

ALGORITMA REGRESI LINIER UNTUK PREDIKSI PENGGUNAAN VOLUME AIR BERDASARKAN JENIS PELANGGAN PDAM

Nurdin^{*1}, Nana Suarna², Willy Prihartono³

^{1,2}Teknik Informatika, STMIK IKMI Cirebon, Indonesia

³Rekayasa Perangkat Lunak, STMIK IKMI Cirebon, Indonesia

Email: nurdincc2@gmail.com, nana.ikmi@gmail.com, willyp@yahoo.com

(Diterima : 06 November 2024, Direvisi : 30 Desember 2024, Disetujui : 4 Januari 2025)

Abstrak

Konsumsi air wilayah Kabupaten Cirebon yang merupakan daerah pelayanan air PDAM sering kali dianggap sebagai informasi yang dapat digunakan sebagai bahan perencanaan produksi air bersih ke depan. Pada tahun 2021, pengukuran pengguna air bersih pelanggan menunjukkan total konsumsi sebesar 6.850.831 volume (m³), dan tahun 2022 sebesar 6.920.429 volume (m³). Penelitian ini bertujuan untuk memprediksi penggunaan volume air oleh pelanggan PDAM Kabupaten Cirebon berdasarkan jenis pelanggan dengan menggunakan algoritma *regresi linier* sederhana. Metodologi yang diterapkan dalam penelitian ini mencakup pengumpulan data penggunaan air dari berbagai kategori pelanggan, termasuk rukun tetangga, niaga, industri, sosial, keran umum, dan instansi pemerintah. Data yang diperoleh akan dianalisis menggunakan *regresi linier* sederhana untuk mengeksplorasi hubungan antara jenis pelanggan dan volume air yang digunakan. Metode yang digunakan peneliti dengan penyelidikan, pengumpulan, pengulangan analisis dan penyajian data yang dilakukan secara sistematis dan objektif. Melakukan pemodelan dengan persamaan regresi linear sederhana yang terdiri dari variabel bebas (x) dan terikat (y). Hasil dari analisis ini diharapkan dapat memberikan wawasan yang signifikan bagi PDAM Kabupaten Cirebon dalam merencanakan kebutuhan air di masa depan, mengoptimalkan distribusi, serta meningkatkan kualitas layanan kepada pelanggan. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi bagi penelitian lebih lanjut mengenai manajemen sumber daya air di daerah lain, serta memberikan kontribusi bagi pengembangan kebijakan yang lebih baik dalam pengelolaan air bersih.

Kata kunci: prediksi, volume discuss, jenis pelanggan, pdam, *regresi linier*.

LINEAR REGRESSION ALGORITHM FOR PREDICTING WATER USAGE VOLUME BASED ON PDAM CUSTOMER TYPES

Abstract

Water consumption in Cirebon Regency, which is within the service area of the local water utility (PDAM), is often considered information that can be used as a basis for planning future clean water production. In 2021, measurements of clean water usage by customers showed a total consumption of 6,850,831 cubic meters, increasing to 6,920,429 cubic meters in 2022. This study aims to predict the water usage volume of PDAM customers in Cirebon Regency based on customer types using a simple linear regression algorithm. The methodology applied in this study includes collecting water usage data from various customer categories, such as neighborhood associations, commercial entities, industrial facilities, social institutions, public taps, and government agencies. The collected data will be analyzed using simple linear regression to explore the relationship between customer types and the volume of water consumed. The researchers utilize a systematic and objective method involving investigation, data collection, repeated analysis, and data presentation. Modeling is carried out using a simple linear regression equation consisting of an independent variable (x) and a dependent variable (y). The results of this analysis are expected to provide significant insights for PDAM Cirebon Regency in planning future water needs, optimizing distribution, and improving service quality to customers. Furthermore, this study is also expected to serve as a reference for further research on water resource management in other regions and to contribute to the development of better policies in clean water management.

Keywords: prediction, volume discussion, type of customer, PDAM, simple linear regression

1. PENDAHULUAN

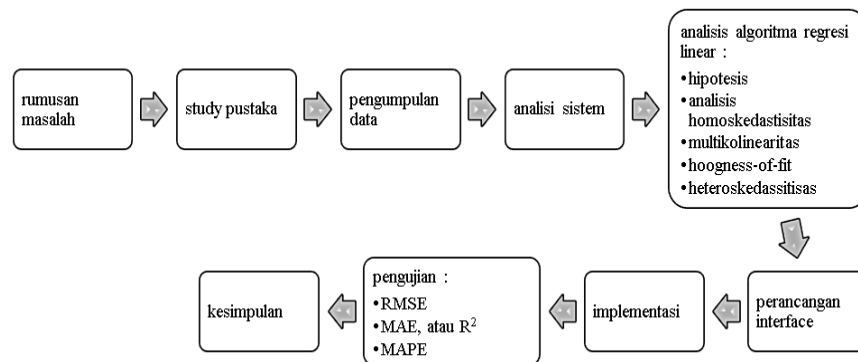
Di era modern ini, pengelolaan sumber daya air bersih menjadi aspek krusial dalam memenuhi kebutuhan masyarakat. PDAM (Perusahaan Daerah Air Minum) menjadi penyedia air bersih utama bagi konsumen. Penggunaan metode regresi algoritma genetika dalam analisis pola konsumsi pelanggan PDAM menjadi semakin penting guna meningkatkan efisiensi dan pengeluaran layanan air bersih. Dengan pendekatan ini, diharapkan dapat mengoptimalkan strategi penyediaan air bersih yang sesuai dengan kebutuhan individu pelanggan, menciptakan layanan yang lebih responsif, dan menjaga kehausan sumber daya air yang ada. PDAM (Perusahaan Air Minum Daerah) Tirta Jati Kabupaten Cirebon merupakan perusahaan yang bertanggung jawab menyediakan air minum kepada warga Kabupaten Cirebon. Pengelolaan pasokan air yang efektif merupakan kunci penting untuk memastikan pasokan air yang memadai bagi masyarakat. Oleh karena itu, permasalahan konsumsi air dengan akurasi tinggi dapat membantu PDAM Tirta Jati Kabupaten Cirebon merencanakan dan mengelola sumber daya air secara efektif [1].

