

Analisis Marketplace Shopee Untuk Memprediksi Penjualan dengan Algoritma Regresi Linier

Yusuf Syakir¹, Teguh Iman Hermanto², Yudhi Raymond Ramadhan³

^{1,2,3}Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana Purwakarta, Indonesia

e-mail: yusufsyakir23@wastukencana.ac.id¹, teguhiman@wastukencana.ac.id², yudhi.raymond@wastukencana.ac.id³

Abstract

Many methods can be used to predict sales, one of which is the processing of sales data using the method of data mining with a linear regression algorithm. The data in this study used is data on sales of the Ariqa Collection Boutique in the Shopee marketplace starting from May 2020 to April 2022. By using a linear regression algorithm, the Ariqa Collection Boutique can predict sales estimates based on total visitors and total orders. The data mining method used is SEMMA (Sample, Explore, Modify, Model, Assess). With the Rapidminer Studio 9.10 tools the test results Mean Square Error (MSE) value is 5.172.628.212.404, Root Mean Square Error (RMSE) is 2.274.341, and Mean Absolute Percentage Error (MAPE) is 4.34%. Based on the MAPE value obtained, the accuracy of the linear regression algorithm in predicting sales of Ariqa Collection Boutique in the Shopee marketplace provides high accuracy.

Keywords: Data Mining, Linier Regression, Sales Prediction, SEMMA

Abstrak

Banyak metode yang dapat digunakan untuk memprediksi penjualan, salah satunya adalah pengolahan data penjualan menggunakan metode data mining dengan algoritma regresi linier. Data pada penelitian ini yang digunakan adalah data penjualan Butik Ariqa Collection pada marketplace Shopee terhitung dari bulan Mei 2020 hingga April 2022. Dengan menggunakan algoritma regresi linier pihak Butik Ariqa Collection dapat memprediksi perkiraan penjualan berdasarkan total pengunjung, dan total pesanan. Adapun metode Data Mining yang digunakan yaitu SEMMA (Sample, Explore, Modify, Model, Assess). Dengan menggunakan RapidMiner Studio 9.10 jumlah nilai hasil pengujian Mean Square Error (MSE) adalah 5.172.628.212.404, Root Mean Square Error (RMSE) adalah 2.274.341, dan Mean Absolute Percentage Error (MAPE) adalah 4,34%. Berdasarkan nilai MAPE yang didapat, maka akurasi algoritma regresi linier dalam memprediksi penjualan Butik Ariqa Collection pada marketplace Shopee memberikan akurasi yang tinggi.

Kata kunci: Data Mining, Regresi Linier, Prediksi Penjualan, SEMMA

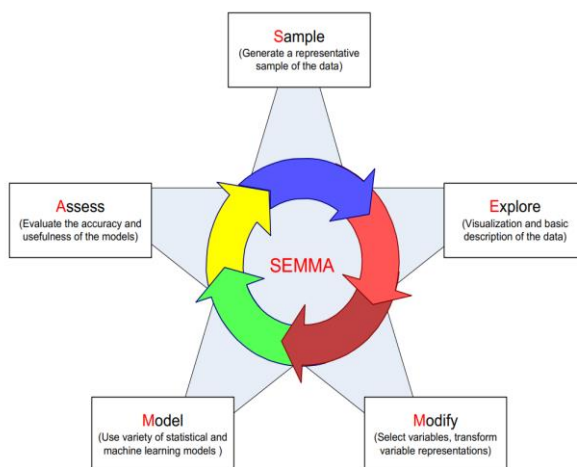
1. PENDAHULUAN

Ariqa Collection merupakan salah satu dari banyak bisnis yang terdampak oleh pandemi covid-19 lalu berhasil bertahan dan bahkan meningkatkan performa bisnisnya melalui *e-commerce*, dapat dilihat berdasarkan data *Seller Centre* Shopee dari peningkatan penjualannya pada tahun 2020 dengan total penjualan Rp37.503.321 dan meningkat pesat pada tahun 2021 dengan total penjualan mencapai Rp6.14M. Dengan membuka *online store* pada salah satu *marketplace*, bisnis ini dapat terus berkembang di saat sedang pandemi covid-19.

