

Prediksi Jumlah Tagihan Air Pdam Tirta Kualo Menggunakan Metode Regresi Sederhana

Isdalina¹, Putri Indriani², Saddam Adnan Manurung³

^{1,2,3} Mahasiswa Program Studi Sistem Informasi, Universitas Royal

¹isdalinaisda@gmail.com; ²putri.indriani692479@gmail.com; ³saddammanurung344@gmail.com

* putri.indriani692479@gmail.com

INFO ARTIKEL

ABSTRAK

PDAM Tirta Kualo adalah perusahaan wilayah yang menyuplai air bersih untuk masyarakat sekitar. Perkiraan jumlah tagihan air yang tepat sangat krusial untuk membantu PDAM dalam mengatur sumber daya dan keuangan dengan efisien. Studi ini bertujuan untuk menciptakan model prediksi total tagihan air dengan menggunakan metode regresi linear. Data yang dimanfaatkan merupakan data historis tagihan pelanggan yang dianalisis untuk mengidentifikasi keterkaitan antara volume pemakaian air dan total tagihan.

Hasil penelitian menunjukkan bahwa model regresi sederhana mampu memprediksi jumlah tagihan air dengan tingkat akurasi yang sangat tinggi, yaitu 0.9926. Dengan akurasi tersebut, model ini dapat digunakan secara efektif dalam perencanaan keuangan PDAM serta membantu pelanggan dalam mengestimasi konsumsi air mereka.

Kata kunci: prediksi tagihan air, PDAM, regresi sederhana, analisis data.



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

Copyright ©2024 by Author. Published by PT Beranda Teknologi Academia

ABSTRACT

PDAM Tirta Kualo is a regional company that supplies clean water to the surrounding community. Accurate estimation of water bill amounts is crucial to assist PDAMs in managing resources and finances efficiently. This study aims to create a prediction model for total water bills using the linear regression method. The data used is historical customer billing data which is analyzed to identify the relationship between air usage volume and total billing.

The findings show that a simple regression model can describe the water bill amount with an impressive accuracy of 0.9926. This precise model allows it to be used effectively in PDAM financial planning and assists customers in estimating their water usage.

Keywords: water bill prediction, PDAM, simple regression, data analysis.

PENDAHULUAN

Air bersih adalah salah satu kebutuhan utama yang sangat krusial untuk kehidupan manusia [1]. PDAM Tirta Kualo sebagai penyedia layanan air bersih memiliki tanggung jawab untuk menjamin ketersediaan air bersih kepada masyarakat dengan kualitas yang baik dan harga yang terjangkau. PDAM Tirta Kualo adalah perusahaan daerah yang memberikan air bersih kepada orang-orang di sekitarnya. Dalam proses pengelolaannya, PDAM menghadapi kendala berupa efek negatif pemakaian air oleh pelanggan yang berdampak langsung pada total tagihan yang diterima. Oleh karena itu, diperlukan sebuah sistem yang dapat menyampaikan total tagihan air dengan tepat untuk mendukung perencanaan serta pengambilan keputusan yang lebih efektif [2].

Pembelajaran mesin, sebagai salah satu sektor kecerdasan buatan, telah diterapkan secara luas dalam berbagai bidang, termasuk dalam prediksi yang mengandalkan data masa lalu. Regresi sederhana, sebuah metode dalam pembelajaran terawasi, adalah teknik yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen [3]. Dalam penelitian ini, volume penggunaan air dijadikan variabel independen, sementara jumlah tagihan air menjadi variabel dependen yang akan diprediksi. Metode ini dipilih karena karakteristiknya yang sederhana dan penjelasannya yang mudah dipahami oleh pihak pengelola PDAM [4].

Dalam penelitian ini, data yang digunakan merupakan data historis tagihan air pelanggan yang diperoleh dari PDAM Tirta Kualo. Data ini berisi jumlah penggunaan air bulanan serta total tagihan yang dipesan oleh pelanggan. Data tersebut selanjutnya dianalisis dengan metode regresi sederhana guna memodelkan hubungan antara total pemakaian air dan jumlah tagihan. Tujuan penelitian ini adalah untuk mengembangkan model prediksi jumlah tagihan air menggunakan machine learning yang dapat menghasilkan hasil yang tepat dengan tingkat kesalahan yang minimal. Hasil penelitian ini diharapkan dapat mendukung PDAM Tirta Kualo dalam mengatur keuangan dan sumber daya air dengan lebih efektif [5].

