

RANCANG BANGUN APLIKASI PREDIKSI TAGIHAN AIR BERBASIS WEB MENGUNAKAN REGRESI LINIER BERGANDA

Arum Fatmawati^{1*}; Afril Efan Pajri²; Sahri³

Teknik Informatika^{1,3}, Sistem Komputer²
Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia^{1,2,3}
<http://unugiri.ac.id>^{1,2,3}
arumfatma16@gmail.com^{1*}, afril@unugiri.ac.id², sahriunugiri@gmail.com³

(*) Corresponding Author



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-Non Komersial 4.0 Internasional.

Abstract—The management of water billing in the PAMSIMAS service in Sidobandung Village is still conducted manually and does not provide early information regarding bill estimates, often resulting in delayed payments by customers. This study aims to design and develop a web-based water bill prediction application using the Multiple Linear Regression (MLR) method, capable of delivering fast, accurate, and accessible billing estimates. The dataset used in this research consists of historical monthly water usage and billing data from January to December 2024, with a structure comprising 231 rows of customer data and 30 feature columns. The research stages include data preprocessing, model training using MLR, integration of the model into a web-based system, and evaluation of prediction results using the Mean Squared Error (MSE) and R-squared (R^2) metrics. Evaluation results showed that the model achieved an MSE of 18,882 and an R^2 of 0,8, indicating a fairly good and stable prediction performance. The system allows customers to log in, view predicted water bills for the 13th month based on previous data, and access graphical visualizations of usage and cost trends. Meanwhile, the admin can efficiently manage customer data through a dedicated dashboard. With the implementation of this application, the management and prediction process of water billing becomes more transparent, efficient, and helps customers in planning their water expenses more precisely.

Keywords: MSE, Multiple Linear Regression, R^2 , Water Bill Prediction, Web Application

Abstrak—Pengelolaan tagihan air pada layanan PAMSIMAS di Desa Sidobandung masih dilakukan secara manual dan belum menyediakan informasi awal mengenai estimasi tagihan, sehingga seringkali menyebabkan keterlambatan pembayaran oleh pelanggan. Penelitian ini bertujuan untuk merancang dan membangun aplikasi prediksi tagihan air berbasis *website* menggunakan metode Regresi Linier Berganda (RLB) yang mampu memberikan estimasi tagihan air secara cepat, akurat, dan dapat diakses oleh masyarakat. *Dataset* yang digunakan merupakan data historis pemakaian air dan biaya tagihan bulanan pelanggan dari bulan Januari hingga Desember tahun 2024, dengan struktur terdiri atas 231 baris data pelanggan dan 30 kolom fitur. Tahapan penelitian meliputi *preprocessing* data, pelatihan model RLB, evaluasi hasil prediksi menggunakan metrik *Mean Squared Error* (MSE) dan *R-squared* (R^2), serta integrasi model ke dalam sistem *web*. Hasil evaluasi menunjukkan bahwa model memiliki nilai MSE sebesar 18,882 dan R^2 sebesar 0,8, yang mengindikasikan performa prediksi yang cukup baik dan stabil. Sistem ini memungkinkan pelanggan untuk *login*, melihat hasil prediksi tagihan bulan ke-13 berdasarkan data bulan sebelumnya, serta mengakses grafik visualisasi tren pemakaian air dan biaya. Sementara itu, *admin* dapat mengelola data pelanggan secara efisien melalui *dashboard* khusus. Dengan adanya aplikasi ini, proses pengelolaan dan prediksi tagihan air menjadi lebih transparan, efisien, dan membantu pelanggan dalam merencanakan pengeluaran air secara lebih terukur.

Kata kunci: MSE, Multiple Linear Regression, R^2 , Prediksi Tagihan Air, Aplikasi Web

PENDAHULUAN

Kebutuhan akan air bersih merupakan hal mendasar bagi masyarakat, khususnya di wilayah pedesaan yang masih memiliki keterbatasan akses terhadap layanan air (Jannah et al., 2021). Program Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat (PAMSIMAS) hadir sebagai solusi untuk meningkatkan pelayanan air bersih melalui partisipasi aktif masyarakat. Salah satu desa yang telah mengimplementasikan program ini adalah Desa Sidobandung, Kecamatan Balen, Kabupaten Bojonegoro. Namun, dalam pelaksanaannya, pengelolaan tagihan air masih dilakukan secara manual, tanpa adanya sistem prediksi tagihan bulanan yang dapat membantu pelanggan merencanakan pembayaran.

Ketidaktelitian dalam memperkirakan jumlah tagihan air seringkali menyebabkan keterlambatan pembayaran, munculnya tunggakan, serta kesulitan dalam manajemen distribusi air bersih oleh pengelola. Oleh karena itu, diperlukan sistem prediksi yang dapat memperkirakan jumlah tagihan berdasarkan pola pemakaian sebelumnya secara akurat, efisien, dan mudah diakses. Pentingnya penerapan *forecasting* dalam sistem pelayanan telah banyak dibuktikan dalam penelitian sebelumnya. Penelitian yang dilakukan oleh Dani et al. menegaskan bahwa proses peramalan yang akurat berperan penting dalam mendukung pengambilan keputusan yang lebih tepat dan efisien di berbagai bidang (Deni et al., 2023).

