

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Pada bab ini akan dijelaskan mengenai proses, hasil, dan pembahasan penelitian yang dilakukan dalam rangka menganalisis kinerja model Regresi Linier untuk memprediksi kepuasan pelanggan pada Sambal Sinyar-Nyar. Sebagai alat bantu pengolahan data, digunakan *software* RapidMiner.

4.2 Implementasi Data Mining Regresi Linier

Implementasi data mining menggunakan algoritma regresi linier dilakukan melalui beberapa tahapan sesuai dengan proses KDD (*Knowledge Discovery in Database*), yang mencakup pemilihan data (*data selection*), pembersihan dan transformasi data, pemodelan (*modeling*), serta evaluasi model. Dalam penelitian ini, proses implementasi dilakukan menggunakan *platform* RapidMiner.

4.2.1 Data Selection

Pada tahap ini, data yang digunakan dalam penelitian ini dipilih berdasarkan relevansinya dengan tujuan untuk memprediksi tingkat kepuasan pelanggan sambal Sinyar-Nyar. Data yang dipilih meliputi informasi mengenai harga, kualitas rasa, dan pelayanan, yang dianggap sebagai faktor-faktor yang mempengaruhi kepuasan pelanggan. Data ini terdiri dari variabel independen yang menjadi faktor penentu dalam prediksi tingkat kepuasan. Data yang digunakan dalam analisis ini adalah hasil perhitungan rata-rata dari respons para responden pada masing-masing pertanyaan dalam kuesioner. Setiap variabel diukur dengan menggunakan skala Likert, yang kemudian dihitung rata-ratanya.

Data rata-rata ini kemudian diproses lebih lanjut dengan menggunakan aplikasi RapidMiner untuk dianalisis. Berikut adalah tabel yang menunjukkan hasil rata-rata untuk setiap variabel yang diukur berdasarkan kuesioner yang disebarkan kepada pelanggan sambal Sinyar-Nyar.

4.2.2 Preprocessing

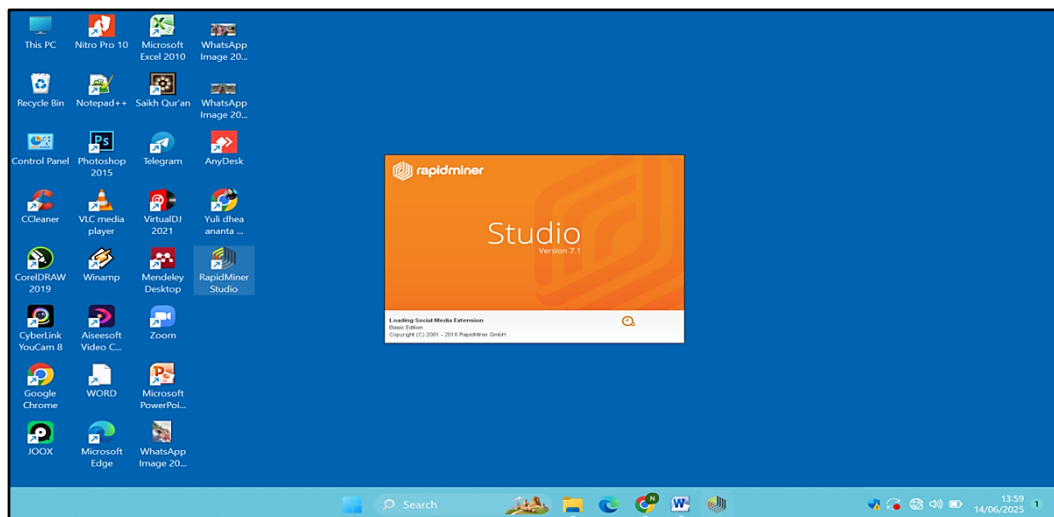
Tahap *preprocessing* atau pembersihan data memiliki tujuan utama untuk menghilangkan data yang tidak relevan atau noise, yaitu data yang tidak memiliki hubungan langsung dengan tujuan akhir dari proses data mining. Langkah-langkah yang dilakukan dalam proses pembersihan data mencakup identifikasi dan penghapusan data duplikat, pemeriksaan terhadap inkonsistensi data yang mungkin terjadi, serta perbaikan terhadap kesalahan-kesalahan yang ada, seperti kesalahan pengetikan atau data yang tidak sesuai dengan format yang diinginkan. Proses ini sangat penting untuk memastikan bahwa data yang digunakan dalam penelitian ini memiliki kualitas yang baik dan dapat memberikan hasil yang akurat dalam memprediksi tingkat kepuasan pelanggan sambal Sinyar-Nyar.

