

PREDIKSI VOLUME PENJUALAN GAS PT PGN (PERSERO) MENGUNAKAN REGRESI LINEAR BERGANDA

Edison ^{1*}, David ^{2*}, Amora Antonio Pangestu^{3*}, Efanly^{4*}, Rindiany^{5*}

* Sistem Informasi, Universitas Universal

edison.81883@gmail.com ¹, davidisen00@gmail.com ², amora250618@gmail.com ³, Fanly.f01@gmail.com ⁴, rindiany27@gmail.com ⁵

Article Info

Article history:

Received 18-06-2021

Revised 05-07-2021

Accepted 10-07-2021

Keyword:

CRISP-DM, Data Mining, Multiple Linear Regression Algorithm, Rstudio, Sales Volume Predict.

ABSTRACT

PT PGN (Persero) gets an erratic gas sales volume every year, this can affect the existing gas supply. If the existing gas supply can't fulfill the demand or vice versa, the existing supply becomes excessive and affect company performance. Of these problems, it is necessary to predict sales volume to determine future inventory. This research uses Multiple Linear Regression Algorithm by applying the Cross-Industry Standard Process for Data Mining (CRISP-DM) method. The Multiple Linear Regression Algorithm aims to find the value of the regression equation, after getting the regression equation, the next step is to do it. Error analysis to determine the accuracy of predictions using MAD, MSE, and MAPE through R.Studio software. From the processing results, the results obtained from the sales volume in 2016 amounted to 109 443.97, 2017 amounted to 79 521.42, 2018 amounted to 102 059.01 and in 2019 amounted to 86 799.89 at PT.PGN (Persero). Then with the resulting error analysis, the MAD value is 27741.58, the MSE value is 791516224.16 and the MAPE value is 27.18%.



Copyright © 2021. This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

I. PENDAHULUAN

Pada zaman yang berkembang sangat pesat ini dimana infrastruktur yang semakin banyak dan berkembang yang tentunya juga diikuti dengan meningkatnya energi yang dibutuhkan, salah satu energi tersebut yaitu gas alam. Gas alam merupakan salah satu jenis sumber energi yang berasal dari fosil tanaman, hewan, mikroorganisme yang tersimpan dibawah tanah selama ribuan bahkan jutaan tahun. Gas alam memiliki peran yang sangat penting bagi kehidupan masyarakat, terutama masyarakat Indonesia dalam hal mendukung perindustrian, pembangkit listrik, komersil, dan untuk kehidupan sehari-hari.

Salah satu perusahaan yang memproduksi, mentransmisi dan mendistribusi gas di Indonesia yaitu PT.PGN (Persero). Perusahaan mencatatkan volume penjualan yang selalu mengalami perubahan setiap tahunnya, hal tersebut tentu dapat membuat persediaan gas menjadi berlebihan atau kekurangan. Ketersediaan gas yang berlebihan ataupun kekurangan dapat membuat kurang maksimalnya kinerja perusahaan.

Dari permasalahan tersebut, diperlukannya suatu cara untuk memprediksi kebutuhan gas yang akan datang agar persediaan gas dapat mencukupi volume penjualan dan juga

persediaan gas tidak terlalu berlebihan. Salah satunya ialah dengan menggunakan model Regresi Linear Berganda. Model Regresi Linear ini digunakan untuk mengestimasi hubungan antara dua atau lebih *variable independent* (X) dan satu *variable dependent* (Y). Pada penelitian ini menggunakan Jumlah Pelanggan dan Nilai Penjualan gas sebagai *variable* yang mempengaruhi atau *independent* (X), sedangkan Volume Penjualan gas sebagai *variable* yang terpengaruhi atau *dependent* (Y).

Beberapa penelitian sudah pernah dilakukan untuk prediksi, hasil penelitian yang dilakukan Mona et al. (2015) dengan pengujian hipotesis menggunakan uji F diketahui bahwa jumlah produksi buah kelapa, biaya, jumlah pohon kelapa, luas lahan, dan jumlah anggota keluarga secara bersama-sama berpengaruh terhadap pendapatan petani kelapa, hal ini ditunjukkan dari nilai F hitung sebesar 85.075 dengan angka signifikansi sebesar 0.000. Nilai koefisien determinasi ganda yang dihasilkan adalah 0.918 dan nilai adalah 0.907 atau 90.7% [1].