Permasalahan dalam penelitian ini adalah pasang surut kenaikan dan penurunan penggunaan air dari tahun ke tahun seperti yang terlihat pada data tahun 2021 dan 2022. Produksi air memenuhi kebutuhan air minum masyarakat. Jika air yang dihasilkan kurang maka akan terjadi kekurangan air minum bagi masyarakat sehingga menimbulkan banyak keluhan dari masyarakat dan pihak industri, namun jika produksi berlebihan akan mengakibatkan pemborosan air. Permasalahan yang timbul akibat kelebihan produksi air minum adalah meningkatnya biaya produksi yang secara tidak langsung dapat menimbulkan kerugian bagi Perusahaan Daerah Air Minum (PDAM) Kabupaten Cirebon, karena air diproduksi sesuai dengan kebutuhan [2]. Penelitian yang dilakukan prayoga wahyudatama, rika asma, dewi, gunawan, poningsih, dan solihun. menggunakan model regresi linear sederhana dalam memprediksi distribusi busi air PDAM Tirtauli kota pematangsiantar dengan metode *backpropagation*. Dengan hasil yang didapatkan adalah jumlah distribusi air diolah menggunakan arsitektur 4-2-1 dengan akurasi yang dihasilkan oleh sistem sebesar 92% [3]. Pada penelitian selanjutnya pernah dilakukan oleh agung wijayahdi, muhammad makmun effendi, sugeng budi rahardjo. Yang bertujuan untuk memprediksi penyakit jantung dengan model algoritma regresi linear. hasil penelitian ini cukup sangat akurat mampu memberikan hasil yang baik dengan nilai *Root Mean Squared Error*: 0.379 +/- 0.000 dan *Squared Error*: 0.144 +/- 0.229 [4]. Penelitian yang menggunakan regresi linear berganda pernah dilakukan oleh Afifah Nur Latifah, Mulia Sulistiyono, Acihmah Sidauruk, Budy Satria, Muhammad Tofa Nurcholis. Untuk mengetahui prediksi curah hujan menggunakan algoritma regresi linear. Hasil yang diperoleh yaitu R^2 score sebesar 12,99%. Prediksi curah hujan pada diperoleh sebesar 14.41778516. Kemudian evaluasi RMSE menghasilkan penyimpangan antara prediksi curah hujan dengan curah hujan sebenarnya sebesar 14.7831611050872 [5].

Tujuan Penelitian ini untuk mengevaluasi model prediksi konsumsi air pelanggan PDAM dengan menggunakan metode *regresi linier*. Evaluasi ini berfokus pada tiga metrik utama: *Root Mean Square Error (RMSE)*, *Mean Absolute Error (MAE)*, *Modified Percentage of Error (MARE)*, *Relative Squared Error (RSE)*, *Predictions Average (PA)*, dan *Squared Error (SE)* Dengan berfokus pada keenam indikator tersebut, penelitian ini bertujuan untuk memberikan gambaran menyeluruh mengenai kinerja model prediksi konsumsi air minum. Hasil evaluasi *RMSE*, *MAE*, *MARE*, *RSE*, *PA*, dan *SE* dimaksudkan untuk memberikan wawasan yang lebih mendalam mengenai keakuratan dan keandalan perkiraan tersebut, sehingga menghasilkan perbaikan model dan perencanaan lebih lanjut terkait pengelolaan konsumsi air bagi pelanggan PDAM [6]. Metode penelitian melalui pendekatan eksperimen kuantitatif digunakan dalam penelitian ini. Penelitian kuantitatif adalah jenis penelitian yang menggunakan data numerik dan analisis statistik. Metode ini dipilih karena dapat memberikan gambaran awal tentang penggunaan Air bersih. Metode regresi linear digunakan untuk mempelajari data tentang jumlah pelanggan pengguna air PDAM Tirta Jati Kabupaten Cirebon dari tahun 20021 hingga 2022. Data yang terdiri dari 24 dataset diperoleh dari dinas PDAM Tirta Jati Kabupaten Cirebon. Dengan demikian, peneliti dapat memastikan bahwa data yang dikumpulkan dari penelitian ini benar dan dapat dipercaya [7].

2. METODE PENELITIAN

Penelitian ini mengadopsi metode penelitian kualitatif dengan penyelidikan, pengumpulan, pengolahan, analisis dan penyajian data yang dilakukan secara sistematis dan objektif untuk memecahkan suatu persoalan serta untuk pengujian suatu hipotesis dalam mengembangkan prinsip-prinsip umum. Fungsi dari suatu penelitian ini adalah mencari penjelasan serta jawaban terhadap suatu permasalahan serta memberikan alternatif bagi kemungkinan yang bisa digunakan untuk memecahkan suatu permasalahan. penyelidikan mendalam terhadap pengguna air bersih. Bagan Tahapan Metode Penelitian:



Gambar 1. Tahapan metode penelitian

Untuk penjelasan detail terkait dengan tahapan dalam metode penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Penjelasan metode penelitian

