat adalah Regresi Linear (Prasetyo et al., 2021). Dalam Ilmu Komputer atau Ilmu Statistika, dikenal sebuah cara untuk mengetahui sebuah prediksi, ilmu tersebut adalah Data Mining (Siregar et al., 2021). Data Mining sendiri adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data (Boy, 2020). Untuk dapat membantu memprediksi penjualan kedepannya lebih baik, lebih tepat dan lebih cepat, maka dibutuhkan Analisis Data Mining dengan dukungan metode Regresi Linier (Adiguno et al., 2022). Regresi linear adalah teknik yang digunakan untuk memperoleh model hubungan antara 1 variabel dependen dengan 1 variabel atau lebih variabel independen (Jhonson et al., 2020), variabel dependen merupakan variabel akibat atau variabel yang dipengaruhi sedangkan variabel independen merupakan variabel sebab atau variabel yang mempengaruhi (Ayuni & Fitriana, 2019).

Penelitian ini bertujuan untuk mencari nilai *Root Mean Squared Error* (RMSE) dan nilai *Absolute Error* terhadap data penjualan album K-Pop. Serta memperkirakan dengan akurat tingkat penjualan album berdasarkan variabel-variabel yang mempengaruhi penjualan album melalui analisis Regresi Linear.

Dalam penelitian ini Peneliti berharap dapat membantu perusahaan dalam menentukan jumlah produksi album ditahun-tahun berikutnya, dan membantu perusahaan dalam memenuhi banyaknya jumlah permintaan album di berbagai Negara.

PENELITIAN RELEVAN

Beberapa penelitian sebelumnya yang menggunakan berbagai metode prediksi seperti penelitian (Luh et al., 2021) dengan judul “Penerapan Metode Regresi Linier untuk Memprediksi Permohonan ITAS” yang menggunakan metode Regresi Linear untuk menghitung akurasi peramalan digunakan metode *Mean Absolute Percentage Error* (MAPE) karena dari hasil perbandingan metode akurasi perhitungan peramalan dengan data yang tersedia, MAPE memiliki akurasi eror yang kecil dibandingkan metode lainnya. Penelitian (Hamdanah & Fitriana, 2021) dengan judul “Analisis Performansi Algoritma Linear Regression Dengan Generalized Linear Model Untuk Prediksi Penjualan Pada Usaha Mikro, Kecil, Dan Menengah” untuk mendapatkan prediksi penjualan barang untuk tahun berikutnya berdasarkan data penjualan yang telah diinputkan sebelumnya. Data penjualan dimanfaatkan menjadi informasi dalam meningkatkan penjualan barang. Penerapan untuk penelitian ini menggunakan sebuah metode prediksi dengan perbandingan algoritma *Linear Regression* dan *Generalized Linear*. Metode *Linear Regression* dan *Generalized Linear* yang didukung dengan perhitungan RMSE. RMSE (*Root Mean Square Error*) berfungsi untuk mendapatkan besaran tingkat kesalahan dari hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka semakin akurat nilai prediksinya. Dan penelitian (Tampubolon & Saripurna, 2020) dengan judul “Implementasi Regresi Linier Berganda Untuk Memprediksi Tingkat Penjualan Alat Kelistrikan” memprediksi penjualan yang akan datang. Akan tetapi prediksi ini selalu kurang tepat karena hanya melihat perkiraan berdasarkan penjualan yang telah terjadi yang dilakukan secara manual yang ditangani oleh seorang admin pada perusahaan tersebut. Dengan metode yang digunakan dalam penelitian ini adalah Regresi Linier Berganda. metode ini cocok untuk digunakan dalam penelitian ini karena dapat mengetahui seberapa besar pengaruh variabel bebas terhadap variabel terikat baik secara parsial maupun secara bersama-sama..