Salah satu pendekatan yang banyak digunakan dalam prediksi berbasis data historis adalah metode Regresi Linier Berganda. Regresi linier berganda merupakan metode yang sering digunakan dalam prediksi berbasis data numerik, seperti pada penelitian oleh Hariyanto et al. yang memprediksi harga ponsel berdasarkan spesifikasi teknis dan memperoleh hasil prediksi yang cukup akurat menggunakan evaluasi RMSE (Hariyanto et al., 2025). Metode ini memungkinkan pemodelan hubungan antara beberapa variabel *independen*, seperti volume pemakaian air tiap bulan, dengan variabel *dependen* berupa total biaya tagihan. Sejumlah penelitian sebelumnya telah memanfaatkan metode ini dalam berbagai konteks. Penelitian yang dilakukan oleh Prasetyo et al. memprediksi hasil produksi kelapa sawit pada PT Perkebunan Nusantara I menggunakan RLB dengan nilai MAPE sebesar 14,28%, yang membantu perusahaan dalam merumuskan strategi produksi (Prasetyo et al., 2021). Penelitian yang dilakukan oleh Salma et al. memprediksi dosis koagulan optimal pada Perumda Air Minum Tirta Moedal menggunakan Regresi Linear Berganda. Hasil

penelitian menunjukkan efisiensi pencapaian 94,72%, yang mendukung perusahaan dalam menjaga kualitas air bersih untuk masyarakat Kota Semarang (Salma Rahima Ahmad, 2023).

Penelitian lain oleh Diah Retnowati et al. memanfaatkan RLB untuk menganalisis permintaan air PDAM di Kota Purwokerto dengan mempertimbangkan variabel pendapatan keluarga, jumlah anggota keluarga, serta biaya pemakaian air (Diah Retnowati, Eli Susanti, Ratna Puji Astuti, Sodik Dwi Purnomo, Herwiek Diah Lestari, 2023). Penelitian serupa juga dilakukan oleh Dimas Dwi Al Hakim di Surabaya (Dimas Dwi Al Hakim, 2023) yang menunjukkan hubungan variabel harga air, pendapatan, dan jumlah anggota keluarga dengan tingkat permintaan air. Sementara itu, penelitian oleh Nandita Afrilia et al. menggunakan RLB untuk memprediksi hasil panen wortel yang diintegrasikan ke dalam sistem berbasis web dengan tingkat akurasi tinggi (Nandita Afrilia S, Fathia Frazna Az-Zahra, 2024). Selain itu, Jafar et al. (Jafar et al., 2023) serta Fernandes et al. (Fernandes et al., 2023) menerapkan regresi linier pada prediksi kualitas air, masing-masing memperoleh nilai R^2 yang sangat baik, menunjukkan keandalan metode ini dalam memodelkan hubungan antar variabel untuk tujuan prediksi.

Penelitian ini bertujuan untuk merancang dan membangun sistem prediksi tagihan air berbasis *website* dengan menggunakan metode Regresi Linier Berganda. Sistem ini diharapkan dapat memberikan estimasi tagihan secara akurat kepada pelanggan, sekaligus membantu pengelola dalam menganalisis pola konsumsi dan meningkatkan efisiensi pelayanan air bersih di tingkat desa.

BAHAN DAN METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen, yang bertujuan untuk mengembangkan sistem prediksi tagihan air berbasis *web* dengan menggunakan algoritma Regresi Linier Berganda (RLB). Sistem ini dirancang untuk memberikan estimasi tagihan air secara otomatis berdasarkan data historis pemakaian pelanggan layanan PAMSIMAS. Berikut merupakan kerangka pemikiran yang dapat dilihat pada Gambar 1 dibawah ini.



Sumber: (Hasil Penelitian, 2025)

Gambar 1. Kerangka Pemikiran

Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari pencatatan manual oleh pengelola PAMSIMAS di Desa Sidobandung, Kecamatan Balen, Kabupaten Bojonegoro. Teknik pengumpulan data dilakukan melalui metode dokumentasi, yaitu dengan mengambil data catatan pemakaian air dan tagihan bulanan yang telah didokumentasikan secara manual oleh pengelola PAMSIMAS dalam format fisik, lalu dikompilasi ke dalam file Excel untuk keperluan analisis lebih lanjut. *Dataset* terdiri atas 231 entri pelanggan, masing-masing memuat informasi jumlah pemakaian air dalam satuan meter kubik (m³) dan nilai tagihan dalam satuan rupiah (Rp), yang tercatat setiap bulan dari Januari hingga Desember 2024. Data awal dikompilasi dalam format Excel (.xlsx) dan kemudian diproses menggunakan platform *Google Colab* untuk keperluan pengolahan dan analisis lebih lanjut.