Penelitian yang dilakukan oleh Wahyudin et al. (2020) mendapatkan hasil akhir dari algoritma regresi linear berganda dapat mengestimasi penjualan di tahun yang akan datang, dari hasil persamaan regresi yang didapat bahwa

penjualan mobil PT Astra Isuzu International Tbk-Isuzu di tahun 2020 adalah 12223 mobil, dimana angka penjualan tersebut turun sebanyak 1350 dibandingkan dengan tahun sebelumnya, dengan hasil perhitungan *error* estimasi dengan menggunakan metode MAD memiliki nilai yang relatif kecil yaitu 0,0607 [4].

Penelitian yang dilakukan oleh Sulistyono et al. (2017) mendapatkan hasil bahwa persamaan regresi yang mempengaruhi jumlah produksi adalah variable Kerusakan Mesin (KM) dan Harga Bahan Baku (HBB) serta Jumlah Tenaga Kerja (JTK) nilai konstanta 500308 menyatakan bahwa jika tidak ada variable kerusakan mesin, harga bahan baku dan jumlah tenaga kerja, maka jumlah produksi sebesar 500300. Dengan mengasumsikan diabaikannya *variable independent* lainnya, jika kedua variable Kerusakan Mesin (KM) bernilai positif sebesar 47869 dan Harga Bahan Baku (HBB) bernilai positif sebesar 7.2700000, maka jumlah produksi meningkat sebesar 1% dan jika variable Jumlah Tenaga Kerja (JTK) bernilai negative -3460, jumlah produksi mengalami penurunan 1% sebesar 3640 [3].

Adapun penelitian ini bertujuan untuk memprediksi volume penjualan gas PT PGN (Persero) sehingga dapat digunakan sebagai penunjang pengambilan keputusan dalam hal menentukan persediaan gas dan juga terkait ekspansi pembangunan pipa gas untuk memenuhi kebutuhan gas yang akan datang.

II. METODE

Tahapan penelitian yang dilakukan dalam penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining (CRISP-DM)*, berikut tahapannya:

A. Business Understanding

Pada tahapan pemahaman bisnis ada beberapa hal yang dilakukan antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan data mining.

B. Data Understanding

Pada tahapan pemahaman data, ada beberapa hal yang dilakukan antara lain, mengumpulkan data awal, mendeskripsikan data, mengeksplorasi data, dan membangun kerangka pemikiran.

C. Data Preparation

Setelah data dikumpulkan, data-data perlu diidentifikasi, dipilih, dibagi, kemudian dibangun dalam bentuk atau format yang diinginkan.

D. Modeling

Setelah data siap, tahap selanjutnya adalah melakukan modeling menggunakan aplikasi algoritma untuk mencari, mengidentifikasi, dan menampilkan pola. Pemilihan algoritma berdasarkan pada tipe data, beberapa algoritma umumnya digunakan untuk estimasi, prediksi, klasifikasi, clustering atau melihat hubungan asosiatif. Adapun dalam

Penelitian ini menggunakan algoritma regresi linear berganda yang bertujuan untuk melakukan prediksi.

E. Evaluation

Melakukan evaluasi akurasi prediksi menggunakan metode *Mean Absolute Deviation (MAD)*, *Mean Squared Error (MSE)*, dan *Mean Absolute Percent Error (MAPE)*. Hasil evaluasi dapat digunakan untuk mendapatkan kesimpulan dari kualitas hasil prediksi.

III. HASIL DAN PEMBAHASAN

A. Business Understanding

a. Menentukan Tujuan Bisnis

Hasil prediksi volume penjualan gas tahun 2016-2019 didapat dengan menggunakan variable bebas (*independent*) dan variable terikat (*dependen*). Manfaat dari hasil prediksi menggunakan algoritma regresi linear berganda menggunakan software R.Studio bagi perusahaan adalah untuk meminimalisir terjadinya kelebihan atau kekurangan persediaan gas.

b. Menilai Situasi

Penelitian ini berkaitan dengan penjualan gas di PT PGN (Persero). Dari data penjualan gas PT PGN (Persero) bahwa setiap tahunnya volume penjualan gas selalu mengalami perubahan, dikarenakan volume penjualan yang selalu berubah setiap tahunnya maka perlu dilakukannya prediksi volume penjualan yang akurat. Penelitian ini bertujuan untuk melakukan pengolahan data terhadap atribut antara lain, jumlah pelanggan, volume penjualan dan nilai penjualan.

c. Menentukan Tujuan Data Mining

Tujuan penelitian ini adalah untuk mengetahui cara penerapan algoritma regresi linear berganda dalam melakukan prediksi volume penjualan gas PT PGN (Persero) dan mengetahui hasil dari prediksi volume penjualan gas serta mengetahui kesalahan atau error prediksi menggunakan metode MAD, MSE, dan MAPE.