Dalam penelitian ini penulis melakukan proses analisis prediksi terhadap data penjualan Butik Ariqa Collection pada *marketplace* Shopee menggunakan algoritma regresi linier. Regresi linier adalah metode pengumpulan data untuk menentukan apakah ada hubungan antara variabel tertentu dengan variabel yang ingin diprediksi, dan skema peningkatan data perkiraan historis ini linier. bahkan jika tidak, itu didasarkan pada asumsi bahwa itu linier [1].

Data mining adalah metode yang menggunakan sejumlah besar data untuk mengekstrak data berharga yang dapat digunakan demi membuat keputusan penting [2]. Sederhananya, data mining atau pengembangan data dapat didefinisikan sebagai proses memilih, mengeksplorasi, dan memodelkan sejumlah besar data untuk menemukan pola atau tren yang biasanya tidak dikenali [3].

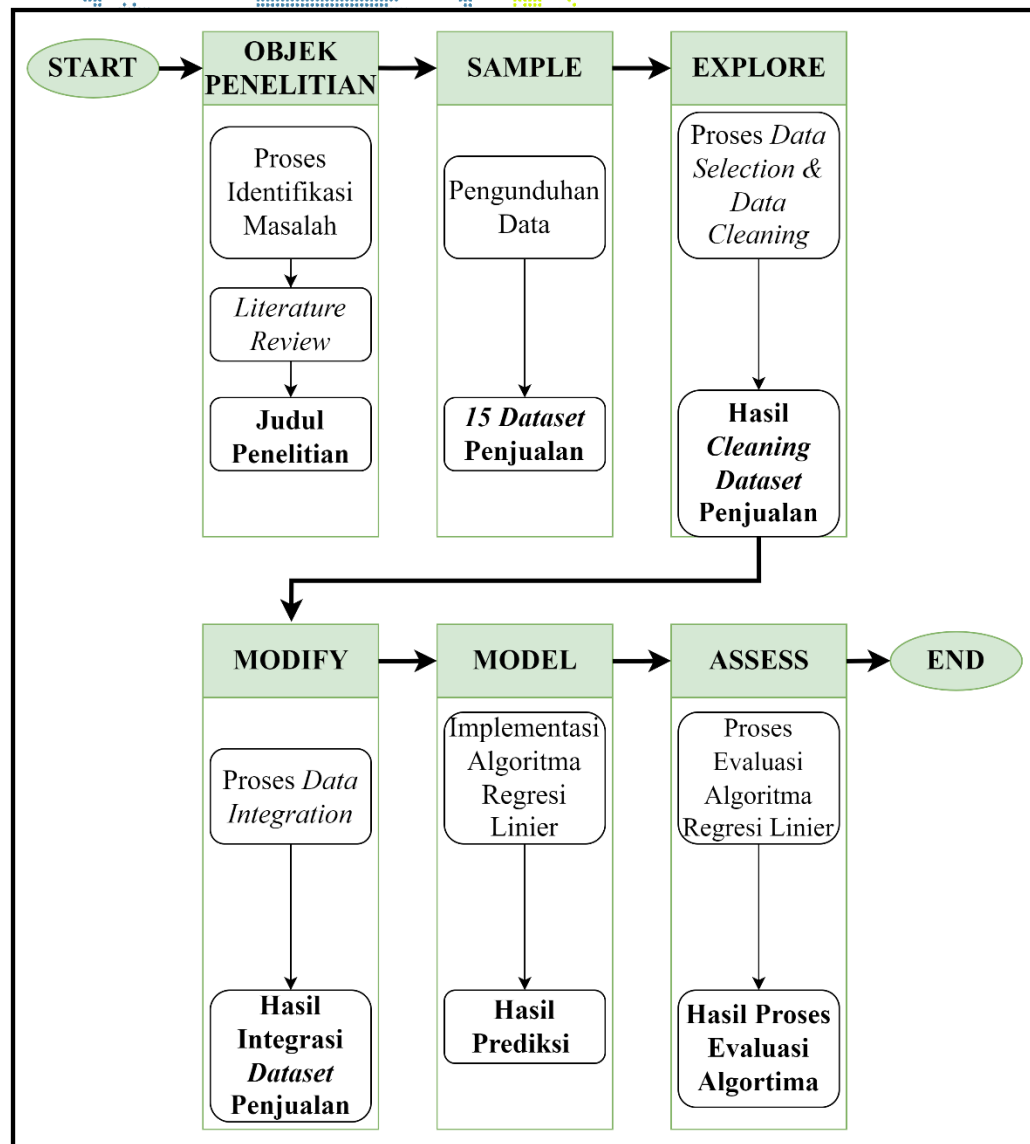
SEMMA merupakan singkatan dari *Sample, Explore, Modify, Model, Assess*. Singkatan itu mengacu terhadap proses dalam melakukan suatu objek *data mining* [4]. Metode SEMMA tertuju pada tugas modifikasi terhadap proyek penambangan data juga pemodelan data dan mengabaikan aspek kebutuhan bisnis dirancang untuk membantu pengguna *software SAS enterprise miner* [5].