Tahap Preprocessing

Sebelum proses pemodelan dilakukan, data yang diperoleh terlebih dahulu melalui tahapan *preprocessing* agar siap untuk dianalisis (Pii et al., 2023). Baris-baris dengan nilai kosong yang terlalu banyak dihapus, sedangkan nilai kosong yang bersifat minor diisi menggunakan nilai median dari kolom terkait untuk menjaga kualitas data (Yulian Pamuji et al., 2024). Selanjutnya, proses deteksi dan penanganan *outlier* dilakukan menggunakan metode *Interquartile Range (IQR)* guna memastikan distribusi data tetap stabil dan tidak bias (Santoso & Priyadi, 2024). Setelah pembersihan data selesai, dilakukan proses standarisasi menggunakan metode *Z-score normalization*. Tujuannya adalah menyamakan skala antar variabel agar tidak ada variabel yang mendominasi dalam proses pelatihan model. Data yang sudah bersih kemudian dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%.

Pemodelan

Model prediksi yang dibangun dalam penelitian ini menggunakan algoritma Regresi Linier Berganda. Model ini memanfaatkan sebelas variabel *independen* berupa data pemakaian atau tagihan air dari bulan Januari hingga November, untuk memprediksi satu variabel *dependen* yaitu tagihan pada bulan Desember. Algoritma diterapkan menggunakan pustaka *scikit-learn* dalam bahasa pemrograman *Python*, dengan fungsi utama *LinearRegression()* untuk proses pelatihan. Secara umum, model regresi ini dirumuskan dalam bentuk persamaan linear :

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

Di mana *Y* merupakan nilai prediksi tagihan dan *X₁* hingga *X_n* adalah variabel input berdasarkan bulan-bulan sebelumnya.

Evaluasi Model

Untuk mengevaluasi kinerja model prediksi, digunakan dua metrik utama yaitu *Mean Squared Error (MSE)* dan *R-squared (R²)*. MSE digunakan untuk mengukur seberapa besar rata-rata kesalahan kuadrat antara nilai aktual dan hasil prediksi (Amansyah et al., 2024), sedangkan *R²* digunakan untuk mengukur proporsi variansi dari variabel target yang dapat dijelaskan oleh variabel input (Nuris, 2024). Kedua metrik ini dihitung menggunakan pustaka *scikit-learn* setelah model dilatih.

Implementasi Sistem

Model prediksi yang telah dibangun kemudian diintegrasikan ke dalam sistem berbasis *web*. Tujuan dari integrasi ini adalah mempermudah pengguna dalam mengakses fitur prediksi tanpa perlu memahami teknis perhitungan model. Sistem ini dikembangkan menggunakan *framework Flask (Python)* sebagai *backend*, dan *MySQL* sebagai basis data. Antarmuka pengguna dirancang agar pelanggan dapat login, melihat estimasi tagihan air bulan berikutnya berdasarkan riwayat konsumsi, serta menampilkan grafik tren pemakaian dan biaya tagihan. Visualisasi grafik dibangun menggunakan pustaka *Pandas* dan *Matplotlib*. Selain itu, disediakan panel khusus untuk admin guna mengelola data pelanggan dan histori pemakaian air secara efisien.

HASIL DAN PEMBAHASAN

Struktur Dataset

Dataset yang digunakan dalam penelitian ini memiliki struktur sebanyak 231 baris dan 30 kolom. Setiap baris merepresentasikan satu pelanggan PAMSIMAS di Desa Sidobandung, sedangkan setiap kolom merupakan atribut yang berkaitan dengan identitas pelanggan, parameter umum yang berkaitan dengan pemakaian air, volume pemakaian air bulanan, serta biaya tagihan air bulanan. Struktur kolom dalam *dataset* dijelaskan pada Tabel 1 berikut.

Tabel 1. Struktur Dataset

Nama Kolom	Tipe Data	Keterangan
Nama	String	Nama pelanggan
Alamat	String	Alamat tempat tinggal pelanggan
Bulan Lalu	Float	Volume air yang tercatat di bulan sebelumnya
Saat Ini (m ³)	Float	Volume air yang tercatat saat ini

Nama Kolom	Type Data	Keterangan
Harga Rp/m ³	Float	Harga pemakaian air per meter kubik
Beban	Float	Biaya beban tetap penggunaan air
Pemakaian Januari	Float	Volume pemakaian air bulan Januari
Pemakaian Februari	Float	Volume pemakaian air bulan Februari
Pemakaian Maret	Float	Volume pemakaian air bulan Maret
Pemakaian April	Float	Volume pemakaian air bulan April
Pemakaian Mei	Float	Volume pemakaian air bulan Mei
Pemakaian Juni	Float	Volume pemakaian air bulan Juni
Pemakaian Juli	Float	Volume pemakaian air bulan Juli
Pemakaian Agustus	Float	Volume pemakaian air bulan Agustus
Pemakaian September	Float	Volume pemakaian air bulan September
Pemakaian Oktober	Float	Volume pemakaian air bulan Oktober
Pemakaian November	Float	Volume pemakaian air bulan November
Pemakaian Desember	Float	Volume pemakaian air bulan Desember
Biaya Januari	Float	Biaya tagihan air bulan Januari
Biaya Februari	Float	Biaya tagihan air bulan Februari
Biaya Maret	Float	Biaya tagihan air bulan Maret
Biaya April	Float	Biaya tagihan air bulan April
Biaya Mei	Float	Biaya tagihan air bulan Mei
Biaya Juni	Float	Biaya tagihan air bulan Juni
Biaya Juli	Float	Biaya tagihan air bulan Juli
Biaya Agustus	Float	Biaya tagihan air bulan Agustus
Biaya September	Float	Biaya tagihan air bulan September
Biaya Oktober	Float	Biaya tagihan air bulan Oktober
Biaya November	Float	Biaya tagihan air bulan November
Biaya Desember	Float	Biaya tagihan air bulan Desember