4.3 Implementasi RapidMiner pada Metode Regresi Linier

Setelah data dipersiapkan melalui tahap *data selection* dan *preprocessing* yang telah dijelaskan sebelumnya, langkah selanjutnya adalah melakukan implementasi model dalam perangkat lunak RapidMiner. Pada penelitian ini, model yang dipilih untuk dianalisis adalah *Linear Regression*, yang dipilih karena kemampuannya untuk memodelkan hubungan antara variabel independen dan dependen secara linier, serta kemudahan dalam menginterpretasi hasilnya. Adapun langkah-langkah dalam implementasi model *Linear Regression* ini sebagai berikut

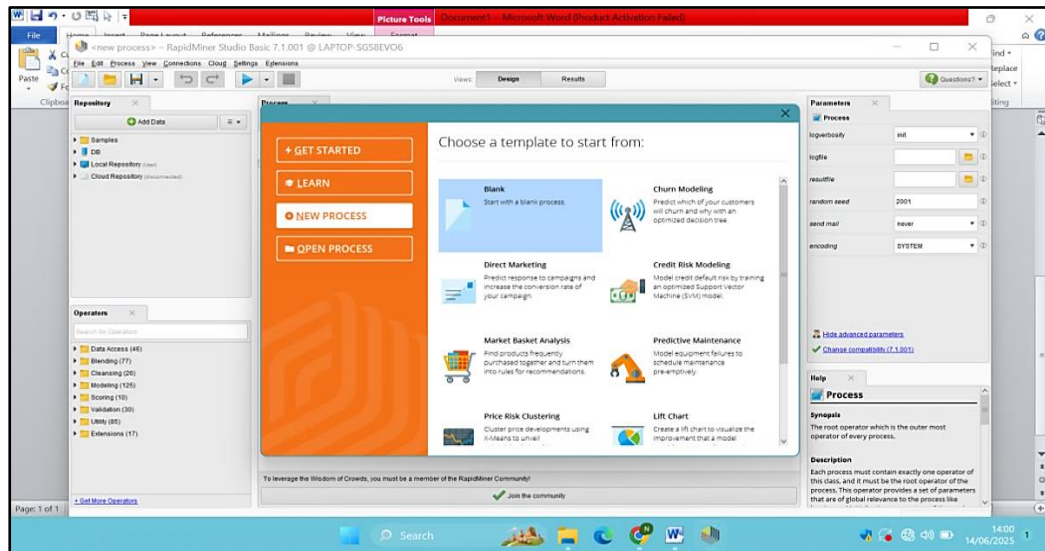
4.3.1 Proses Input Data ke RapidMiner

Berikut tampilan awal ketika aplikasi rapidminer ketika dijalankan.



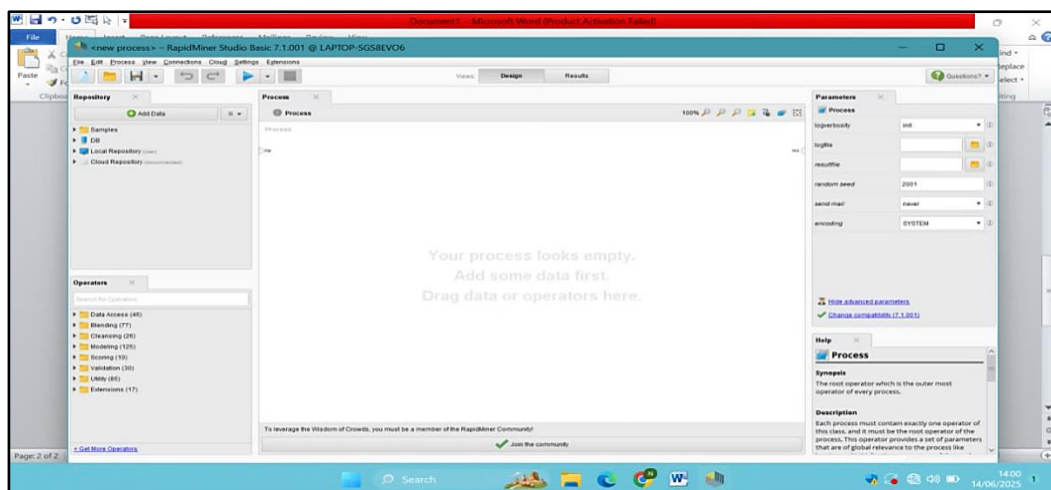
Gambar 4.1 Tampilan Awal RapidMiner

Setelah tampilan awal RapidMiner muncul yang berisi logo dan informasi perangkat lunak, pengguna akan diarahkan ke antarmuka utama yang menampilkan pilihan untuk membuat proyek baru atau membuka file yang sudah ada. Pada tahap ini, pengguna dapat memulai proses analisis dengan memilih untuk membuat *process* baru yang akan digunakan.