Gambar 1. Tahapan SEMMA

2. METODOLOGI PENELITIAN

Dalam penelitian ini penulis menggunakan metode SEMMA untuk pengolahan data menggunakan teknik *data mining*. Untuk mempermudah pengerjaan penelitian ini, sehingga penulis membuat kerangka penelitian sebagai berikut.



Gambar 2. Kerangka Penelitian

2.1. Tahap Objek Penelitian

Pada proses ini penulis menentukan data penjualan Butik Ariqa Collection pada *marketplace* Shopee sebagai objek penelitian, selain itu penulis juga melakukan *literature review* dengan membaca beberapa jurnal yang berkaitan dengan *Data Mining* dan Regresi Linier.

2.2. Tahap Sample

Tahap ini terdiri dari pengambilan data dengan mengekstraksi sebagian dari kumpulan data besar yang cukup besar untuk memuat informasi yang signifikan [6]. Proses yang dilakukan pada tahap *sample* adalah melakukan pengunduhan sampel data penjualan dari bulan Mei 2020 sampai bulan April 2022 melalui *website Seller Centre* Shopee. Contoh data penjualan dapat

dilihat pada tabel 1:

Tabel 1. Contoh Data Penjualan

Tanggal	Total Penjualan (IDR)	Total Pesanan	Penjualan per Pesanan	Produk Dilihat	Total Pengunjung
01-08-2020	0	0	0,00	50	11
02-08-2020	0	0	0,00	33	11
03-08-2020	0	0	0,00	37	12
04-08-2020	132.919	1	132.919,00	51	13
05-08-2020	76.460	2	38.230,00	806	259
06-08-2020	256.500	2	128.250,00	778	275

2.3. Tahap *Explore*

Didalam tahap *explore* akan dilakukan eksplorasi data dengan cara menelaah tren atau anomali yang tidak terduga [7]. Pada tahap ini *dataset* berhasil didapatkan, lalu dilakukan *data cleaning* dan *data selection* pada *dataset* [8]. Pada proses *data cleaning* dilakukan penghapusan data dengan nilai penjualan kosong pada data penjualan dan diperoleh sebanyak 156 baris data yang bernilai kosong dari total 730 baris data sehingga menyisakan sebanyak 574 data yang dapat diolah sebagai objek penelitian. Setelah melalui *data cleaning* selanjutnya dilakukan proses *data selection* dengan melakukan penyeleksian atribut terhadap data untuk memilih atribut yang relevan terhadap analisis yang akan dilakukan, dalam hal ini penulis memilih atribut tanggal, total penjualan, total pesanan dan total pengunjung lalu menghapus atribut selain atribut tersebut. Contoh data penjualan setelah melalui tahap *explore* dapat dilihat pada tabel 2 berikut:

Tabel 2. Tahap *Explore*

Tanggal	Total Penjualan (IDR)	Total Pesanan	Total Pengunjung
04-08-2020	132.919	1	13
05-08-2020	76.460	2	259
06-08-2020	256.500	2	275

2.4. Tahap *Modify*

Proses yang dilakukan pada tahapan *modify* adalah dilakukan proses *data integration* dengan melakukan penggabungan data-data [9]. Penggabungan data dilakukan pada data penjualan dari bulan Mei 202 sampai Maret 2022 yang akan digunakan untuk *data training*, dan untuk data bulan April 2022 akan digunakan sebagai *data testing*. Contoh data penjualan setelah melalui tahap *data integration* dapat dilihat pada tabel 3 berikut ini:

Tabel 3. *Data Integration*

Tanggal	Total Penjualan (IDR)	Total Pesanan	Total Pengunjung
28/05/2020	65.500	1	21
04/06/2020	114.130	1	11
10/06/2020	63.000	1	24
31/07/2020	102.500	2	13
04/08/2020	132.919	1	13
05/08/2020	76.460	2	259
06/08/2020	256.500	2	275
07/08/2020	50.000	1	208
08/08/2020	410.580	5	240
09/08/2020	734.000	6	136
10/08/2020	640.671	5	273
12/08/2020	60.500	1	45
13/08/2020	142.633	2	56
14/08/2020	795.016	4	48
15/08/2020	643.500	3	40
16/08/2020	305.500	2	28
17/08/2020	338.333	3	42
21/08/2020	63.945	1	31
22/08/2020	1.808.972	24	1455
23/08/2020	864.600	8	173
24/08/2020	372.500	4	333
25/08/2020	1.181.526	15	442
26/08/2020	745.121	10	480
27/08/2020	517.671	6	581
28/08/2020	1.087.580	9	572
29/08/2020	671.248	7	526
30/08/2020	74.700	2	385
31/08/2020	877.068	8	303

2.5. Tahap Model

Dalam tahap ini hasil *data training* dan *data testing* yang telah didapatkan akan dilakukan proses pemodelan dengan algoritma Regresi Linier Berganda [10]. Rumus perhitungan Regresi Linier Berganda yang digunakan sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Dimana :

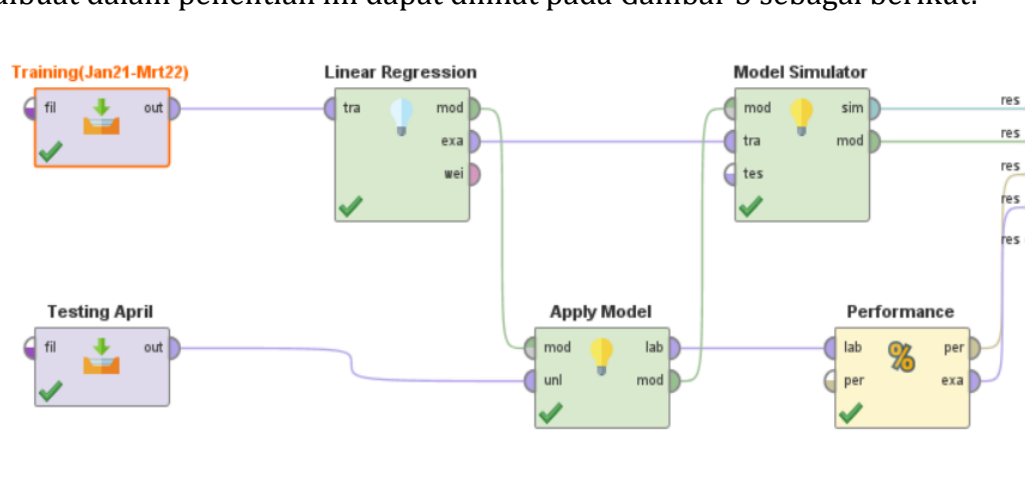
Y = Variabel yang akan diramalkan

X = Variabel bebas

a = Intercep (nilai Y pada saat X = 0)

b = Slope (perubahan rata-rata Y pada perubahan satu unit X)
 n = Banyaknya data

Pada proses pemodelan digunakan *software* RapidMiner Studio 9.10 dengan dataset bulan April 2022 sebagai *data testing* untuk menguji apakah perhitungan regresi linier yang telah diterapkan pada *data training* sudah sesuai dan memiliki tingkat *error* yang kecil. Adapun *design* RapidMiner yang dibuat dalam penelitian ini dapat dilihat pada Gambar 3 sebagai berikut:



Gambar 3. Tahap Model

2.6. Tahap Assess

Pada tahap ini dilakukan pengujian menggunakan perhitungan *Mean Square Error* (MSE) [11], *Root Mean Square Error* (RMSE) [12], dan *Mean Absolute Percentage Error* (MAPE) untuk mengetahui keandalan model dari proses yang telah dilakukan [13]. Persamaan 2 merupakan rumus perhitungan MSE, persamaan 3 merupakan rumus perhitungan RMSE dan persamaan 4 merupakan rumus perhitungan MAPE.