Sumber: (Hasil Penelitian, 2025)

Preprocessing Data

Tahap *preprocessing* dilakukan untuk mempersiapkan data agar siap digunakan dalam proses pelatihan model Regresi Linier Berganda. Adapun tahapan *preprocessing* yang diterapkan secara berurutan meliputi penanganan *missing value*, pembersihan data, deteksi dan penanganan *outlier*, serta standarisasi data.

1. Penanganan *Missing Value*

Berdasarkan hasil analisis terhadap 231 baris data awal, beberapa kolom seperti pemakaian february, biaya february, hingga biaya desember memiliki jumlah nilai kosong yang bervariasi. Salah satu contohnya dapat dilihat pada Tabel 2 berikut, di mana pelanggan bernama Taman Waringin memiliki nilai kosong pada beberapa kolom tagihan dan pemakaian.

Tabel 2. Contoh Data Missing Value

Nama	Pemakaian Januari	Biaya Januari	Pemakaian Februari	Biaya Februari
Taman Waringin	-	-	-	-

Sumber: (Hasil Penelitian, 2025)

Setelah proses ini dilakukan, jumlah data berkurang dari 231 menjadi 221 baris, menunjukkan bahwa 10 entri dengan terlalu banyak *missing value* telah dihapus. Proses ini memastikan bahwa data yang digunakan dalam pemodelan berada dalam kondisi bersih dan layak untuk dilatih menggunakan model Regresi Linier Berganda. Perbandingan jumlah data sebelum dan sesudah penanganan *missing value* disajikan dalam Gambar 2.

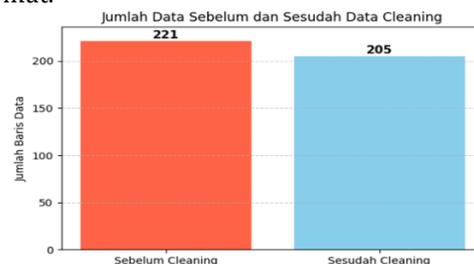


Sumber: (Hasil Penelitian, 2025)

Gambar 2. Data Setelah Missing Value

2. Pembersihan Data (Data Cleaning)

Poses data *cleaning* atau pembersihan data, bertujuan untuk menghilangkan entri-entri yang tidak valid atau berpotensi mengganggu hasil analisis. Setelah dilakukan pemeriksaan, sebanyak 16 baris data dibersihkan. Dengan demikian, jumlah data yang akan digunakan pada proses pemodelan menjadi 205 baris. Ringkasan perubahan jumlah data pada tahap ini dapat dilihat pada Gambar 3 berikut.



Sumber: (Hasil Penelitian, 2025)

Gambar 3. Data Setelah Cleaning

3. Deteksi dan Penanganan Outlier

Proses ini dilakukan untuk mengurangi pengaruh nilai-nilai ekstrem terhadap pemodelan. Berdasarkan hasil deteksi, ditemukan beberapa baris data yang mengandung nilai pemakaian air atau biaya yang jauh di luar rentang wajar.

Penanganan terhadap *outlier* dilakukan dengan mengganti nilai-nilai tersebut menggunakan median dari masing-masing kolom terkait. Pendekatan ini dilakukan secara hati-hati untuk menjaga agar pola utama dalam data tetap terjaga. Perbandingan data sebelum dan sesudah penyesuaian *outlier* ditampilkan pada Tabel 3 dan Tabel 4 berikut.

Tabel 3. Sebelum Penanganan Outlier

Pemakaian Januari	Pemakaian Februari	Pemakaian Maret	Pemakaian Desember
13	19	10	27
25	25	21	25
5	10	11	14

Sumber: (Hasil Penelitian, 2025)

Tabel 4. Sesudah Penanganan Outlier

Pemakaian Januari	Pemakaian Februari	Pemakaian Maret	Pemakaian Desember
13	14	14	14
25	25	21	25
10	10	11	10

Sumber: (Hasil Penelitian, 2025)

4. Standarisasi Data

Proses standarisasi dilakukan menggunakan metode *Z-score*, di mana setiap nilai dikonversi berdasarkan rata-rata dan standar deviasi kolomnya. Variabel target (Y) tetap dipertahankan dalam bentuk asli agar prediksi yang dihasilkan dapat dibaca dalam satuan rupiah. Contoh hasil transformasi *Z-score* untuk kolom pemakaian air bulan Januari dapat dilihat pada Tabel 5 berikut.