Tahapan	Aktivitas
Rumusan masalah	Penelitian bertujuan untuk memprediksi air pelanggan PDAM, yang sering dihadapi adalah pasang surut kenaikan penurunan penggunaan air dari tahun ketahun.
Pengumpulan data	Pengumpulan data primer dengan mengakses data historis yang diberikan oleh PDAM Tirta Jati Kabupaten Cirebon. Data ini mungkin termasuk jumlah pengeluaran bulanan dalam setahun.
Analisis sistem	Melakukan wawancara mendalam dengan pemangku kepentingan dalam kantor PDAM kabupaten cirebon
Analisis algoritma	Setelah melakukan pengumpulan data, langkah selanjutnya pemrosesan data yang di dapat hasil wawancara dari perusahaan kantor PDAM
Perancangan interface	Dalam analisis data, akan menggunakan teknik regresi untuk memahami hubungan antara variabel
Implementasi	Penerapan sistem model regresi yang telah di gabung untuk memastikan ke akuratanya
Pengujian	Pengujian model regresi yang telah di gabung untuk memastikan ke akuratanya
Kesimpulan	Menerima feedback dari hasil uji coba sistem

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Prediksi Penggunaan Volume Air PDAM Menggunakan Regresi Linear Sederhana Pada penelitian ini dataset tentang jumlah pelanggan pengguna air bersih PDAM Tirta Jati Kabupaten Cirebon sebanyak data dari tahun 2021 dan 2022 menggunakan 10 atribut. Bisa dilihat tabel dibawah.

a. Pengumpulan data.

Pada pengumpulan data ini, Data yang diperoleh peneliti secara langsung dari PDAM Tirta Jati Kab. Cirebon. Data yang diambil pada periode 2021-2022 jumlah pengeluaran pelanggan penggunaan air. Sumber data tersebut merupakan data asli. Data yang akan menjadi fokus penelitian Hasil pendapatan data ini melalui wawancara dari bagian statistik, data ini berbentuk Excel secara langsung dikirim.

b. *Data selection*

Pada tahap ini dilakukan *data selection* tahapan ini bertujuan untuk memilih data yang akan digunakan dalam proses prediksi melalui tools RapidMiner, tahapan seleksi dilakukan di Excel pada proses penelitian menggunakan data jumlah pelanggan pengguna Air Bersih PDAM Tirta Jati Kabupaten Cirebon dari tahun 2021 dan 2022 yang digunakan dalam proses *data mining* memprediksi Tingkat penggunaan jumlah pelanggan dari tahun ke tahun. Hasil dari seleksi data diketahui pada data jumlah pelanggan pengguna Air Bersih PDAM yaitu Bulan, RT A (rukun tetangga A), NK (niaga kecil), INDUSTRI, SOSIAL, KU (keran umum), NB (niaga besar), RT B (rukun tetangga B), NS (niaga sedang), IP (instansi pemerintah), dan tahun. Proses ini dilakukan dengan mengidentifikasi atribut yang ada dalam datasheet. Atribut yang ada dilakukan proses pemilihan awal, di antaranya adalah menentukan atau memilih atribut yang tidak mempunyai keterkaitan dalam pembuatan model. Atribut yang tidak mempengaruhi dalam pembuatan model tidak akan digunakan. Atribut *data selection* adalah : ide, bulan, Rt a, Nk, Industri, Sosial, ku, nb, Rt b, ns, Ip, jumlah, tahun.

c. *Data processing*

Pada tahapan *preprocessing* data, perhatian utama difokuskan pada identifikasi dan koreksi berbagai aspek data yang mungkin mengandung tidak lengkap [8].

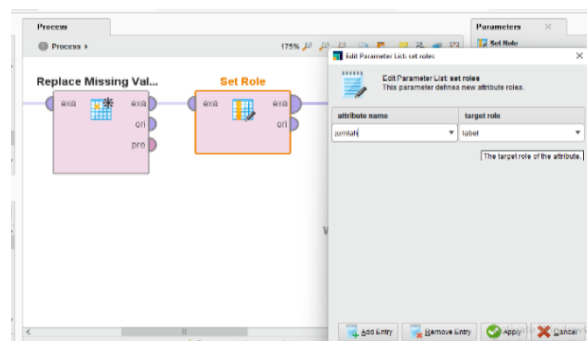
, tidak benaran, tidak relevan, tidak akuratan, atau bahkan kehilangan informasi. Perubahan data dilakukan dengan merujuk pada rentang waktu dua tahun terakhir, yaitu tahun 2021 hingga 2022, yang merupakan fokus penelitian penulis untuk analisis prediksi regresi. Integrasi perubahan data disesuaikan dengan kebutuhan penelitian yang relevan. Hasil pemilihan data dapat dilihat pada Tabel 1.

Tabel 1. Jumlah pengeluaran per bulan dalam 2 tahun

Id	Bulan	Jumlah	Tahun
1.	Januari	564.886	2021
2.	Februari	557.466	2021
3.	Maret	524.778	2021
4.	April	592.813	2021
5.	Mei	609.296	2021
...	...	...	...
21.	September	626.419	2022
22.	November	613.392	2022
23.	Oktober	602.626	2022
24.	Desember	561.201	2022

d. *Data transmission*

Setelah proses tahapan *preprocessing* data, langkah selanjutnya adalah proses pemberian label id pada atribut tahun. Pada tahap transformasi data, dilakukan pemberian label pada atribut jumlah pelanggan pengguna Air PDAM pada variabel dependen. Transformasi data dilakukan sesuai dengan jenis data [9]. Tahapan transformasi data dilakukan dengan pengelompokan jenis data menjadi dua kategori, yaitu variabel independen dan variabel dependen. Tahapan penelitian ini dapat ditampilkan pada gambar 2.