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

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (Y_t - Y'_t)^2} \quad (3)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - Y'_t|}{Y_t} \quad (4)$$

Dimana:

- t = Nilai periode ke t
- Y_t = Nilai sebenarnya atau aktual
- Y'_t = Nilai prediksi
- n = Jumlah data

3. HASIL DAN PEMBAHASAN

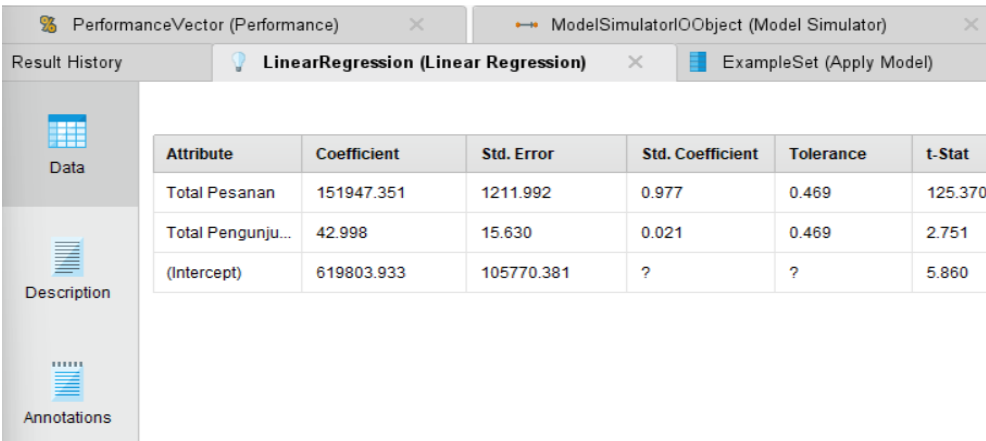
Penelitian ini menggunakan data penjualan bulan April 2022 sebagai *data testing* untuk mengukur performa model regresi linier yang dilakukan. Contoh *data testing* dapat dilihat pada tabel 4 sebagai berikut:

Tabel 4. Contoh *Data Testing*

Tanggal	Total Penjualan (IDR)	Total Pesanan	Total Pengunjung
01/04/2022	23.716.163	150	11309
02/04/2022	22.901.501	135	13221
03/04/2022	28.422.133	165	12352
04/04/2022	80.003.098	544	19069
05/04/2022	42.492.783	287	13191
06/04/2022	43.803.681	254	15189
07/04/2022	38.988.669	267	15501
08/04/2022	39.454.629	260	15249
09/04/2022	43.275.545	275	18281
10/04/2022	50.116.991	314	22849
11/04/2022	52.555.973	317	18787
12/04/2022	45.538.908	289	19282
13/04/2022	38.170.652	243	19535
14/04/2022	44.968.799	290	19001
15/04/2022	55.372.364	352	22197

3.1. Hasil Perhitungan Algoritma Regresi Linier

Hasil perhitungan algoritma regresi linier yang diperoleh menggunakan *software* RapidMiner Studio dapat dilihat pada Gambar 4 berikut:



Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat
Total Pesanan	151947.351	1211.992	0.977	0.469	125.370
Total Pengunju...	42.998	15.630	0.021	0.469	2.751
(Intercept)	619803.933	105770.381	?	?	5.860

Gambar 4. Hasil Perhitungan Algoritma Regresi Linier

Dapat dilihat pada Gambar 4 di atas sehingga dapat disimpulkan nilai persamaan regresinya yaitu seperti berikut:

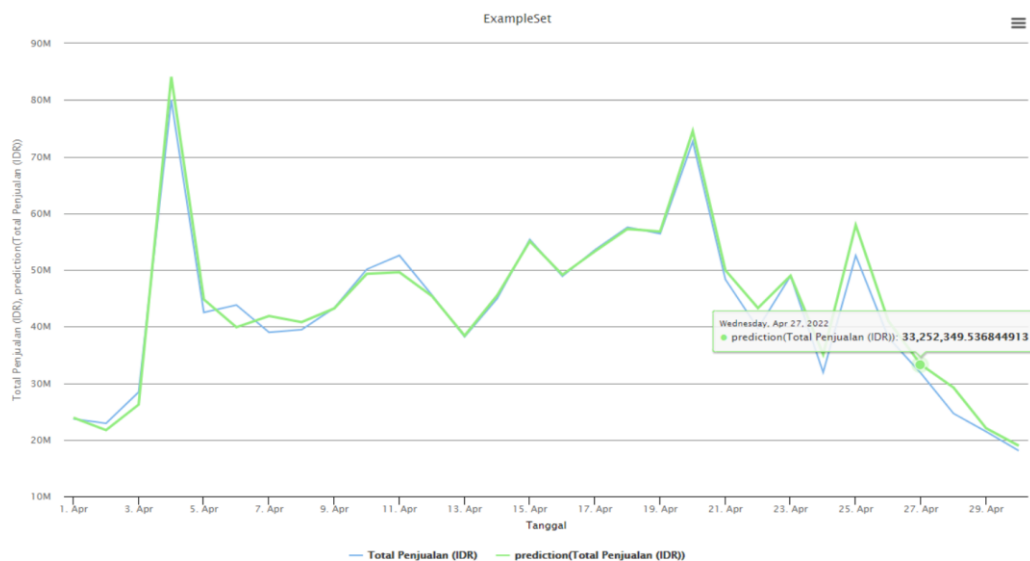
$$Y = 619803,93 + 151947,35X_1 + 42,99X_2$$

Dimana :
Y = Variabel yang akan diramalkan
X1 = Total Pesanan
X2 = Total Pengunjung

Hasil prediksi yang diperoleh dari perhitungan menggunakan algoritma regresi linier pada *software* RapidMiner Studio dapat dilihat pada Gambar 5 dan 6 berikut:

Row No.	Total Penjua...	prediction(T...	Tanggal	Total Pesan...	Total Pengu...
1	23716163	23898175.285	Apr 1, 2022	150	11309
2	22901501	21701177.934	Apr 2, 2022	135	13221
3	28422133	26222232.862	Apr 3, 2022	165	12352
4	80003098	84099098.970	Apr 4, 2022	544	19069
5	42492783	44795885.307	Apr 5, 2022	287	13191
6	43803681	39867533.502	Apr 6, 2022	254	15189
7	38988669	41856264.558	Apr 7, 2022	267	15501
8	39454629	40781797.510	Apr 8, 2022	260	15249

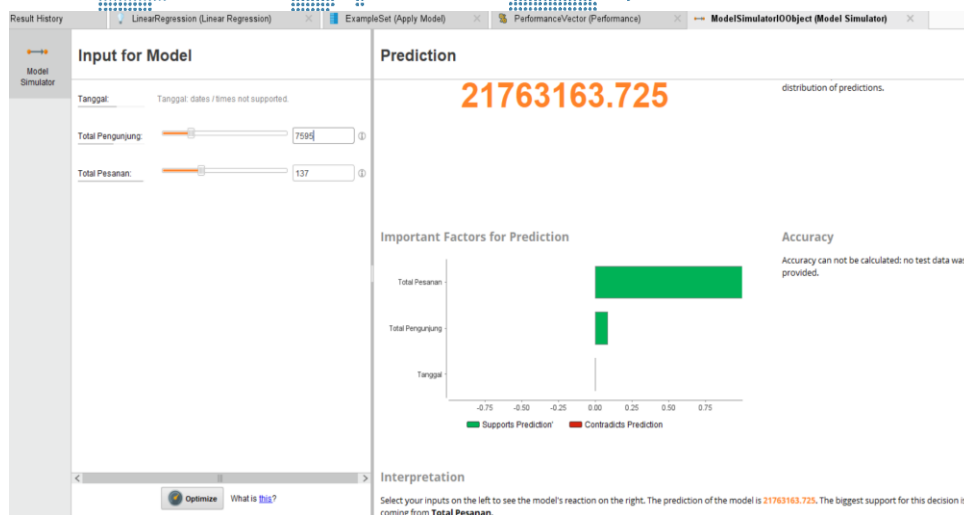
Gambar 5. Nilai Prediksi Pada RapidMiner