METODE

Pada penelitian ini menggunakan machine learning sebagai dasar keilmuan. Machine learning merupakan adalah mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunaannya [6]. Adapun definisi dari machine learning itu sendiri adalah cabang kecerdasan buatan (AI) yang memungkinkan sistem komputer untuk belajar dan melakukan tugas kompleks secara mandiri.

Model regresi adalah salah satu model dalam memprediksi variabel terikat (dependen) melalui variabel bebas independen [7]. Tujuan dari analisis regresi ini adalah untuk mengetahui pola ikatan matematis antara parameter X dengan parameter Y, untuk mengetahui seberapa besar pengaruh parameter X terhadap parameter Y, dan untuk memprediksi nilai parameter Y setelah mengetahui nilai parameter X [8]. Dalam regresi sederhana, prinsip utama adalah bahwa parameter dependen (Y) dan parameter independen (X) harus menunjukkan adanya hubungan sebab akibat atau kausalitas. Ini didasarkan pada teori, temuan penelitian sebelumnya, atau penjelasan logistik khusus. Model regresi ini banyak diterapkan di berbagai bidang, termasuk ekonomi. Model regresi sederhana atau lebih sering dikenal dengan regresi linier sederhana adalah regresi yang sering digunakan dalam kehidupan sehari-hari untuk melihat pengaruh variabel bebas (X) terhadap variabel terikat (Y). Persamaan regresi linier sederhana adalah sebagai berikut:

$$Y = a + bX \text{-----}(1)$$

Y = Parameter yang tergantung (parameter yang terikat)

X = Parameter yang tidak bergantung (parameter yang bebas)

a = Tetapan (nilai y ketika x = 0)

b = Bilangan regresi (dampak positif (+) atau negatif (-))

Data Understanding

Data understanding adalah tahapan dalam metodologi sains data dan pengembangan AI untuk memahami data yang akan digunakan untuk memecahkan masalah [9]. Data understanding dilakukan dengan mengeksplorasi data dan melakukan persiapan data, Tujuan data understanding adalah untuk: Mengidentifikasi masalah pada data, Membuat ringkasan data, Menentukan kualitas data, Mengetahui kesesuaian data dengan masalah bisnis [10]. Berikut ini adalah data understanding pada dataset perediksi jumlah tagihan air PDAM tirta kualo menggunakan metode regresi sederhana ialah sebagai berikut:

	Pelanggan No.	Samb - Nama	Jenis	Rek	Awal	Akhir	Pakai	Harga_Air	ADM	DPM	Meterai	Jumlah
0	01010002	- PIKER SINAGA	R2	0	0	25	25	96290	5000	0	0	101290
1	01010004	- ZULHAM LBS	R2	0	0	30	30	120770	5000	0	0	125770
2	01010005	- BURHANUDDIN NASUTION	R2	0	0	25	25	96290	5000	0	0	101290
3	01010008	- A. NUHIN/RIDWAN NST	R2	0	0	20	20	71810	5000	0	0	76810
4	01010009	- GAZALI	R2	0	0	10	10	32640	5000	0	0	37640
...	...	...	...	...	...	...	...	...	...	...	...	...
295	01010625	- MAS BULAN	R2	0	235	249	14	48308	5000	0	0	53308
296	01010626	- ARMAJOSNIDA FAJRINI	R2	0	391	425	34	140354	5000	0	0	145354
297	01010627	- JULIANDY DALIMUNTHE	R2	0	0	22	22	81602	5000	0	0	86602
298	01010628	- MUHAMMAD AL ARIS	R2	0	0	10	10	32640	5000	0	0	37640
299	01010629	- CANDRA GUFTA TOHA AZHAR	R2	0	0	20	20	71810	5000	0	0	76810