METODE PENELITIAN

Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data. Data mining dibagi menjadi beberapa metode yaitu 1) Asosiasi (*Association*), 2) Pengklusteran (*Clustering*), 3) Prediksi (*Prediction*), 4) Klasifikasi (*Classification*), dan 5) Estimasi (*Estimation*) (Boy, 2020).

Data dalam bentuk file Excel yang bernama Data Penjualan Album K-Pop diperoleh dari situs website *Circle Chart*. Jumlah *record* sebanyak 55 *record* dengan penjualan sebanyak 2,522,971 keping album pada bulan September-Desember 2022 dan 219,338 keping album pada bulan Januari-September 2023. Banyaknya data yang akan dipakai ditunjukkan pada tabel 1.1 berikut.

Tabel 1. Dataset

No	Data Minggu Ke-	Tahun	Jumlah Terjual
1	1	2022	2,141,281
2	2	2022	242,962
3	3	2022	72,963
4	4	2022	3,271
5	5	2022	7,961
6	6	2022	9,241
7	7	2022	3,530
8	8	2022	4,286
9	9	2022	4,216
10	10	2022	2,839
11	11	2022	2,528
12	12	2022	6,333
13	13	2022	6,441
14	14	2022	5,996
15	15	2022	5,776
16	16	2022	3,347
17	17	2023	4,102
18	18	2023	3,399
19	19	2023	5,472
20	20	2023	4,411
21	21	2023	11,658
22	22	2023	9,598
23	23	2023	7,814
24	24	2023	4,530
25	25	2023	14,846

26	26	2023	3,179
27	27	2023	6,367
28	28	2023	3,976
29	29	2023	5,390
30	30	2023	4,770
31	31	2023	3,776
32	32	2023	3,528
33	33	2023	2,794
34	34	2023	4,159
35	35	2023	3,918
36	36	2023	4,554
37	37	2023	3,541
38	38	2023	4,479
39	39	2023	4,568
40	40	2023	5,043
41	41	2023	5,281
42	42	2023	5,331
43	43	2023	3,011
44	44	2023	5,892
45	45	2023	4,531
46	46	2023	4,288
47	47	2023	15,435
48	48	2023	4,301
49	49	2023	4,309
50	50	2023	5,707
51	51	2023	2,917
52	52	2023	8,363
53	53	2023	2,124
54	54	2023	5,304
55	55	2023	12,662

Metode yang digunakan dalam penelitian ini menggunakan analisis Regresi Linear. Analisis Regresi Linear merupakan sebuah teknik analisis data yang digunakan untuk menentukan hubungan antara satu variabel terikat dengan variabel bebas. Metode utama untuk prediksi adalah membangun model regresi dengan mencari hubungan antara satu atau lebih variabel bebas atau prediktor (X) dan variabel terikat atau respons (Y) (Prasetyo et al., 2021).

Rumus persamaan Regresi Linear

$$Y = a + bX$$

Dimana :

Y = Variabel dependen (variabel terikat)

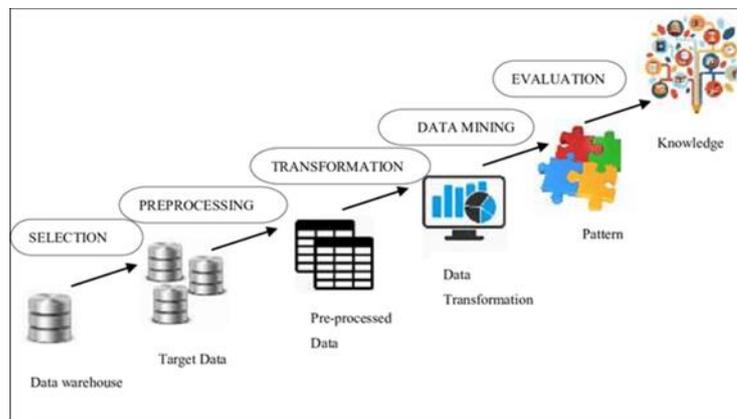
X = Variabel Independen (variabel bebas)

a = Konstanta (nilai dari Y apabila X = 0)

b = Koefisien regresi (pengaruh positif atau negatif)