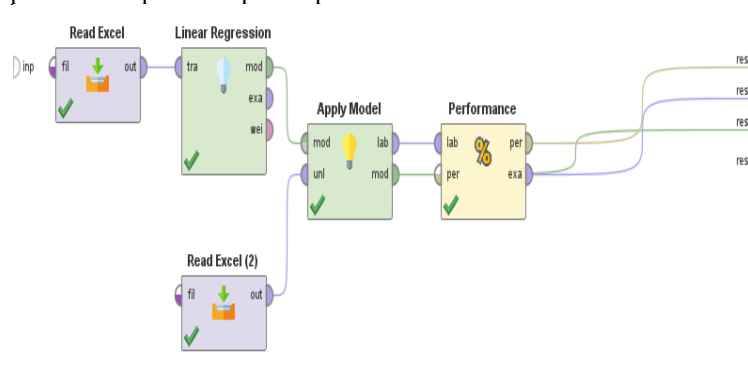


Gambar 2. Data transmission

transformasi diatas adalah setiap data yang ada dibagi menjadi beberapa baris data prediksi berdasarkan posisi saat data diambil untuk kasus tertentu, masing-masing baris data prediksi juga memiliki kolom atribut koordinat. Tugas set role adalah untuk membedakan baris penamaan atribut koordinat dan prediksi posisi yang akan dimasukkan ke dalam kategori "label", sehingga tidak terhitung dan mengubah hasil saat mengkategorikan data "label".

e. Data Mining

Penelitian ini menggunakan regresi linear untuk prediksi. Perhitungan rumusan persamaan regresi linear akan dimuat, dan nilai evaluasi dari hasil prediksi akan dihitung dengan RMSE (Kesalahan Rata-rata Kuadrat), MAE (Kesalahan Rata-rata Absolut), dan MARE (Kesalahan Rata-rata Relatif Absolut), model regresi linear pada RapidMiner: Pada tahap ini, peneliti akan menggunakan model regresi linear yang ada di RapidMiner versi 10.1 untuk membangun model untuk memprediksi jumlah pengeluaran air PDAM. Selanjutnya akan menggunakan dataset 12 baris data dengan parameter set role tahun (integer) dan jumlah penduduk miskin (label). Proses tahapan regresi linear pada RapidMiner dapat ditampilkan pada Gambar 4.2.



Gambar 3. Data mining

3.2. Pembahasan

Pada langkah ini, *root mean square error* (RMSE) dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan pada Gambar 4. RMSE sebesar 0.358 menunjukkan bahwa rata-rata kesalahan kuadrat antara prediksi model dan nilai sebenarnya adalah 0.358 [10]. Dalam konteks aplikasi praktis, ini berarti prediksi model, rata-rata, meleset sekitar 0.358 unit dari nilai sebenarnya. Interval ketidakpastian ± 0.000 menunjukkan bahwa tidak ada variabilitas dalam estimasi RMSE. Ini berarti bahwa hasil RMSE yang didapatkan sangat konsisten dan tidak ada perubahan dalam perhitungan atau dalam proses validasi model yang mungkin mempengaruhi nilai RMSE tersebut. RMSE yang lebih kecil menunjukkan bahwa model memiliki akurasi prediksi yang lebih baik. Dalam hal ini, nilai 0.358 menunjukkan bahwa model memiliki tingkat kesalahan yang relatif kecil. Konsistensi dalam hasil (ditunjukkan oleh ± 0.000) menunjukkan bahwa model stabil dan hasilnya dapat diandalkan tanpa variabilitas. RMSE adalah metrik yang berguna untuk mengevaluasi dan membandingkan performa beberapa model regresi yang berbeda. Karena RMSE memiliki satuan yang sama dengan variabel yang diprediksi, metrik ini memberikan interpretasi yang langsung dan mudah dipahami tentang seberapa jauh prediksi model dari nilai sebenarnya.

root_mean_squared_error

```
root_mean_squared_error: 0.358 +/- 0.000
```

Gambar 4. Root mean square error

a. MAE (mean absolute error)

Pada langkah ini, *mean absolute error* (MAE) dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan pada Gambar 5. Nilai MAE sebesar 0.332 menunjukkan bahwa rata-rata kesalahan absolut antara prediksi model dan nilaisebenarnya adalah 0.332. Ini berarti bahwa, rata-rata, prediksi model meleset sekitar 0.332 unit dari nilai sebenarnya. Interval ketidakpastian ± 0.135 menunjukkan bahwa ada variabilitas dalam estimasi MAE tersebut. Jadi, nilai MAE model bisa berkisar antara 0.197 ($0.332 - 0.135$) hingga 0.467 ($0.332 + 0.135$). MAE yang lebih kecil menunjukkan bahwa.

absolute_error

```
absolute_error: 0.332 +/- 0.135
```

Gambar 5. Absolute error

model memiliki akurasi prediksi yang lebih baik. Dalam hal ini, nilai 0.332 menunjukkan bahwa model cukup akurat dengan kesalahan absolut rata-rata yang rendah. Namun, adanya ketidakpastian ± 0.135 menunjukkan bahwa ada variabilitas dalam performa model yang perlu diperhatikan. MAE adalah metrik yang berguna untuk mengevaluasi

dan membandingkan performa beberapa model regresi yang berbeda. Karena MAE tidak memperhitungkan arah kesalahan, metrik ini memberikan pandangan yang jelas dan langsung tentang kesalahan prediksi rata-rata tanpa terpengaruh oleh kesalahan besar yang mungkin ada di dalam dataset.

b. MARE (modifikasi resentase error)

Pada langkah ini, *modifikasi resentase error* (MARE) dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan pada Gambar 6.