Gambar 4.2 Tampilan New Process

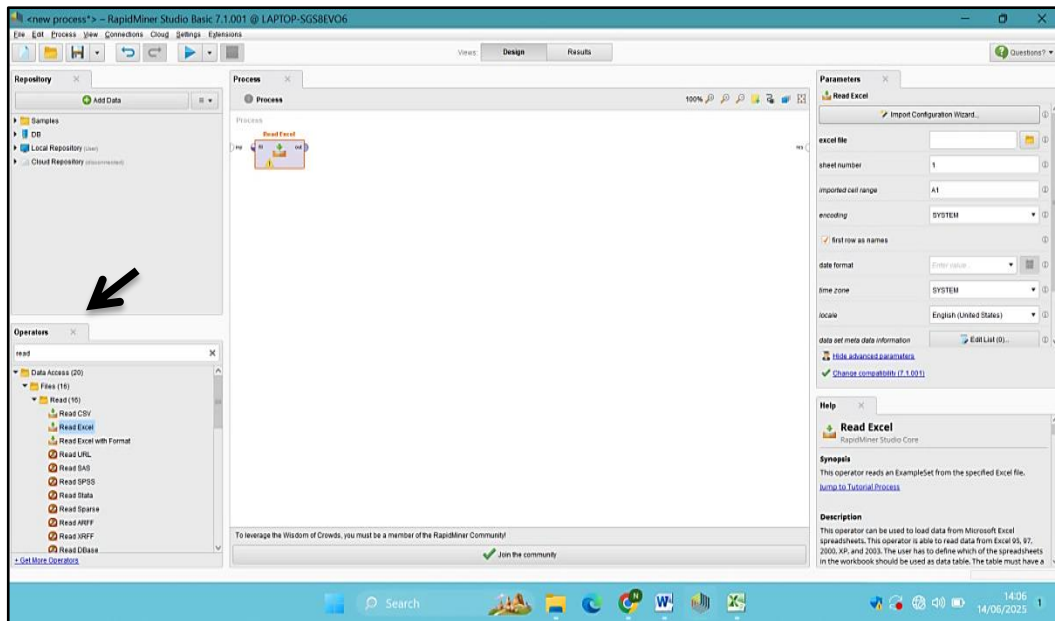
Tampilan awal ini menjadi pintu masuk sebelum melanjutkan ke tahap pengambilan data, seperti penggunaan operator *Read Excel* yang akan dibahas pada langkah berikutnya.