300 rows × 11 columns

Gambar 1. Deskripsi Data

Data Preparation

Data preparation adalah proses menyiapkan data mentah agar bisa diproses dan dianalisis lebih lanjut, Data preparation penting untuk dilakukan sebelum data diproses dan dianalisis. Proses ini membantu untuk menghilangkan bias yang disebabkan oleh kualitas data yang buruk [11]. Berikut ini adalah proses melakukan data preparation:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Pelanggan No. Samb - Nama  300 non-null    object
 1   Jenis                  300 non-null    object
 2   Rek                    300 non-null    int64
 3   Awal                   300 non-null    int64
 4   Akhir                  300 non-null    int64
 5   Pakai                   300 non-null    int64
 6   Harga_Air               300 non-null    int64
 7   ADM                     300 non-null    int64
 8   DPM                     300 non-null    int64
 9   Meterai                 300 non-null    int64
10  Jumlah                  300 non-null    int64
dtypes: int64(9), object(2)
memory usage: 25.9+ KB
```

Gambar 2. Deskripsi Tipe Data

Modelling

Modelling adalah proses membuat representasi matematis atau statistik dari hubungan antara variabel-variabel yang ada dalam sebuah data [12]. Dalam konteks regresi sederhana, pemodelan bertujuan untuk menggambarkan hubungan linier antara variabel independen (X) dan variabel dependen (Y) dalam persamaan bentuk linier. di mana hubungan antara variabel divisualisasikan dan dirangkum dalam bentuk persamaan matematis yang dapat digunakan untuk berbagai keperluan analisis atau prediksi [13].

Evaluation

Evaluation dalam metode regresi sederhana adalah proses menilai dan mengukur kinerja model yang telah dibuat untuk memastikan bahwa model tersebut sesuai, akurat, dan mampu memberikan

prediksi yang Andal. Proses evaluasi dilakukan menggunakan metrik tertentu dan uji statistik untuk menilai hubungan antara variabel independen (X) dan tergantung (Y) [14].

Depeloment

Depeloment adalah proses pembuatan laporan atau artikel jurnal menggunakan hasil penelitian. Hasil dari penerapan metode regresi sederhana di interprestasikan dalam Bahasa pemrograman yang memudahkan Perusahaan PDAM Tirta Kualo untuk membaca serta memahami informasinya dengan lebih cepat dan efektif [15].

HASIL DAN PEMBAHASAN

Tahap pertama adalah import dataset yang digunakan untuk diprediksi.

```
In [1]: #Load library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#Load dataset
df = pd.read_excel('pelanggan_PDAM.xlsx')

Out[1]:
```

	Pelanggan No.	Samb	Nama	Jenis	Rek	Awal	Akhir	Pakai	Harga_Air	ADM	DPM	Meterai	Jumlah
0	01010002	-	PIKER SINAGA	R2	0	0	25	25	96290	5000	0	0	101290
1	01010004	-	ZULHAM LBS	R2	0	0	30	30	120770	5000	0	0	125770
2	01010005	-	BURHANUDDIN NASUTION	R2	0	0	25	25	96290	5000	0	0	101290
3	01010008	-	A. NUHIN/RIDWAN NST	R2	0	0	20	20	71810	5000	0	0	76810
4	01010009	-	GAZALI	R2	0	0	10	10	32640	5000	0	0	37640
...	...	...	...	...	...	...	...	...	...	...	...	...	...
295	01010625	-	MAS BULAN	R2	0	235	249	14	48308	5000	0	0	53308
296	01010626	-	ARMAJOSNIDA FAJRINI	R2	0	391	425	34	140354	5000	0	0	145354
297	01010627	-	JULIANDY DALIMUNTHE	R2	0	0	22	22	81602	5000	0	0	86602
298	01010628	-	MUHAMMAD AL ARIS	R2	0	0	10	10	32640	5000	0	0	37640
299	01010629	-	CANDRA GUFTATOHAZHAR	R2	0	0	20	20	71810	5000	0	0	76810

300 rows x 11 columns

Gambar 3. Deskripsi 5 Data Awal

Tahap kedua adalah menampilkan statistik deskriptif data

```
In [5]: #statistik deskriptif data
df.describe()