Penelitian ini juga memanfaatkan aplikasi RapidMiner dalam mengimplementasikan metode Regresi Linear. RapidMiner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu Data Mining. Platform dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran. RapidMiner memiliki sekitar 100 solusi pembelajaran untuk pengelompokan, klasifikasi dan analisis regresi. RapidMiner juga mendukung sekitar 22 format file, seperti .xls, .csv, dan sebagainya (Prasetyo et al., 2021).

Populasi yang akan digunakan adalah sebagai penelitian adalah idol K-Pop yang melakukan penjualan album. Sedangkan sampel adalah yang akan digunakan adalah idol group Blackpink. Penelitian ini juga menggunakan teknik analisis *Knowledge Discovery in Database* (KDD) adalah proses yang bertujuan untuk menggali dan menganalisis data yang sangat besar menjadi informasi yang berguna untuk pengetahuan. Berikut di bawah ini Gambar 1.1 alur Knowledge Discovery in Database (KDD)

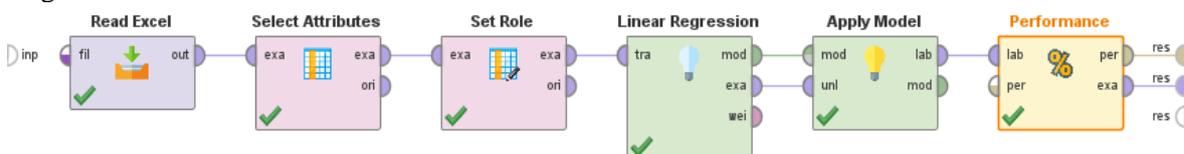


Gambar 1. Alur Knowledge Discovery in Database (KDD)

HASIL DAN PEMBAHASAN

Proses Uji Coba dan Evaluasi

Proses uji coba dilakukan dengan menggunakan cara import data dan menggunakan operator Regresi Linear.



Gambar 1. Proses uji coba data

Hasil dari uji coba data pada RapidMiner, didapatkan hasil seperti pada tabel di bawah ini

Tabel 2. Hasil prediksi menggunakan RapidMiner

No	Jumlah Terjual	Prediction (Jumlah Terjual)
1	2141281	183200.204
2	242962	179798.268
3	72963	176396.333
4	3271	172994.397
5	7961	169592.462
6	9241	166190.526
7	3530	162788.591
8	4286	159386.655
9	4216	155984.720
10	2839	152582.784
11	2528	149180.849
12	6333	145778.913
13	6441	142376.978
14	5996	138975.042
15	5776	135573.107
16	3347	132171.171
17	4102	70260.570
18	3399	66858.634
19	5472	63456.699
20	4411	60054.763
21	11658	56652.828
22	9598	53250.892
23	7814	49848.957
24	4530	46447.021
25	14846	43045.086
26	3179	39643.150
27	6367	36241.215
28	3976	32839.279
29	5390	29437.343
30	4770	26035.408
31	3776	22633.472
32	3528	19231.537
33	2794	15829.601
34	4159	12427.666

35	3918	9025.730
36	4554	5623.795
37	3541	2221.859
38	4479	-1180.076
39	4568	-4582.012
40	5043	-7983.947
41	5281	-11385.883
42	5331	-14787.818
43	3011	-18189.754
44	5892	-21591.689
45	4531	-24993.625
46	4288	-28395.560
47	15435	-31797.496
48	4301	-35199.431
49	4309	-38601.367
50	5707	-42003.302
51	2917	-45405.238
52	8363	-48807.173
53	2124	-52209.109
54	5304	-55611.044
55	12662	-59012.980

Dari hasil nilai yang telah di prediksi, terdapat beberapa hasil nilai yang negatif (-) dikarenakan hasil nilai prediksi dengan nilai yang sebenarnya memiliki perbandingan atau selisih nilai yang cukup jauh. Maka nilai yang dihasilkan bersifat negatif (-).