Tabel 5. Data Sebelum dan Sesudah Standarisasi

Pemakaian Januari	Z-score januari
10	-0,219969834
11	-0,083684176
14	0,325172798
8	-0,49254115
12	0,052601482
14	0,325172798
8	-0,49254115
8	-0,49254115

Sumber: (Hasil Penelitian, 2025)

Pemodelan Regresi Linier Berganda

1. Hasil Koefisien Model

Hasil pelatihan menghasilkan nilai *intercept* dan koefisien untuk setiap variabel X_1 hingga X_{11} . Nilai-nilai ini menunjukkan seberapa besar

pengaruh masing-masing variabel input terhadap output pemakaian bulan Desember, dengan asumsi variabel lain tetap konstan. Tabel 6 menunjukkan hasil koefisien dari model yang sudah dilatih.

Tabel 6. Hasil Koefisien RLB

No	Variabel	Koefisien (b)
1	Intercept (a)	2,36277618
2	X_1	0,10873866
3	X_2	-0,04097166
4	X_3	0,08608559
5	X_4	0,02863141
6	X_5	0,11865414
7	X_6	0,00070712
8	X_7	0,05917757
9	X_8	0,22631548
10	X_9	-0,14155648
11	X_{10}	-0,11997684
12	X_{11}	0,44150701

Sumber: (Hasil Penelitian, 2025)

2. Interpretasi Model

Untuk menguji bagaimana model ini bekerja, dilakukan perhitungan manual terhadap satu pelanggan yang memiliki riwayat pemakaian air dari bulan Januari hingga November sebagai dalam Tabel 7 berikut.

Tabel 7. Data Pemakaian Air

Bulan	Pemakaian (m ³)
Januari (X_1)	10
Februari (X_2)	11
Maret (X_3)	8
April (X_4)	17
Mei (X_5)	14
Juni (X_6)	13
Juli (X_7)	12
Agustus (X_8)	8
September (X_9)	7
Oktober (X_{10})	15
November (X_{11})	15

Sumber: (Hasil Penelitian, 2025)

Substitusi ke dalam persamaan :

$$\begin{aligned}
 Y &= 2,36277618 + (0,10873866 \times 10) + (-0,04097166 \times 11) \\
 &\quad + (0,08608559 \times 8) + (0,02863141 \times 17) \\
 &\quad + (0,11865414 \times 14) + (0,00070712 \times 13) \\
 &\quad + (0,05917757 \times 12) + (0,22631548 \times 8) \\
 &\quad + (-0,14155648 \times 7) \\
 &\quad + (-0,11997684 \times 15) \\
 &\quad + (0,44150701 \times 15) \\
 Y &= 2,36277618 + (1,087387) + (-0,45069) + (0,688685) \\
 &\quad + (0,486734) + (1,661158) + (0,009193) \\
 &\quad + (0,710131) + (1,81052388) \\
 &\quad + (-0,990895344) + (-1,79965267) \\
 &\quad + (6,62260511) = 12,197 \text{ m}^3
 \end{aligned}$$

Setelah diperoleh hasil prediksi pemakaian air untuk bulan Desember sebesar 12,197 m³, tahap selanjutnya adalah menghitung estimasi biaya yang harus dibayarkan oleh pelanggan. Perhitungan biaya ini mengikuti struktur tarif PAMSIMAS Desa Sidobandung, di mana tarif penggunaan air

ditetapkan sebesar Rp. 2,500 per meter kubik, ditambah dengan biaya beban tetap sebesar Rp. 3,000 per bulan. Dengan demikian, prediksi biaya tagihan air bulan Desember dapat dihitung menggunakan rumus :

$$Biaya = (Pemakaian\ Prediksi \times 2,500) + 3,000 \dots (2)$$

Substitusi nilai prediksi biaya pemakaian air :

$$Biaya = (12,197 \times 2,500) + 3,000 = 30,492 + 3000 = Rp. 33,492$$

Dengan demikian, hasil prediksi menunjukkan bahwa pelanggan diperkirakan akan mengonsumsi air sebesar 12,197 m³ dan menerima tagihan sebesar Rp. 33,493 pada bulan berikutnya. Nilai prediksi ini dapat dibandingkan secara langsung dengan data aktual untuk mengevaluasi akurasi model dalam memproyeksikan pemakaian dan biaya tagihan air.

Prediksi dilakukan terhadap seluruh baris data pada *dataset* uji, dan hasil prediksi dibandingkan secara langsung dengan nilai aktual pada kolom pemakaian bulan Desember. Perbandingan ini menjadi dasar dalam mengevaluasi kinerja model secara kuantitatif. Tabel 8 berikut menyajikan hasil prediksi dibandingkan dengan nilai aktual.