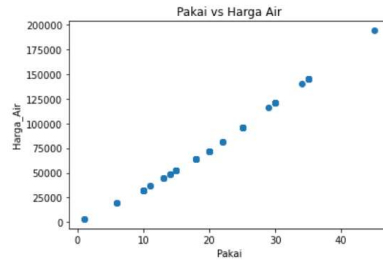
Out[5]:
```

	Rek	Awal	Akhir	Pakai	Harga_Air	ADM	DPM	Meterai	Jumlah
count	300.0	300.000000	300.000000	300.000000	300.000000	300.0	300.0	300.0	300.000000
mean	0.0	154.163333	173.583333	19.420000	71900.826667	5000.0	0.0	0.0	76900.826667
std	0.0	737.175698	736.955417	7.034811	30911.999468	0.0	0.0	0.0	30911.999468
min	0.0	0.000000	7.000000	1.000000	3264.000000	5000.0	0.0	0.0	8264.000000
25%	0.0	0.000000	15.000000	15.000000	52225.000000	5000.0	0.0	0.0	57225.000000
50%	0.0	0.000000	20.000000	20.000000	71810.000000	5000.0	0.0	0.0	76810.000000
75%	0.0	0.000000	25.000000	25.000000	96290.000000	5000.0	0.0	0.0	101290.000000
max	0.0	6744.000000	6769.000000	45.000000	194210.000000	5000.0	0.0	0.0	199210.000000

Gambar 4. Statistik Deskriptif Data

Tahap ketiga adalah tahap analysis (analisa hubungan antar variabel) jenis dengan harga_air menggunakan scatter plot.

```
In [8]: #bivariate analysis (analisa hubungan antar variabel) pakai dengan harga_air menggunakan scatter plot
import matplotlib.pyplot as plt
plt.scatter(df['Pakai'], df['Harga_Air'])
plt.xlabel('Pakai')
plt.ylabel('Harga_Air')
plt.title('Pakai vs Harga Air')
plt.show()
```

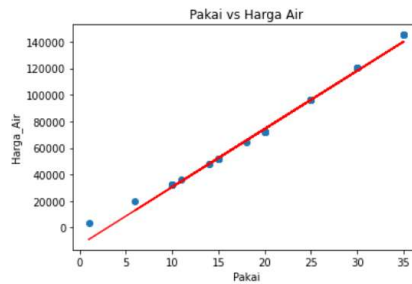


Gambar 6. Analisa Hubungan Antar Variable

Tahap ke empat adalah visualisasi aktual vs prediksi menggunakan data testing. Garis merah merupakan garis regresi dari persamaan yang kita dapat tadi

```
In [21]: #visualisasi aktual vs prediksi menggunakan data testing
#garis merah merupakan garis regresi dari persamaan yang kita dapat tadi