Hasil penilaian *Root Mean Square Error* (RMSE) dan *Absolute Error* yang dilakukan dengan memanfaatkan operator performance pada RapidMiner, didapatkan nilai sebesar 276051.973 untuk RMSE, 97012.624 untuk *Absolute Error*.

Hasil penilaian *Root Mean Square Error* (RMSE) dan *Absolute Error* yang dilakukan dengan memanfaatkan operator performance pada RapidMiner, dapat disimpulkan bahwa nilai parameter yang relatif lebih kecil dan lebih baik diantara *Root Mean Square Error* (RMSE) dan *Absolute Error* adalah *Absolute Error* dengan nilai sebesar 97012.624.

Tabel 1. Hasil dengan menggunakan operator performance pada RapidMiner

<i>Performance</i>	Hasil
<i>Root Mean Square Error</i> (RMSE)	276051.973
<i>Absolute Error</i>	97012.624

SIMPULAN

Berdasarkan hasil percobaan yang dilakukan sebelumnya, dapat disimpulkan bahwa metode Regresi Linear cukup dapat diterapkan untuk memprediksi terhadap Penjualan Album K-Pop. Meskipun terdapat beberapa hasil yang nilainya tidak sesuai. Dan dari hasil perhitungan dengan menggunakan operator *Performance* di RapidMiner diperoleh nilai *Root Mean Square Error* (RMSE) sebesar 276051.973. Nilai tersebut terbilang cukup besar dikarenakan jauh dari mendekati nilai 0. Karena perhitungan RMSE adalah ketika nilai yang dihasilkan lebih kecil dan mendekati 0, maka hasil tersebut lebih akurat. Untuk hasil dari *Absolute Error*, diperoleh nilai sebesar 97012.624. Diantara parameter *Root mean Square Error* (RMSE) dengan *Absolute Error*, nilai *Absolute Error* yang lebih kecil dibandingkan dengan nilai RMSE.

DAFTAR PUSTAKA

- Adiguno, S., Syahra, Y., & Yetri, M. (2022). *Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda*. 1, 275–281.
- Ayuni, G. N., & Fitriyah, D. (2019). *Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ*. 14(2), 79–86.
- Boy, A. F. (2020). *Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)* *Journal of Science and Social Research* ISSN 2615 – 3262 (Online . 4307(August), 78–85.
- Hamdanah, F. H., & Fitriyah, D. (2021). *ANALISIS PERFORMANSI ALGORITMA LINEAR REGRESSION DENGAN GENERALIZED LINEAR MODEL UNTUK PREDIKSI PENJUALAN PADA USAHA MIKRO, KECIL, DAN MENENGAH*. 10, 23–32.
- Jhonson, I., Saragih, A., Rumahorbo, I., & Yulistira, R. (2020). *PREDIKSI CURAH HUJAN BULANAN DI DELI SERDANG*. 7(2), 6–14.
- Luh, N., Arya, W., Aurelius, R., Diaz, N., Dwi, K., Novianti, P., & Informatika, F. (2021). *Penerapan Metode Regresi Linier untuk Memprediksi Permohonan ITAS*. 92–100. <https://doi.org/10.30864/eksplora.v10i2.380>
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). *Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression*. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- Siregar, N. N., Syahra, Y., & Syaifudin, M. (2021). *Penerapan Data Mining Untuk Memprediksi Penggunaan Daya Listrik Pada PT . PLN (Persero) Rayon Medan Selatan Dengan Menggunakan Metode Regresi Linier Berganda*. 20(1), 20–27.
- Tampubolon, D., & Saripurna, D. (2020). *Implementasi Regresi Linier Berganda Untuk Memprediksi Tingkat Penjualan Alat Kelistrikan*. 3(1), 176–185.