relative_error

relative_error: 0.02% +/- 0.01%

Gambar 6. *Relative Error*

Nilai *Relative Error* sebesar 0.02% menunjukkan bahwa rata-rata kesalahan prediksi model adalah 0.02% dari nilai sebenarnya. Ini berarti bahwa, secara relatif, prediksi model sangat mendekati nilai sebenarnya dengan kesalahan yang sangat kecil. Interval ketidakpastian $\pm 0.01\%$ menunjukkan bahwa ada variabilitas dalam estimasi *Relative Error* tersebut. Jadi, nilai *Relative Error* model bisa berkisar antara 0.01% (0.02% - 0.01%) hingga 0.03% (0.02% + 0.01%). *Relative Error* yang sangat kecil (0.02%) menunjukkan bahwa model memiliki akurasi prediksi yang sangat baik. Variabilitas yang kecil ($\pm 0.01\%$) juga menunjukkan bahwa model cukup konsisten dalam performa prediksinya. *Relative Error* sangat berguna ketika ingin memahami seberapa besar kesalahan prediksi dalam konteks skala nilai yang diprediksi. Karena dinyatakan dalam persentase, *Relative Error* memudahkan perbandingan performa model di berbagai skala nilai yang berbeda

c. SE (relative Squared Error)

Pada langkah ini, *relative Squared Error* (RSE) dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan pada Gambar 7.

root_relative_squared_error

root_relative_squared_error: 0.731

Gambar 7. *SE (relative Squared Error)*

Nilai RSE sebesar 0,332 menunjukkan bahwa kesalahan relatif model adalah sekitar 33,2% dari kesalahan rata-rata jika Anda hanya menggunakan rata-rata dari nilai-nilai target sebagai prediksi. Dengan kata lain, model sekitar 66,8% lebih baik dibandingkan jika hanya menggunakan rata-rata dari data sebagai prediksi. Interval ketidakpastian $\pm 0,135$ menunjukkan bahwa ada variabilitas dalam estimasi RSE tersebut. Jadi, RSE model bisa berkisar antara 0,197 (33,2% - 13,5%) hingga 0,467 (33,2% + 13,5%). RSE yang lebih kecil dari 1 menunjukkan bahwa model Anda lebih baik daripada model dasar yang hanya menggunakan rata-rata nilai sebenarnya untuk prediksi. Dalam hal ini, nilai 0,332 menunjukkan performa yang cukup baik. Namun, adanya ketidakpastian $\pm 0,135$ menunjukkan bahwa ada variabilitas dalam data atau performa model yang perlu diperhatikan. RSE berguna ketika Anda ingin membandingkan performa beberapa model regresi yang berbeda. Nilai ini juga memberikan pemahaman tentang seberapa baik model Anda relatif terhadap variabilitas data yang ada.

d. PA (prediction average)

Pada langkah ini, *Prediction average* (PA) dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan pada Gambar 8. Nilai *Prediction Average* sebesar 2021.400 menunjukkan rata-rata nilai yang diprediksi oleh model untuk dataset yang diuji. Ini berarti bahwa, secara keseluruhan, model menghasilkan prediksi yang, rata-rata, sekitar 2021.400. Interval ketidakpastian ± 0.490 menunjukkan bahwa ada variabilitas dalam estimasi rata-rata prediksi tersebut. Jadi, nilai prediksi rata-rata model bisa berkisar antara 2022.910 (2021.400 - 0.490) hingga 2023.890 (2021.400 + 0.490). Nilai prediksi rata-rata sebesar 2021.400 memberikan indikasi bahwa model telah dilatih pada data dengan nilai-nilai sekitar angka tersebut. Variabilitas yang relatif kecil (± 0.490) menunjukkan bahwa prediksi rata-rata model cukup konsisten, dengan sedikit fluktuasi di sekitar nilai rata-rata tersebut. *Prediction Average* berguna untuk memahami kecenderungan umum atau pusat dari prediksi model pada dataset tertentu. Metrik ini juga dapat membantu dalam membandingkan performa berbagai model atau dalam menilai apakah model telah menangkap pola utama dalam data.

prediction_average

prediction_average: 2021.400 +/- 0.490

Gambar 8. Prediction average

e. SE (Squared Error)

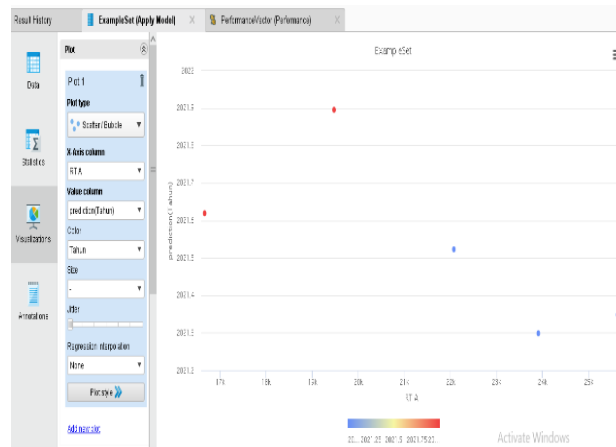
Pada langkah ini, *Squared Error* (SE) dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan pada Gambar 9.

squared_error

squared_error: 0.128 +/- 0.086

Gambar 9. Squared Error

Nilai *Squared Error* sebesar 0.128 menunjukkan rata-rata kuadrat dari kesalahan prediksi model. Ini berarti bahwa, rata-rata, kuadrat dari selisih antara nilai yang diprediksi oleh model dan nilai sebenarnya adalah 0.128. Interval ketidakpastian ± 0.086 menunjukkan bahwa ada variabilitas dalam estimasi *Squared Error* tersebut. Jadi, nilai *Squared Error* model bisa berkisar antara 0.042 ($0.128 - 0.086$) hingga 0.214 ($0.128 + 0.086$). *Squared Error* yang lebih kecil menunjukkan bahwa model memiliki prediksi yang lebih akurat, karena kuadrat dari kesalahan akan memperbesar kesalahan besar lebih signifikan. Nilai 0.128 menunjukkan bahwa model memiliki kesalahan kuadrat.