Tabel 8. Hasil Prediksi Pemakaian Air

Y Pemakaian Aktual	Y Pemakaian Prediksi
73	74,42
11	12,20
11	9,09
15	13,87
3	7,63
26	18,73
22	13,45
0	2,52
8	9,29
25	17,15
32	36,02

Sumber: (Hasil Penelitian, 2025)

Dari Tabel 8, terlihat bahwa sebagian besar prediksi berada dalam rentang yang cukup dekat dengan nilai aktual. Namun, terdapat beberapa prediksi yang menunjukkan deviasi cukup besar, seperti pada pemakaian aktual 26 yang diprediksi 18,73, atau aktual 22 yang diprediksi 13,45. Secara umum, model sudah menunjukkan kecenderungan yang benar, di mana prediksi meningkat seiring meningkatnya pola penggunaan air sebelumnya, dan sebaliknya.

Untuk memberikan pemahaman visual yang lebih baik, hasil prediksi dan nilai aktual juga divisualisasikan dalam bentuk grafik scatter plot seperti pada Gambar 4 berikut.



Sumber: (Hasil Penelitian, 2025)

Gambar 4. Plot Nilai Prediksi dan Nilai Aktual

Evaluasi Model

Model Regresi Linier Berganda yang dibangun dalam penelitian ini dievaluasi menggunakan dua metrik utama, yaitu *Mean Squared Error* (MSE) dan *R-squared* (R^2). Hasil evaluasi model ditunjukkan dalam Tabel 9.

Tabel 9. Hasil Evaluasi Model

Model	MSE	R^2
Regresi Linier Berganda	18,882	0,800

Sumber: (Hasil Penelitian, 2025)

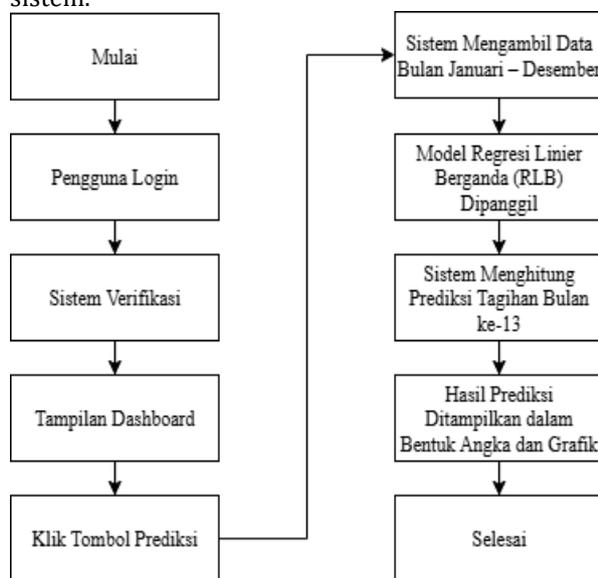
Berdasarkan hasil evaluasi, diperoleh nilai MSE sebesar 18,882 dan nilai R^2 sebesar 0,800, yang mengindikasikan bahwa model mampu menjelaskan sekitar 80% variabilitas dari data target. Nilai ini menunjukkan tingkat akurasi yang cukup baik dan dapat diterima dalam konteks prediksi berbasis data historis.

Jika dibandingkan dengan hasil penelitian sejenis, capaian model dalam penelitian ini tergolong kompetitif. Penelitian yang dilakukan oleh Prasetyo et al. dalam prediksi produksi kelapa sawit menggunakan RLB menghasilkan nilai *Mean Absolute Percentage Error* (MAPE) sebesar 14,28%, yang mencerminkan akurasi model yang cukup tinggi dalam konteks agrikultur (Prasetyo et al., 2021). Penelitian lain oleh Diah Retnowati et al. yang memprediksi permintaan air PDAM berdasarkan faktor sosial-ekonomi juga menunjukkan hubungan yang signifikan antar variabel prediktor, seperti halnya model dalam penelitian ini yang menggunakan data teknis bulanan (Diah Retnowati, Eli Susanti, Ratna Puji Astuti, Sodik Dwi Purnomo, Herwiek Diah Lestari, 2023). Selain itu, regresi linier yang diterapkan oleh Jafar et al. (Jafar et al., 2023) dan Fernandes et al. (P Fernandes et al., 2023) untuk memprediksi kualitas air menunjukkan nilai R^2 di atas 0,85, namun konteks dan kompleksitas data yang berbeda membuat angka R^2 sebesar 0,800 dalam penelitian ini tetap tergolong memadai. Mengacu pada teori

regresi multivariat, nilai R^2 di atas 0,7 umumnya dianggap cukup representatif untuk menggambarkan kekuatan model dalam menjelaskan hubungan antar variabel dalam data nyata.

Integrasi Model ke Sistem Web

Sistem ini dirancang untuk memberikan kemudahan kepada pelanggan PAMSIMAS dalam memperkirakan tagihan air pada bulan berikutnya (bulan ke-13), berdasarkan data historis pemakaian dan biaya pada bulan-bulan sebelumnya. Gambar 5 berikut merupakan diagram alur yang menggambarkan interaksi pengguna dengan sistem.