Gambar 6. Visualisasi Total Penjualan Dengan *Prediction*

Dapat dilihat visualisasi diagram garis pada Gambar 6 di atas bahwa garis berwarna biru merupakan Total Penjualan sedangkan garis berwarna hijau merupakan *Prediction*. Terlihat bahwa pada visualisasi diagram garis tersebut tidak terdapat perbedaan yang signifikan diantara nilai Total Penjualan dengan *Prediction*. Pada proses pemodelan penulis menambahkan operator tambahan yaitu *Model Simulator* yang berguna untuk melakukan

simulasi prediksi Total Penjualan pada variabel bebas yang akan diinputkan. Visualisasi *Model Simulator* dapat dilihat pada Gambar 7 di bawah ini:



Gambar 7. Visualisasi Model Simulator

3.2. Hasil Pengujian

Pengujian dilakukan untuk menilai data dengan mengevaluasi kegunaan dan keandalan pemodelan algoritma regresi linier yang telah dilakukan menggunakan perhitungan *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) [14]. Untuk mengetahui akurasi hasil prediksi penjualan Butik Ariqa Collection maka dilakukan pengujian dengan menggunakan perhitungan MSE, RMSE, dan MAPE dimana semakin kecil nilai *error* maka menunjukkan bahwa metode prediksi yang digunakan semakin baik [15]. Berikut ini merupakan tabel yang dibuat untuk mengetahui nilai Square Error dan Absolute Percentage Error dapat dilihat pada tabel 5.

Tabel 5. Square Error Dan Absolute Percentage Error

Tanggal	aktual	prediksi	Square Error	Absolute Percentage Error
01/04/2022	23716163	23898175	33128471919	0,007674609
02/04/2022	22901501	21701178	1440775463477	0,052412419
03/04/2022	28422133	26222233	4839560616046	0,077400951
04/04/2022	80003098	84099099	16777223949569	0,051198029
05/04/2022	42492783	44795885	5304280234861	0,054199846
06/04/2022	43803681	39867534	15493257129261	0,089858829
07/04/2022	38988669	41856265	8223104285153	0,07354946
08/04/2022	39454629	40781798	1761376252775	0,03363784
09/04/2022	43275545	43191379	7083936892	0,001944889
10/04/2022	50116991	49313742	645208674898	0,016027475

Tanggal	aktual	prediksi	Square Error	Absolute Percentage Error
11/04/2022	52555973	49594925	8767806501161	0,05634085
12/04/2022	45538908	45361683	31408641267	0,003891723
13/04/2022	38170652	38382984	45084717295	0,005562693
14/04/2022	44968799	45501548	283821467463	0,01184708
15/04/2022	55372364	55059707	97754675108	0,005646453
16/04/2022	48883403	49145842	68874451038	0,005368682
17/04/2022	53558300	53280326	77269713201	0,005190126
18/04/2022	57531332	57232462	89323419393	0,005194913
19/04/2022	56395385	56776233	145044988646	0,006753172
20/04/2022	72682318	74511418	3345608171624	0,025165686
21/04/2022	48318114	49943882	2643120934925	0,03364717
22/04/2022	39677919	43264906	12866476531031	0,090402602
23/04/2022	48955008	48963787	77075479	0,000179333
24/04/2022	31896958	35090312	10197510267787	0,10011469
25/04/2022	52524971	57915437	29057128657167	0,102626738
26/04/2022	38064770	40973638	8461515493815	0,076418915
27/04/2022	31753395	33252350	2246864703528	0,047206119
28/04/2022	24636507	29229395	21094621402686	0,186426109
29/04/2022	21437552	22035498	357538955389	0,027892439
30/04/2022	18075045	18956519	776996589281	0,048767464
Total			155178846372132	1,302547306

Berdasarkan hasil perhitungan di atas, nilai evaluasi yang didapatkan dengan perhitungan *Mean Square Error* (MSE) adalah total nilai Square Error yaitu 155.178.846.372.132 dibagi jumlah data (n) sebanyak 30 sehingga menghasilkan nilai MSE sebesar 5.172.628.212.404. nilai evaluasi yang didapatkan dengan perhitungan *Root Mean Square Error* (RMSE) adalah akar dari nilai *Mean Square Error* (MSE) yaitu $\sqrt{5.172.628.212.404}$ sehingga menghasilkan nilai RMSE sebesar 2.274.341,27. nilai evaluasi yang didapatkan dengan perhitungan *Mean Absolute Percentage Error* (MAPE) adalah total dari nilai *Absolute Percentage Error* yaitu 1,302547306 dibagi dengan banyaknya n yaitu sebanyak 30 sehingga menghasilkan nilai MAPE sebesar 0.04341 atau 4,34%.