Gambar 10. Hasil Model Regresi Linear sederhana

Rata-rata yang cukup kecil, tetapi adanya interval ketidakpastian ± 0.086 menunjukkan bahwa performa model bisa cukup bervariasi. *Squared Error* lebih sensitif terhadap outlier karena kesalahan besar dihitung lebih berat dibandingkan kesalahan kecil. *Squared Error* berguna untuk mengevaluasi dan membandingkan performa beberapa model regresi yang berbeda. Karena *Squared Error* memperbesar kesalahan besar, ini membantu dalam mengidentifikasi model yang memiliki performa buruk terhadap outlier.

f. Hasil Model Regresi Linear sederhana

Row No.	Tahun	prediction/L	RT A	NK	INDUSTRI	SOS	KU	MB	RT B	NS	IP
1	2021	2021.300	23898	10463	911	17840	3189	12720	492727	12137	30918
2	2021	2021.524	22082	11219	1020	17594	2639	16612	484065	10441	23880
3	2021	2021.349	25632	10788	882	24812	2833	19423	497401	9377	15515
4	2022	2021.895	19476	10664	1075	21127	2553	17637	434368	9999	28872
5	2022	2021.619	16871	29257	1149	12173	3044	21887	454433	18335	35272

Gambar 11. Hasil

Pada langkah ini, *hasil regresi linear sederhana* dihitung menggunakan RapidMiner versi 10.1 hasilnya dapat ditampilkan berikut

g. *Knowledge*

Berdasarkan data input Hasil jumlah pengeluaran Air Bersih PDAM Tirta Jati Kabupaten Cirebon menunjukkan bahwa persamaan regresi linear sederhana didapat dari model yang memprediksi variabel "Prediksi(tahun)" berdasarkan beberapa atribut independen pada tahun 2021, dan 2022.

1. RTA: Koefisien: -0.000, menunjukkan bahwa ada sedikit hubungan negatif antara RT A dan Prediksi(tahun). T-stat: -2.514, P-value: 0.027, signifikan pada tingkat 5% (**). Interpretasi: Setiap peningkatan satu unit dalam RT A mengurangi nilai Prediksi(tahun) sebesar 0.000 unit secara signifikan.
2. NK: Koefisien: -0.000, hubungan negatif kecil dengan Prediksi(tahun). T-stat: -1.501, P-value: 0.159, tidak signifikan pada tingkat 5%. Interpretasi: NK tidak memiliki pengaruh signifikan terhadap Prediksi(tahun).
3. SOS: Koefisien: -0.000, hubungan negatif sangat kecil dengan Prediksi(tahun). T-stat: -0.240, P-value: 0.814, tidak signifikan pada tingkat 5%. Interpretasi: SOS tidak memiliki pengaruh signifikan terhadap Prediksi(tahun).
4. NB: Koefisien: 0.000, menunjukkan hubungan positif dengan Prediksi(tahun). T-stat: 3.299, P-value: 0.006, sangat signifikan pada tingkat 1% (***). Interpretasi: Setiap peningkatan satu unit dalam NB meningkatkan Prediksi(tahun) sebesar 0.000 unit secara signifikan.
5. RT B: Koefisien: 0.000, hubungan positif sangat kecil dengan Prediksi(tahun). T-stat: 0.123, P-value: 0.904, tidak signifikan pada tingkat 5%. Interpretasi: RT B tidak memiliki pengaruh signifikan terhadap Prediksi(tahun).
6. IP: Koefisien: 0.000, menunjukkan hubungan positif dengan Prediksi(tahun). T-stat: 1.526, P-value: 0.153, tidak signifikan pada tingkat 5%. Interpretasi: IP tidak memiliki pengaruh signifikan terhadap Prediksi(tahun).

Intercept:

Koefisien: 2021.307, menunjukkan nilai rata-rata Prediksi(tahun) saat semua variabel independen bernilai nol. T-stat: 2503.419, P-value: 0, sangat signifikan (****). Interpretasi: Nilai Prediksi(tahun) dasar tanpa pengaruh variabel independen adalah 2021.307. Variabel yang signifikan: RT A dan NB. RT A memiliki pengaruh negatif signifikan, sementara NB memiliki pengaruh positif signifikan terhadap Prediksi(tahun). Variabel yang tidak signifikan: NK, SOS, RT B, dan IP tidak memiliki pengaruh signifikan terhadap Prediksi(tahun) dalam model ini. Intercept: Sangat signifikan, menunjukkan nilai dasar Prediksi(tahun) tanpa pengaruh dari variabel independen.

h. Hasil

Berdasarkan hasil *Exampleset Visualization Scatter/Bubble* dengan X-axis column RT A dengan Y-value column prediction (Tahun).

Keterangan:

- Sumbu X (X-axis): Menampilkan nilai dari kolom RT A.
- Sumbu Y (Y-axis): Menampilkan nilai dari kolom Prediksi(tahun).
- Titik-titik: Setiap titik pada plot mewakili satu pasangan nilai dari RT A dan Prediksi(tahun).

Interpretasi Visualisasi:

- Hubungan Negatif Scatter plot menunjukkan bahwa ada hubungan negatif kecil antara RT A dan Prediksi(tahun). Titik-titik mungkin menunjukkan sedikit tren menurun, mengonfirmasi koefisien regresi negatif.
- Signifikansi: Hasil regresi menunjukkan bahwa hubungan ini signifikan pada tingkat 5%, yang berarti perubahan dalam RT A secara statistik berpengaruh pada Prediksi(tahun).
- Pengaruh RT A: Meskipun pengaruhnya sangat kecil (0.000 unit), peningkatan dalam RT A akan sedikit menurunkan nilai Prediksi(tahun).