Sumber: (Hasil Penelitian, 2025)

Gambar 5. Alur Interaksi Sistem

Sistem prediksi tagihan air yang dibangun memiliki beberapa fitur utama yang terintegrasi dalam platform berbasis *web*. Salah satu fitur penting adalah login dan autentikasi yang memungkinkan dua jenis pengguna, yaitu admin dan pelanggan, untuk mengakses sistem dengan peran masing-masing. Proses autentikasi dilakukan dengan mencocokkan data input berupa *username* dan *password* terhadap data yang tersimpan pada tabel admin dan pelanggan di database *MySQL*.

Jika data yang dimasukkan sesuai, pengguna akan diarahkan ke halaman dashboard yang sesuai dengan hak aksesnya. Sistem juga memanfaatkan *session* untuk menyimpan informasi login sehingga pengguna tetap dapat mengakses halaman dashboard selama sesi berlangsung.

Pada sisi admin, setelah berhasil login, pengguna akan diarahkan ke halaman dashboard admin. Halaman ini menampilkan tabel data pelanggan serta menyediakan form input untuk

menambah data pemakaian dan tagihan pada bulan baru. Admin juga dapat memanfaatkan tombol edit untuk memperbarui data pelanggan pada bulan tertentu. Setiap kali data diinput atau diperbarui, sistem akan langsung menyimpan perubahan tersebut ke database *MySQL*. Selain itu, apabila admin menambahkan data bulan baru, maka data historis akan otomatis digeser sehingga data terbaru berada pada bulan Desember.

Bagi pelanggan, setelah berhasil login, pengguna diarahkan ke dashboard yang secara otomatis menampilkan data nama dan RT. Halaman ini menyediakan beberapa tombol utama, antara lain tombol untuk mengakses fitur prediksi, mengganti *password*, serta logout untuk keluar dari sistem dengan aman. Penggunaan *session* memastikan bahwa sistem hanya menampilkan data yang relevan sesuai akun pelanggan yang sedang aktif.

Fitur prediksi tagihan dan pemakaian air dapat diakses pelanggan dengan menekan tombol Prediksi. Sistem kemudian mengambil data historis pemakaian dan tagihan dari bulan Januari hingga Desember yang tersimpan dalam database. Model Regresi Linier Berganda yang telah dibangun kemudian dipanggil untuk memprediksi total tagihan tahunan serta estimasi biaya dan pemakaian air pada bulan ke-13. Hasil prediksi ini ditampilkan dalam bentuk tabel yang memuat informasi biaya (Rp) dan volume air (m³), serta dilengkapi grafik tren 12 bulan terakhir beserta proyeksi bulan berikutnya.

Visualisasi grafik dibuat menggunakan pustaka *matplotlib*, disimpan dalam format gambar (.png), kemudian ditampilkan pada halaman hasil prediksi. Kehadiran grafik ini memudahkan pelanggan dalam memahami pola konsumsi air mereka sekaligus mempersiapkan pengeluaran pada periode selanjutnya.

Sistem menyediakan fitur riwayat pemakaian dan tagihan air yang dapat difilter berdasarkan durasi waktu, yakni 3 bulan, 6 bulan, 9 bulan, atau 12 bulan terakhir. Setelah pelanggan memilih salah satu rentang waktu, sistem akan mengambil data dari database dan menampilkannya dalam bentuk tabel ringkasan yang berisi informasi bulan ke-n, volume pemakaian air (m³), dan total tagihan (Rp).

Sistem dilengkapi dengan fitur ganti *password* untuk menjaga keamanan akun pelanggan. Pada fitur ini, pengguna akan diminta memasukkan *password* lama dan *password* baru. Sistem kemudian memverifikasi kesesuaian *password* lama dengan data yang tersimpan di database. Apabila sesuai, *password* baru akan disimpan menggantikan *password* sebelumnya sehingga keamanan akun pelanggan tetap terjaga.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa sistem prediksi tagihan air berbasis *web* dengan metode Regresi Linier Berganda yang dikembangkan dalam penelitian ini berhasil menjawab kebutuhan digitalisasi layanan publik, khususnya bagi pelanggan PAMSIMAS di Desa Sidobandung. Sistem ini dibangun menggunakan pendekatan *client-server* dengan teknologi *HTML, CSS, JavaScript, Python (Flask)*, dan *MySQL*, yang mengintegrasikan proses input data historis, pemanggilan model, hingga visualisasi prediksi dalam satu antarmuka *web* yang responsif. Model yang diterapkan mampu memprediksi tagihan bulan ke-13 berdasarkan data pemakaian dan biaya dari Januari hingga Desember dengan performa cukup baik, ditunjukkan oleh nilai MSE sebesar 18,882 dan R^2 sebesar 0,8, yang berarti sekitar 80% variasi data target dapat dijelaskan oleh model. Kehadiran visualisasi hasil prediksi turut memudahkan pelanggan memantau pola konsumsi air serta membantu merencanakan pengeluaran secara lebih terstruktur. Dengan demikian, sistem ini dapat dimanfaatkan sebagai alat bantu analisis yang praktis dan informatif dalam mendukung transformasi digital layanan air bersih di tingkat desa.