Gambar 4.3 Lembar Kerja RapidMiner

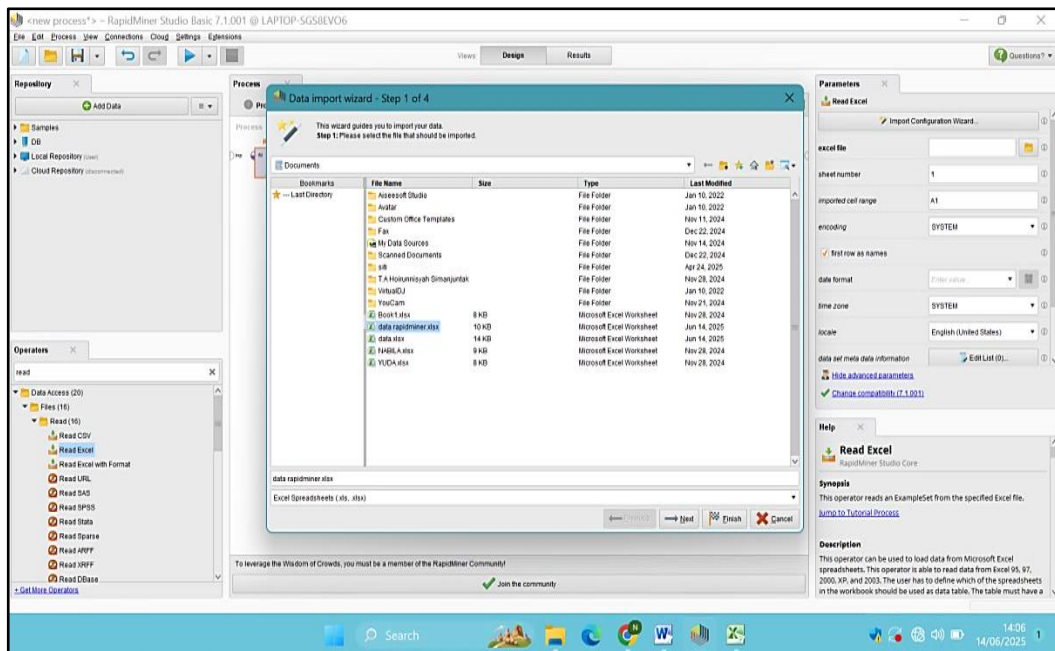
Setelah membuat lembar kerja baru, data dimasukkan menggunakan operator *Read Excel*. Operator ini diambil dari panel kiri, lalu diseret dan dijatuhkan ke area *view process*. Ikon *Read Excel* akan muncul di area proses.

Tampilan ini dapat dilihat pada Gambar 4.4 di bawah ini.



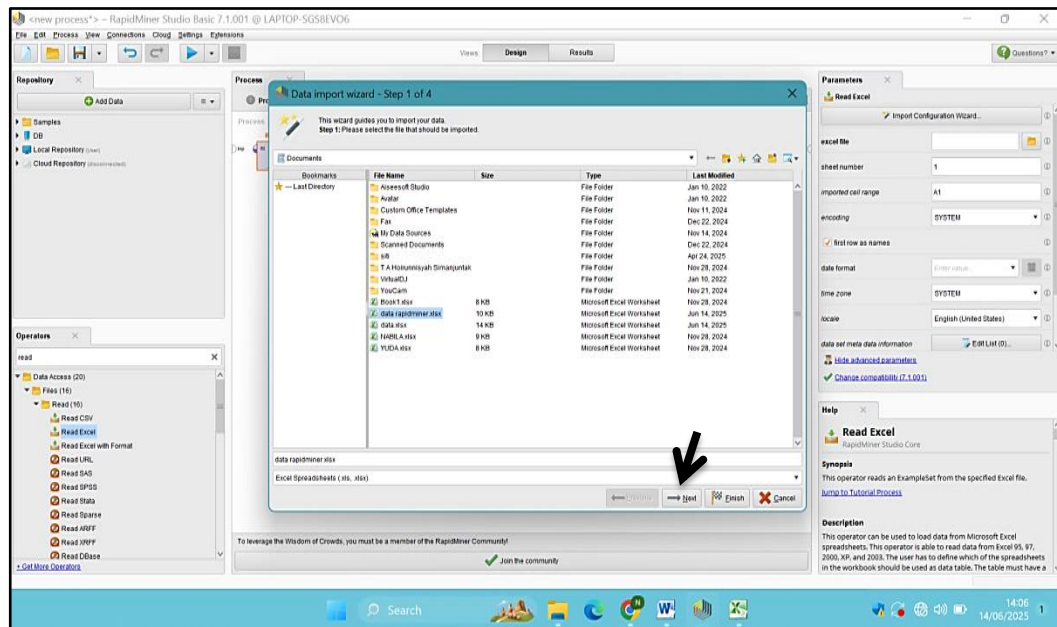
Gambar 4.4 Operator Read Excel

Selanjutnya, untuk mengkonfigurasi sumber data yang akan digunakan, klik dua kali pada operator tersebut atau pilih *Import Configuration Wizard* seperti terlihat pada Gambar 4.5 dibawah ini.