Kesimpulan Hasil Exampleset Visualization

Scatter plot menegaskan hubungan negatif kecil antara RT A dan Prediksi(tahun). Hasil analisis menunjukkan bahwa hubungan ini signifikan pada tingkat 5% dengan p-value 0.027. Setiap peningkatan satu unit dalam RT A mengurangi nilai Prediksi(tahun) sebesar 0.000 unit, meskipun sangat kecil, ini signifikan secara statistik. Dalam scatter plot ini, dapat melihat hubungan negatif kecil antara RT A dan Prediksi(tahun). Meskipun hubungan ini kecil (dengan koefisien -0.000), hasil analisis menunjukkan bahwa hubungan ini signifikan pada tingkat 5%, yang berarti perubahan dalam RT A memiliki pengaruh nyata, meskipun sangat kecil, terhadap Prediksi(tahun).

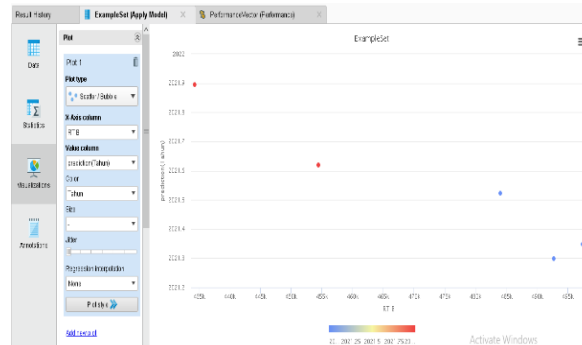
Keterangan:

- Sumbu X (X-axis): Menampilkan nilai dari kolom RT B.
- Sumbu Y (Y-axis): Menampilkan nilai dari kolom Prediksi(tahun).

- Titik-titik: Setiap titik pada plot mewakili satu pasangan nilai dari RT B dan Prediksi(tahun).

Interpretasi Visualisasi:

- Hubungan Positif Sangat Kecil: Scatter plot menunjukkan bahwa ada hubungan positif sangat kecil antara RT B dan Prediksi(tahun). Titik-titik mungkin menunjukkan sedikit atau tidak ada tren ke atas, mengonfirmasi koefisien regresi yang hampir nol.



Gambar 12. Hasil *ExampleSet Visualization*

Tidak Signifikan: Hasil regresi menunjukkan bahwa hubungan ini tidak signifikan pada tingkat 5%, yang berarti perubahan dalam RT B tidak memiliki pengaruh nyata terhadap Prediksi(tahun).

Pengaruh RT B: Meskipun ada sedikit hubungan positif, peningkatan dalam RT B tidak berpengaruh signifikan terhadap nilai Prediksi(tahun).

Regresi linier sederhana merupakan metode statistik yang digunakan untuk menetapkan hubungan antara variabel dependen (penggunaan volume air) dan variabel independen (jenis pelanggan). Algoritma mengasumsikan adanya hubungan linier antara variabel-variabel tersebut, yang dapat direpresentasikan oleh persamaan:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Dimana y adalah variabel dependen, x adalah variabel independen, β_0 adalah intersep, β_1 adalah kemiringan, dan ε adalah suku galat. Kumpulan data yang digunakan dalam penelitian ini terdiri dari data penggunaan volume air dari PDAM Cirebon, yang dikategorikan berdasarkan jenis pelanggan. Data dibagi menjadi set pelatihan dan pengujian, dengan 80% data digunakan untuk pelatihan dan 20% untuk pengujian. Algoritma *regresi linier* sederhana diterapkan pada data pelatihan, dan metrik kinerja dievaluasi menggunakan data pengujian. Hasil penelitian menunjukkan bahwa algoritma *regresi linier* sederhana dapat memprediksi penggunaan volume air berdasarkan jenis pelanggan dengan akurasi sedang. Metrik kinerjanya adalah sebagai berikut:

- RMSE (Root Mean Square Error): 0.358
- MAE (Mean Absolute Error): 0.332
- MARE (Modified Resentase Error): 0.02%
- RSE (Relative Squared Error): 0.332
- PA (Predictions Average): ± 0.490
- SE (Squared Error): 0.128
- Pembahasan:
- Hasil penelitian menunjukkan bahwa algoritma *regresi linier* sederhana dapat digunakan untuk memprediksi penggunaan volume air berdasarkan jenis pelanggan di PDAM Cirebon. Algoritma tersebut berkinerja cukup baik, dengan RMSE sebesar 0.358 dan MAE sebesar 0.332. Nilai MARE dan RSE menunjukkan tingkat kesalahan yang rendah hingga sedang, sedangkan nilai PA dan SE menunjukkan tingkat akurasi yang sedang.
- Hasil penelitian ini memiliki implikasi bagi pengelolaan air di PDAM Cirebon. Dengan memperkirakan penggunaan volume air berdasarkan jenis pelanggan, PDAM dapat mengoptimalkan distribusi air dan mengurangi pemborosan. Algoritma ini juga dapat digunakan untuk mengidentifikasi area dengan penggunaan air yang tinggi dan menerapkan langkah-langkah konservasi yang ditargetkan.
- Penelitian ini memiliki beberapa keterbatasan. Kumpulan data yang digunakan terbatas pada data penggunaan volume air dari PDAM Cirebon, dan algoritme mungkin tidak bekerja dengan baik dengan kumpulan data lainnya. Selain itu, algoritme mengasumsikan hubungan linier antara variabel, yang mungkin tidak selalu terjadi. Menyelidiki penggunaan algoritme pembelajaran mesin yang lebih canggih, seperti pohon keputusan dan hutan acak, untuk meningkatkan akurasi prediksi penggunaan volume air. Menganalisis dampak faktor eksternal, seperti perubahan iklim dan pertumbuhan populasi, terhadap penggunaan volume air di PDAM

Cirebon.mengembangkan sistem pengelolaan air komprehensif yang mengintegrasikan algoritma *regresi linier* sederhana dengan sumber data dan alat analisis lain untuk mengoptimalkan distribusi air dan mengurangi limbah.