y_prediksi = lin_reg.predict(x_test)
plt.scatter(x_test, y_test)
plt.plot(x_test, y_prediksi, c='r')
plt.xlabel('Pakai')
plt.ylabel('Harga_Air')
plt.title('Pakai vs Harga Air')
plt.show()
```



Gambar 7. Analisis Visualisasi Actual vs Prediksi

KESIMPULAN

Berdasarkan analisis prediksi jumlah tagihan air PDAM Tirta Kualo dengan metode regresi sederhana, dapat disimpulkan bahwa terdapat hubungan antara jumlah penggunaan air dan tagihan air, di mana semakin besar penggunaan air, semakin tinggi pula tagihannya. Model regresi sederhana dapat dimanfaatkan untuk memperkirakan total tagihan air dengan cukup efektif. Model ini berguna bagi PDAM dalam merencanakan keuangan dan mengestimasi pendapatan berdasarkan konsumsi air. Agar mendapatkan hasil yang lebih tepat, studi berikutnya sebaiknya mempertimbangkan faktor-faktor tambahan seperti jumlah konsumen atau perubahan tarif air.

DAFTAR PUSTAKA

- [1] D. Christalivea and M. A. I. Pakereng, "Prediksi Volume Penyaluran Air Minum Kota Salatiga

- Tahun 2021 Berdasarkan Time Data Series Menggunakan Regresi Linear,” *J. EMT KITA*, vol. 7, no. 2, pp. 422–429, 2023, doi: 10.35870/emt.v7i2.1031.
- [2] R. Prihatini and A. Sugiarto, “Edukasi Pengamanan Air Bersih Masyarakat Pelanggan PDAM Tirta Kualo Sebagai Perencanaan Wilayah Kota Tanjungbalai,” *J. Pengabd. Kpd. Masy. Nasant.*, vol. 4, no. 3, pp. 2494–2498, 2023.
 - [3] Nurnawaty, F. Latif, E. S. Syawal, and A. . R. Mukhtar, “Implementasi Metode Regresi Linier Sederhana pada Penyajian Hasil Prediksi Pemakaian Air Bersih pada PDAM Kec. Panakkukang Kota Makassar,” *Kohesi J. Multidisiplin Saintek*, vol. 01, no. 04, 2023.
 - [4] C. E. Simbolon, “Penerapan Algoritma Regresi Linier Sederhana Dalam Memprediksi Keuntungan dan Kerugian Kelapa Sawit Pt . Sri Ulina,” *J. Inf. Syst. Res.*, vol. 2, no. 2, pp. 169–172, 2021.
 - [5] R. Eryanto, M. Masril, and F. Herista, “Analisis Kebutuhan Air Bersih Pdam Kota Sawahlunto,” *Ensiklopedia Res. Community Serv. Rev.*, vol. 1, no. 1, pp. 27–37, 2021, doi: 10.33559/err.v1i1.963.
 - [6] Y. Herdiana, “Penerapan Machine Learning Dengan Model Linear Regression Terhadap Analisis Kualitas Hasil Petik the Di Pt. Perkebunan ...,” *Comput. J. Inform.*, vol. 09, pp. 1–9, 2022, [Online]. Available: <https://unibba.ac.id/ejournal/index.php/computing/article/view/855%0Ahttps://unibba.ac.id/ejournal/index.php/computing/article/download/855/710>
 - [7] A. S. Syuhada, A. M. Simanullang, D. S. Lewa, and S. Jefry Marthin, “Makalah Pembelajaran Mesin (Machine Learning),” *Makal. Pembelajaran Mesin (machine Learn.)*, pp. 1–11, 2021.
 - [8] C. Novianti, B. Sadipun, and J. M. Balan, “Pengaruh Motivasi Belajar Terhadap Hasil Belajar Matematika Peserta Didik,” *Sci. Phys. Educ. J.*, vol. 3, no. 2, pp. 57–75, 2020, doi: 10.31539/spej.v3i2.992.
 - [9] M. N. Fahmi, “Implementasi Mechine Learning menggunakan Python Library : Scikit-Learn (Supervised dan Unsupervised Learning),” *Sains Data J. Stud. Mat. dan Teknol.*, vol. 1, no. 2, pp. 87–96, 2023, doi: 10.52620/sainsdata.v1i2.31.
 - [10] D. A. Dzulhijjah, M. B. Herlambang, M. Haifan, P. Studi, P. Insinyur, and T. Selatan, “Implementasi Framework CRISP-DM untuk Proses Data Mining Aplikasi Credit Scoring PT.XYZ,” *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, vol. 1, no. 1, pp. 238–251, 2023.
 - [11] F. Budiman and Y. M. Awaludin, “Optimasi Analisis Kesuburan Tanah dengan Pendekatan Soft Voting Ensemble,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 261–276, 2023, doi: 10.24176/simet.v14i2.11285.
 - [12] N. Supriadi, A. L. Sari, and A. R. JL, “Analisis Hubungan Self-Efficacy dan Representasi Matematis terhadap Pemecahan Masalah Matematis,” *PYTHAGORAS J. Pendidik. Mat.*, vol. 18, no. 2, pp. 148–158, 2023, doi: 10.21831/pythagoras.v18i2.64588.
 - [13] R. Andini, “Analisa Perbandingan Metode Regresi Linier dan Regresi Non Linier Berdasarkan Data Kebutuhan Listrik Di Provinsi Jawa Tengah Pada Tahun 2021 dan 2023,” vol. 17, no. 1, pp. 30–40, 2024.
 - [14] Harsiti, Z. Muttaqin, and E. Srihartini, “Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet,” *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 12–16, 2022, doi: 10.30656/jsii.v9i1.4426.
 - [15] D. S. Purba, V. Tarigan, and W. J. Tarigan, “Analisis Kinerja PDAM Di Sumatera Utara Ditinjau Dari Aspek Keuangan,” *J. Bisnis dan Kewirausahaan*, vol. 17, no. 2, pp. 200–218, 2021, doi:

10.31940/jbk.v17i2.2542.