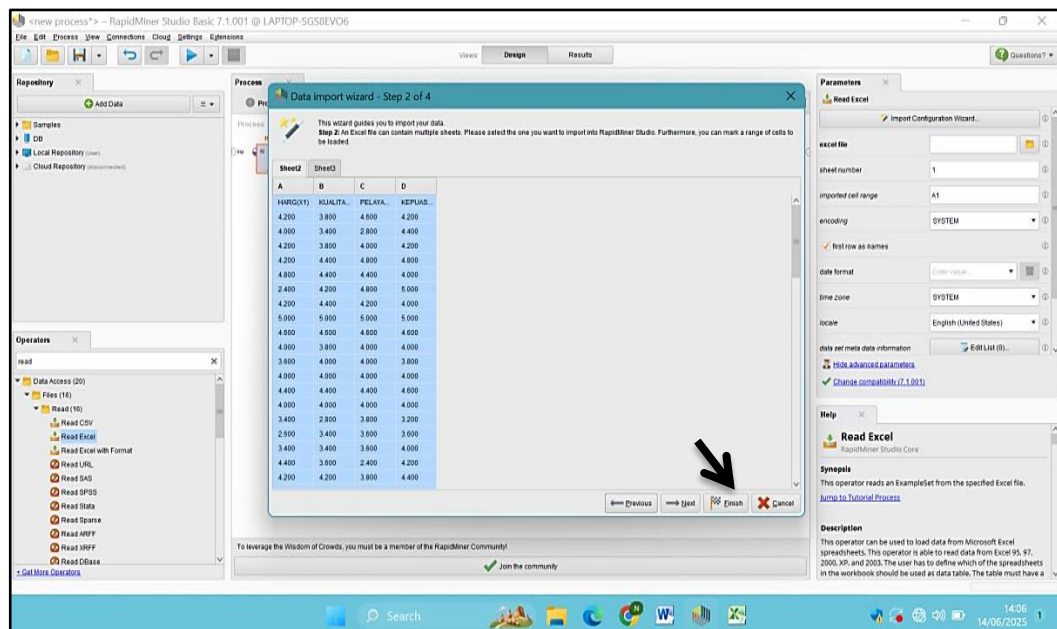
Gambar 4.5 Import Configuration Wizard

Melalui *wizard* ini, pengguna dapat menentukan lokasi file, lembar (*sheet*), serta rentang data yang akan dibaca ke dalam RapidMiner.



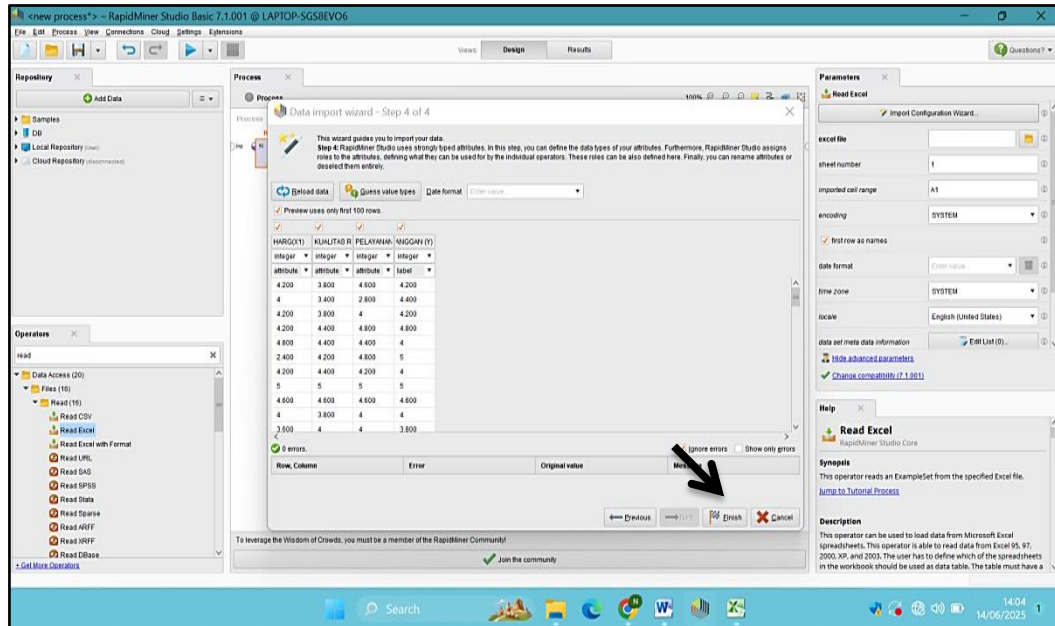
Gambar 4.6 Proses Import Data

Data yang telah dipilih selanjutnya yaitu klik Next kemudian akan muncul form data import wizard tahap 2 seperti Gambar 4.7 di bawah ini.