B. Data Understanding

a. Mengumpulkan Data Awal

Pengumpulan data awal dilakukan dengan observasi melalui website Badan Pusat Statistik, data yang didapat memiliki beberapa atribut, yaitu tahun, jumlah pelanggan, volume penjualan dan nilai penjualan.

Tabel 1. Banyaknya Pelanggan, Volume Penjualan, dan Nilai Penjualan Gas, 2012-2019

Tahun	Jumlah Pelanggan	Volume Penjualan (ribu m ³)	Nilai Penjualan (juta rupiah)
2012	19535	124765.88	524375.38
2013	19713	99202.16	300429.34
2014	19752	88709.29	414006.10
2015	19563	65146.96	388872.94
2016	20428	78616.78	506060.11
2017	20234	109093.38	570859.97
2018	19917	121742.17	672642.60

2019	25470	117683.89	676292.35
------	-------	-----------	-----------

b. Mendeskripsikan Data

Data yang didapat dari website Badan Pusat Statistik adalah data pada tahun 2012-2019 dengan menggunakan format excel, data hasil observasi mendapatkan 4 atribut, yaitu:

Tabel 2. Deskripsi Atribut

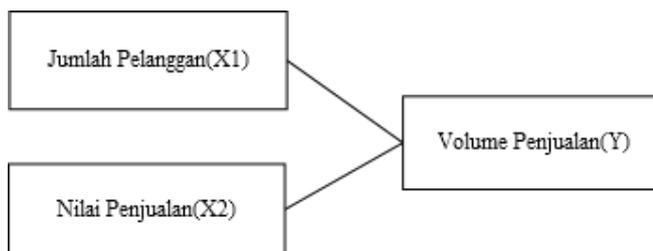
Atribut	Keterangan
Tahun	Keterangan data dalam tahunan
Jumlah Pelanggan	Banyaknya pelanggan dalam pembelian
Volume Penjualan	Jumlah total seluruh penjualan
Nilai Penjualan	Jumlah Nilai dari penjualan

c. Mengeksplorasi Data

Tabel 3. Eksplorasi Data

Atribut	Tipe Data	Keterangan
Tahun	Numeric	Keterangan data dalam tahunan
Jumlah Pelanggan	Numeric	Variabel Independent (X1)
Nilai Penjualan	Numeric	Variabel Independent (X2)
Volume Penjualan	Numeric	Variabel Dependen(Y)

Dari hasil eksplorasi data didapatkan kerangka pemikiran sebagai berikut :



Gambar 1. Kerangka Pemikiran

C. Data Preparation

Ada beberapa tahap persiapan data yang dilakukan, berikut Langkah-langkah yang dilakukan dalam penelitian ini:

a. Select Data

Tahap ini adalah proses untuk memilih data yang dibutuhkan untuk tujuan permodelan. Data yang digunakan adalah data pada Tabel 1.

b. Cleaning Data

pada Tabel 1 dapat dilihat dari masing-masing atribut, yaitu Tahun, Jumlah Pelanggan, Volume Penjualan, dan Nilai Penjualan. Dari atribut-atribut yang dipilih, dalam penelitian

ini akan menghapus atribut Tahun yang bernilai *None* (Tidak diketahui). Tujuan dari menghapus atribut tersebut agar hasil prediksi tidak terdapat nilai *None*.

b. Format Data

Tahap ini adalah menentukan atribut dalam data yang akan digunakan sebagai dataset akhir dalam proses permodelan data mining menggunakan software R.Studio, data yang digunakan sebagai berikut:

Tabel 3. Data Olah

X ₁ = Jumlah Pelanggan	X ₂ = Nilai Penjualan (juta rupiah)	Y = Volume Penjualan (ribu m ³)
19535	524375.38	124765.88
19713	300429.34	99202.16
19752	414006.10	88709.29
19563	388872.94	65146.96
20428	506060.11	78616.78
20234	570859.97	109093.38
19917	672642.60	121742.17
25470	676292.35	117683.89

c. Data Separation

Tahap ini merupakan tahap pembagian data menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membuat permodelan, sedangkan data uji digunakan untuk menguji permodelan. Data latih yang digunakan untuk permodelan diambil dari tahun 2012-2015 sebagai berikut:

Tabel 3. Data Latih

X ₁ = Jumlah Pelanggan	X ₂ = Nilai Penjualan (juta rupiah)	Y = Volume Penjualan (ribu m ³)
19535	524375.38	124765.88
19713	300429.34	99202.16
19752	414006.10	88709.29
19563	388872.94	65146.96

Sedangkan data uji diambil dari tahun 2016-2019 sebagai berikut:

Tabel 4. Data Uji

X ₁ = Jumlah Pelanggan	X ₂ = Nilai Penjualan (juta rupiah)	Y = Volume Penjualan (ribu m ³)
20428	506060.11	78616.78
20234	570859.97	109093.38
19917	672642.60	121742.17
25470	676292.35	117683.89

D. Modeling

Sebelum melakukan prediksi, peneliti terlebih dahulu membentuk model yang nantinya akan digunakan sebagai argument object dalam memprediksi Volume Penjualan Gas pada software R.Studio, Adapun modelnya dapat dilihat pada Persamaan 1:

$$model = lm(Y \sim X_1Latih + X_2Latih) \quad (1)$$

Model yang dibentuk pada Persamaan 1 menggunakan fungsi *lm* (*Linear Models*). Variable yang digunakan

menggunakan dataGasLatih pada Tabel 3. Selanjutnya dilakukan pembentukan fungsi *predict* untuk melakukan prediksi, seperti terlihat pada Persamaan 2:

$$prediksiY = predict(model, dataUji) \tag{2}$$

Dalam fungsi *predict* juga terdapat Data Uji, data tersebut terlihat seperti pada Tabel 4, tetapi atribut yang digunakan yaitu “Jumlah Pelanggan” dan “Nilai Penjualan”. Hasil yang didapat dari perhitungan komputasi prediksi yaitu sebagai berikut:

Tabel 5. Prediksi_VolumePenjualan

Tahun	Hasil Prediksi
2016	109443.97
2017	79521.42
2018	102059.01
2019	86799.89

Selanjutnya hasil pada Tabel 5 akan disisihkan dengan atribut “Volume Penjualan” pada Tabel 4 yang tujuannya untuk mengetahui selisih kesalahan (*error*) dari prediksi. Adapun rumusnya pada Persamaan 3:

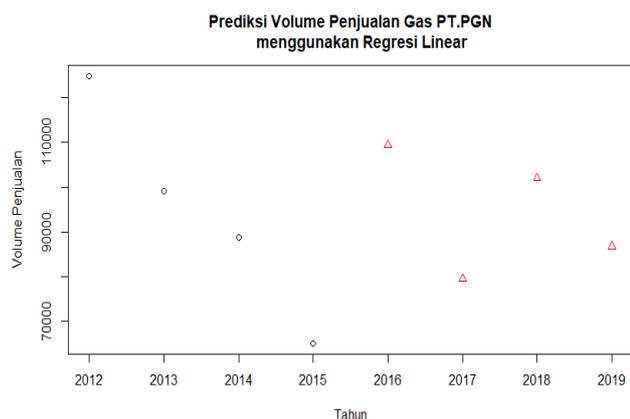
$$selisih = abs(prediksiY - aktualY) \tag{3}$$

Hasil yang didapatkan dari perhitungan selisih dapat dilihat pada Tabel 6 dibawah ini :

Tabel 6. Selisih

Tahun	Selisih
2016	30827.19
2017	29571.96
2018	19683.16
2019	30884.00

Berikut adalah plot grafik yang menggambarkan data aktual dari tahun 2012-2015 dan hasil prediksi dari tahun 2016-2019:



Gambar 2. Plot Grafik

Dapat dilihat pada gambar 5 terdapat dua titik yang berbeda, yaitu titik hitam dan titik merah, titik hitam megambarkan data actual (tahun 2012-2015), sedangkan titik merah

mengambarkan hasil prediksi (2016-2019). Jika dilihat volume penjualan pada tahun 2016 mendapatkan hasil 109443.97 (ribu m³), tahun 2017 mendapatkan hasil sebesar 79521.42 (ribu m³), tahun 2018 mendapatkan hasil sebesar 102059.01 (ribu m³), sedangkan pada tahun 2019 mendapatkan hasil sebesar 86799.89 (ribu m³).

E. Evaluation

Tahap terakhir adalah melakukan evaluasi untuk mengetahui akurasi dari prediksi, karena dalam melakukan prediksi terjadi ketidaksesuaian antara data aktual dengan data hasil prediksi. Tidak ada prediksi yang pasti akurat meskipun menggunakan berbagai macam metode prediksi [2]. Oleh karena itu dilakukan perhitungan kesalahan atau error yang terjadi dengan menggunakan metode analisis kesalahan *Mean Absolute Deviation* (MAD), *Mean Squared Error* (MSE), dan *Mean Absolute Percent Error* (MAPE). berikut penjelasannya:

a. MAD (*Mean Absolute Deviation*), adalah rata-rata kesalahan mutlak selama periode tertentu tanpa memperhatikan apakah hasil prediksi lebih besar atau lebih kecil dibanding aktualnya, dengan kata lain MAD adalah rata-rata dari nilai absolut simpangan. Adapun rumusnya seperti pada Persamaan 4:

$$\sum \frac{|At - Ft|}{n} \tag{4}$$

b. MSE (*Mean Square Error*), adalah rata-rata kesalahan kuadrat yang dihitung dengan semua kesalahan peramalan pada setiap periode dan membagi dengan jumlah periode peramalan. Secara sistematis MSE dirumuskan seperti pada Persamaan 5:

$$\sum \frac{(At - Ft)^2}{n} \tag{5}$$

c. MAPE (*Mean Absolute Percentage Error*), adalah rata-rata presentase kesalahan mutlak. MAPE biasanya lebih berarti dibandingkan MAD karena MAPE menyatakan persentase kesalahan hasil peramalan terhadap volume penjualan aktual selama periode tertentu yang akan memberikan informasi persentase kesalahan, yang kemudian dikalikan 100% agar mendapatkan hasil secara presentase. Secara sistematis MAPE dirumuskan seperti pada Persamaan 6:

$$\left(\frac{100}{n}\right) \sum \left|At - \frac{Ft}{At}\right| \tag{6}$$

Hasil yang didapatkan dari perhitungan kesalahan atau *error* menggunakan software R.Studio pada penelitian ini sebagai berikut:

Tabel 7. Hasil Perhitungan Error

Metode	Hasil
MAD	27741.58

MSE	791516224.16
MAPE	27.18%

Dapat dilihat pada Tabel 7 bahwa nilai MAD yang didapatkan artinya rata-rata kesalahan baik itu kelebihan atau kekurangan dari nilai aktualnya adalah sebesar 27741.58, Sedangkan nilai MSE artinya rata-rata kesalahan berpangkat sebesar 791516224.16, dan nilai MAPE menunjukkan rata-rata kesalahan dalam bentuk presentase sebesar 27.18%. Kesalahan atau error yang dihasilkan cukup tinggi karena dipengaruhi oleh volatilitas variable independent(X) Nilai Penjualan data latih yang cukup tinggi setiap tahun.

Hasil perhitungan dengan menggunakan metode ini diharapkan dapat membantu perusahaan dalam merencanakan persediaan gas yang akan datang untuk memperkecil ketidaksesuaian antara persediaan gas dengan volume penjualan gas yang terjadi. Sehingga perencanaan gas tidak hanya berdasarkan perkiraan saja tanpa ada metode dasar yang dapat menyebabkan ketidaksesuaian antara perencanaan persediaan gas yang telah ditentukan pada awal tahun dengan volume penjualan yang terjadi.

IV. KESIMPULAN

Dari hasil perhitungan prediksi menggunakan algoritma Regresi Linear Berganda pada software R.Studio, didapatkan hasil volume penjualan tahun 2016 sebesar 109443.97, tahun 2017 sebesar 79521.42, tahun 2018 sebesar 102059.01 dan tahun 2019 sebesar 86799.89 di PT PGN (Persero). Kemudian dengan analisis kesalahan yang dihasilkan yaitu nilai MAD sebesar 27741.58, nilai MSE sebesar 791516224.16 dan nilai MAPE sebesar 27.18%.

Hasil perhitungan dengan menggunakan metode ini diharapkan dapat membantu perusahaan dalam merencanakan persediaan gas untuk memperkecil ketidaksesuaian antara persediaan gas dengan volume penjualan yang terjadi. Juga dapat membantu dalam pengambilan keputusan terkait ekspansi pembangunan pipa gas untuk memenuhi kebutuhan gas yang akan datang.

DAFTAR PUSTAKA

- [1] MG. Mona, JS. Kekenusa, JD. Prang, "Penggunaan Regresi Linear Berganda untuk Menganalisis Pendapatan Petani Kelapa Studi Kasus: Petani Kelapa di Desa Beo, Kecamatan Beo Kabupaten Talaud" *d'Cartesian: Jurnal Matematika dan Aplikasi*, vol. 4, no. 2, Sept. 2015.
- [2] A. Saputro, and B. Purwanggono, "Peramalan Perencanaan Produksi Semen dengan Metode Exponential Smoothing pada PT. Semen" *Indonesia Industrial Engineering Online Journal*, vol. 5, no. 4, Nov. 2016.
- [3] Sulistyono, and W. Sulistiyowati, "Peramalan Produksi dengan Metode Regresi Linier Berganda" *Prozima*, vol. 1, no. 2, Dec. 2017.
- [4] AAN. Wahyudin, A. Primajaya, ASY. Irawan, "Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu" *Techno.COM*, vol.19, no.4, November 2020.