REFERENSI

Amansyah, I., Indra, J., Nurlaelasari, E., & Juwita, A. R. (2024). Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia. *INNOVATIVE: Journal Of Social Science Research Volume*, 4(4), 1199-1216. <https://j-innovative.org/index.php/Innovative%0APrediksi>

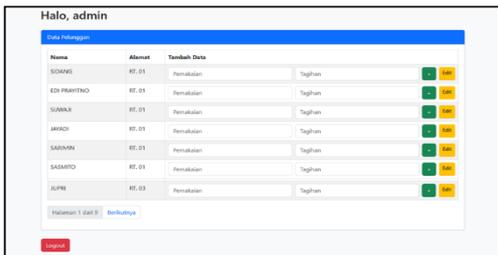
Deni, D. R., Barata, M. A., & Sahri. (2023). Forecasting Metode Single Exponential Smoothing Dalam Meramalkan Penjualan Barang. *Jurnal Informatika Polinema*, 9(4), 435-444. <https://doi.org/10.33795/jip.v9i4.1405>

Diah Retnowati, Eli Susanti, Ratna Puji Astuti, Sodik Dwi Purnomo, Herwiek Diyah Lestari, H. (2023). Analisis Permintaan Air Perusahaan Daerah Air Minum (Studi Empiris Pada PDAM Tirta Satria di Kota Purwokerto). *JIEP: Jurnal Ilmu Ekonomi Dan Pembangunan*, 6(2), 1195-1206.

Dimas Dwi Al Hakim, I. M. S. (2023). Analisis Permintaan Rumah Tangga Terhadap Air Pdam Di Kelurahan Kedungdoro Kota Surabaya. *Jurnal Ekonomi & Bisnis*, 8(2), 129-



Sumber: (Hasil Penelitian, 2025)
 Gambar 6. Tampilan Login



Sumber: (Hasil Penelitian, 2025)
 Gambar 7. Tampilan Admin



Sumber: (Hasil Penelitian, 2025)
 Gambar 8. Tampilan Pelanggan



Sumber: (Hasil Penelitian, 2025)
 Gambar 9. Hasil Prediksi



Sumber: (Hasil Penelitian, 2025)
 Gambar 10. Tampilan Filter Riwayat

138.
Hariyanto, M., Kholiq, M., Yani, A., & Narti. (2025). PREDIKSI HARGA PONSEL BERDASARKAN SPESIFIKASINYA MENGGUNAKAN ALGORITMA LINEAR REGRESSION. *Inti Nusa Mandiri*, 14(2), 133–138.
- Jafar, R., Awad, A., Hatem, I., Jafar, K., Awad, E., & Shahrour, I. (2023). Multiple Linear Regression and Machine Learning for Predicting the Drinking Water Quality Index in Al-Seine Lake. *Smart Cities*, 6(5), 2807–2827. <https://doi.org/10.3390/smartcities6050126>
- Jannah, H. N., Purwadi, O. T., & ... (2021). Potensi Penyediaan Air Bersih Berkelanjutan melalui Pemanenan Air Hujan (Studi Kasus Pulau Pasaran Kecamatan Teluk Betung Timur Kota Bandar Lampung). *Jurnal Rekayasa Sipil Dan ...*, 9(4), 809–818. <http://journal.eng.unila.ac.id/index.php/jrsd/article/view/2166>
- Nandita Afrilia S, Fathia Frazna Az-Zahra, P. (2024). Prediksi hasil panen wortel menggunakan algoritma regresi linear berganda. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10255–10262.
- Nuris, N. (2024). Analisis Prediksi Harga Rumah Pada Machine Learning Metode Regresi Linear. *Explore*, 14(2), 108–112. <https://doi.org/10.35200/ex.v14i2.123>
- P Fernandes, A. C., R Fonseca, A., Pacheco, F. A. L., & Sanches Fernandes, L. F. (2023). Water quality predictions through linear regression - A brute force algorithm approach. *MethodsX*, 10(March), 102153. <https://doi.org/10.1016/j.mex.2023.102153>
- Pii, I., Suarna, N., & Rahaningsih, N. (2023). Penerapan Data Mining Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 423–430. <https://doi.org/10.36040/jati.v7i1.6336>
- Prasetyo, A., Salahuddin, S., & Amirullah, A. (2021). Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda. *Jurnal Infomedia*, 6(2), 76. <https://doi.org/10.30811/jim.v6i2.2343>
- Salma Rahima Ahmad, H. S. D. (2023). Penerapan Regresi Linier Berganda dalam Penentuan Dosis Koagulan Optimal pada Instalasi Pengolahan Air Kaligarang III (Studi Kasus : Perumda Air Minum Tirta Moedal Kota Semarang). *Fakultas Teknik, Universitas Diponegoro*, 3–4.
- Santoso, L., & Priyadi. (2024). Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI. *Jurnal Elektronika Dan Komputer*, 17(2), 657–666.
- Yulian Pamuji, F., Ahmad Rofiqul Muslikh, Rizza Muhammad Arief, & Delviana Muti. (2024). Komparasi Metode Mean dan KNN Imputation dalam Mengatasi Missing Value pada Dataset Kecil. *Jurnal Informatika Polinema*, 10(2), 257–264. <https://doi.org/10.33795/jip.v10i2.5031>