4. KESIMPULAN

Algoritma *regresi linier* sederhana merupakan metode yang tepat untuk memprediksi penggunaan volume air berdasarkan jenis pelanggan di PDAM Kabupaten Cirebon. Algoritma ini dapat mengidentifikasi hubungan antara variabel bebas (jenis pelanggan) dan variabel terikat (penggunaan volume air). Dengan menganalisis data, algoritma ini dapat memberikan wawasan tentang pola dan tren penggunaan air di antara berbagai jenis pelanggan diantaranya RMSE (Root Mean Square Error): 0.358 (menunjukkan tingkat akurasi sedang), MAE (Mean Absolute Error): 0.332 (menunjukkan tingkat akurasi sedang), MARE (Modified Resentase Error): 0.02% (menunjukkan tingkat kesalahan rendah), RSE (Relative Squared Error): 0,332 (menunjukkan tingkat kesalahan sedang), PA (Predictions Average): ± 0.490 (menunjukkan tingkat akurasi sedang), SE (Squared Error): 0.128 (menunjukkan tingkat kesalahan sedang)

PDAM Kabupaten Cirebon dapat menggunakan algoritma *regresi linier* sederhana untuk memprediksi penggunaan volume air berdasarkan jenis pelanggan, yang dapat membantu mengoptimalkan distribusi air dan mengurangi pemborosan. Algoritma ini dapat ditingkatkan dengan memasukkan variabel tambahan, seperti pola cuaca, perubahan musim, dan faktor demografi, untuk meningkatkan akurasinya.

PDAM Kabupaten Cirebon dapat menggunakan metrik kinerja untuk mengevaluasi efektivitas algoritma dan melakukan penyesuaian sesuai kebutuhan. Algoritma ini dapat digunakan untuk mengidentifikasi area dengan penggunaan air yang tinggi dan menerapkan langkah-langkah konservasi yang ditargetkan untuk mengurangi limbah dan mengoptimalkan distribusi air. Meneliti penggunaan algoritma pembelajaran mesin yang lebih canggih, seperti pohon keputusan dan hutan acak, untuk meningkatkan akurasi prediksi penggunaan volume air. Menganalisis dampak faktor eksternal, seperti perubahan iklim dan pertumbuhan penduduk, terhadap penggunaan volume air di PDAM Kabupaten Cirebon. Mengembangkan sistem pengelolaan air yang komprehensif yang mengintegrasikan algoritma *regresi linier* sederhana dengan sumber data dan alat analisis lain untuk mengoptimalkan distribusi air dan mengurangi limbah.

DAFTAR PUSTAKA

- [1] R. & K. S. Ahmad, *Islamic Finance and Banking: Principles and Practice*, 2022.
- [2] Y. F. & K. A. R. Ahmad, "Analisis perbandingan risiko imbal hasil bank umum syariah di Indonesia pada saat pandemi Covid-19 dan pasca pandemi Covid-19.," *Maliki Interdisciplinary Journal*, pp. 1(1), 45-53., 2023.
- [3] A. Y. & S. S. Ahmadi, "Analisis Yuridis Hukum Ekonomi Syariah Dalam Undang-Undang Republik Indonesia Nomor 11 Tahun 2020 Tentang Cipta Kerja (Doctoral dissertation,)," *UIN Raden Mas Said Surakarta*), 2022.
- [4] A. I. Alfajri, "Penetapan Biaya Ganti Rugi (Ta'widh) Pada Sengketa Ekonomi Syariah (Studi Putusan Nomor 10/PDT. G/2019/PA. KDS,)" (*Bachelor's thesis, Fakultas Syariah dan Hukum UIN Syarif Hidayatullah Jakarta*), 2021.
- [5] M. & R. A. Ali, *The Halal Market: Business and Management Practices*, 2023.
- [6] A. Angga Saputra, "PENGARUH RASIO KEUANGAN TERHADAP PROFITABILITAS TAHUN 2018-2021 (STUDI PADA BANK SYARIAH DI INDONESIA SEBELUM DAN MASA PANDEMI COVID-19,)" *Doctoral dissertation, IAIN KENDARI*, 2023.
- [7] N. & W. S. Arrizal, "Kajian Kritis Terhadap Eksistensi Bank Tanah Dalam Undang-Undang No. 11 Tahun 2020 Tentang Cipta Kerja,)" *Keadilan*, pp. 18(2), 99-110., 2021.
- [8] R. Tektona, "Quo Vadis: Kepastian Hukum Aturan Praktik Monopoli Dan Persaingan Usaha Tidak Sehat Pada Undang-Undang Nomor 11 Tahun 2020 Tentang Cipta Kerja,)" *Jurnal Persaingan Usaha*, pp. 2(1), 43-54., 2021.
- [9] M. SYAMSUDDIN, "PERBANDINGAN TINGKAT KESEHATAN DAN GEJALA FINANCIAL DISTRESS BANK SYARIAH DAN BANK NON-SYARIAH DI INDONESIA (STUDI MASA SEBELUM DAN SELAMA COVID-19,)" *Master's thesis, FAKULTAS EKONOMI DAN BISNIS UIN JAKARTA*, 2023.
- [10] S. P. L. A. M. Y. M. & C. P. A. Sutrisno, "Literature Review: Mitigasi Resiko dan Prosedur Penyelamatan pada Sistem Perkreditan Rakyat,)" *journal of Economic, Bussines and Accounting (COSTING)*, pp. 6(2), 1154-1167., 2023.