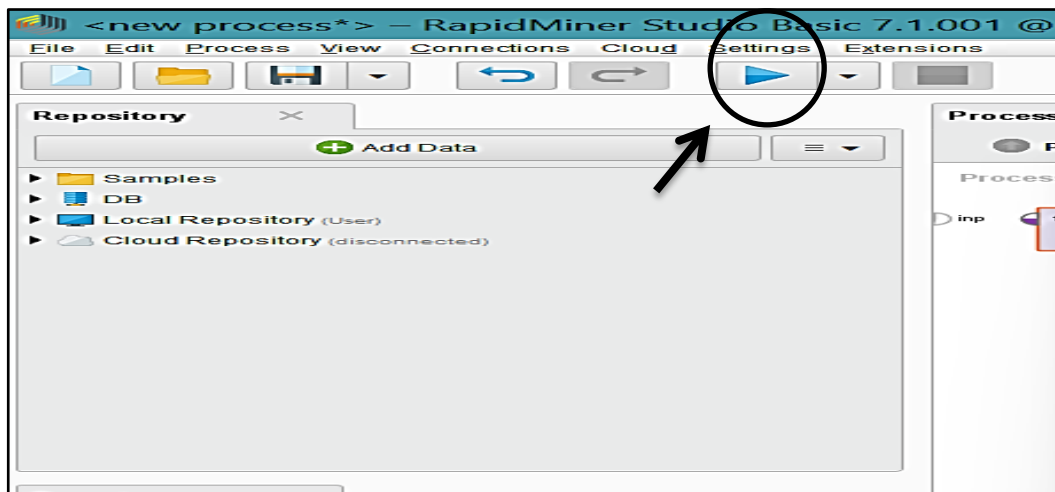
Gambar 4.7 Alur Proses Import

Selanjutnya tahap ke 3 *Data Import Wizard* di RapidMiner, menentukan tipe data untuk setiap kolom, seperti mengubah kolom yang terdeteksi sebagai *string* menjadi *integer* atau *double* jika diperlukan. Setelah memastikan semua data telah sesuai, klik *Finish* untuk menyelesaikan proses impor.



Gambar 4.8 Pemilihan Tipe

Klik tombol Run yang terletak di bagian atas antarmuka RapidMiner. Tombol Run akan memulai proses eksekusi.



Gambar 4.9 Ikon Run

RapidMiner akan membaca serta memproses data sesuai dengan langkah-langkah yang telah Anda tentukan dalam *view process*. Setelah proses selesai, hasil analisis atau output akan muncul di tab **Results**, yang dapat Anda lihat untuk mengevaluasi data yang telah diproses. Setelah setiap langkah atau operator selesai disusun dalam alur kerja, pastikan untuk melakukan **Run** pada setiap proses untuk memastikan bahwa data diproses dengan benar.

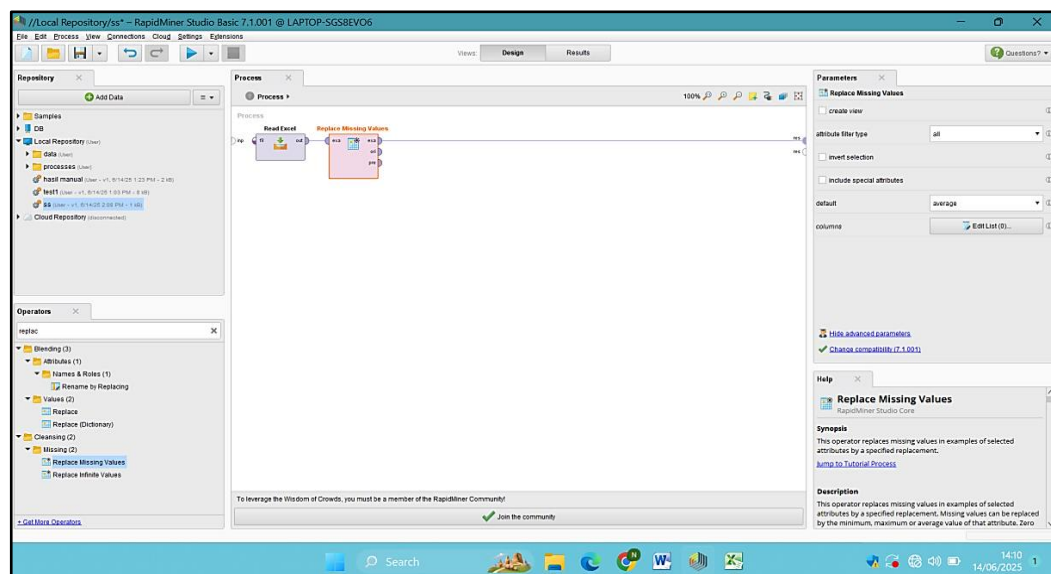
Row No.	KEPUASAN	HARGA	KUALITAS R.	PELAYANAN
1	4.200	4.200	3.800	4.600
2	4.400	4	3.400	2.800
3	4.200	4.200	3.800	4
4	4.800	4.200	4.400	4.800
5	4	4.800	4.400	4.400
6	5	2.400	4.200	4.800
7	4	4.200	4.400	4.200
8	5	5	5	5
9	4.600	4.600	4.600	4.600
10	4	4	3.800	4
11	3.800	3.600	4	4
12	4	4	4	4
13	4.600	4.400	4.400	4.400
14	4	4	4	4
15	3.200	3.400	2.800	3.800
16	3.600	2.600	3.400	3.600
17	4	3.400	3.400	3.600
18	4.200	4.400	3.600	2.400
19	4.400	4.200	4.200	3.800
20	3.600	4	3.600	3.800
21	4	4	4	4
22	5	4.800	4.400	4.200
23	5	4.400	4	4.200
24	4	4.800	4.400	4.600
25	4.600	4.200	4	5
26	4.800	4.400	4.400	4