4. SIMPULAN

Berdasarkan hasil analisis yang dilakukan dalam memprediksi penjualan menggunakan algoritma regresi linier menghasilkan nilai MSE sebesar 5.172.628.212.404, RMSE sebesar 2.274.341,27, dan MAPE sebesar 4,34% sehingga dapat disimpulkan bahwa penelitian ini mempunyai tingkat akurasi yang tinggi. Meski begitu dalam penelitian ini pengembang harus dapat menentukan estimasi Total Pesanan dan Total Pengunjung untuk dapat memprediksi nilai Total Penjualan. Informasi ini dapat menjadi salah satu pertimbangan dalam pengambilan keputusan untuk meminimalisir produksi

yang berlebih atau meningkatkan strategi promosi bagi pihak Butik Ariqa Collection.

DAFTAR PUSTAKA

- [1] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 55, 2019.
- [2] M. winny Amelia, A. S. . Lumenta, and A. Jacobus, "Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes," *J. Tek. Inform.*, vol. 11, no. 1, 2017, doi: 10.35793/jti.11.1.2017.17652.
- [3] N. A. Hasibuan *et al.*, "Implementasi Data Mining Untuk Pengaturan Layout Mini Market Dengan Menerapkan Association Rule," vol. 4, no. 4, pp. 6–11, 2017.
- [4] T. I. Hermanto and Y. Muhyidin, "Analisis Data Sebaran Bandwidth Menggunakan Algoritma Dbscan Untuk Menentukan Tingkat Kebutuhan Bandwidth Di Kabupaten Purwakarta," *Rabit J. Teknol. dan Sist. Inbformasi Univrab*, vol. 5, no. 2, pp. 130–137, 2020, doi: 10.36341/rabit.v5i2.1388.
- [5] U. Ependi and A. Putra, "Solusi Prediksi Persediaan Barang dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang)," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 139, 2019, doi: 10.26418/jp.v5i2.32648.
- [6] A. Azevedo and M. F. Santos, "Kdd , Semma And Crisp-Dm : A Parallel Overview," *IADIS Eur. Conf. Data Min.*, pp. 182–185, 2008.
- [7] M. C. Aksama and A. Wahyuniati, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan," vol. 17, no. 1, pp. 9–18, 2022.
- [8] A. P. Giovani, T. Haryanti, and L. Kurniawati, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. TEKNOINFO*, vol. 14, no. 2, pp. 116–124, 2020.
- [9] M. H. Adiya and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. 01, pp. 17–24, 2019.
- [10] A. N. Maharadja, I. Maulana, and B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021.
- [11] N. Nafi'iyah and E. Rakhmawati, "Analisis Regresi Linear Dan Moving Average Dalam Memprediksi Data Penjualan Supermarket," *J. Teknol. Inf. Dan Komun.*, vol. 12, no. 1, pp. 44–50, 2021, doi: 10.51903/jtikp.v12i1.230.
- [12] R. D. Shaputra and S. Hidayat, "Implementasi regresi linear untuk prediksi penjualan pada aplikasi point of sales restoran," *J. Autom. (Ajang Unjuk Tugas Akhir oleh Mhs. Inform.*, 2021.

- [13] Y. Aryo Jatmiko, R. Luciani R, and G. Darmawan, "Perbandingan Keakuratan Hasil Peramalan Produksi Bawang Merah Metode Holt-Winters dengan Singular Spectrum Analysis (SSA)," *J. Mat.*, vol. 03, p. 11, 2017.
- [14] G. N. Ayuni and D. Fitrianah, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79-86, 2019.
- [15] E. Analysis, T. Sales, and D. Forecasting, "Analisis Error Terhadap Peramalan Data Penjualan," *Techno.Com*, vol. 20, no. 1, pp. 1-9, 2021.