Gambar 4.10 Hasil Import Data

Gambar 4.10 diatas adalah hasil Run Read excel dan data yang akan digunakan dalam melakukan analisis selanjutnya.

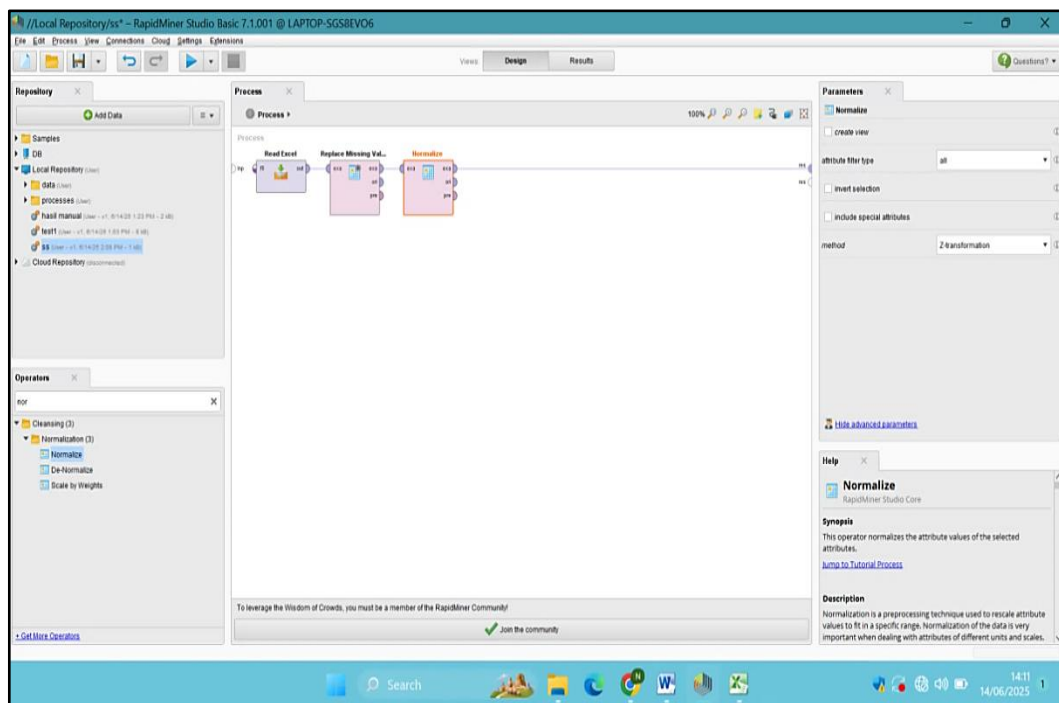
4.3.2 Pra-Pemrosesan Data (Data Preprocessing)

Setelah menjalankan proses dan memverifikasi data, langkah selanjutnya adalah melakukan *Preprocessing* yaitu dengan menangani *missing values* atau nilai yang hilang dan Normalisasi. Untuk itu, Anda dapat menggunakan operator *Replace Missing Values* di RapidMiner.



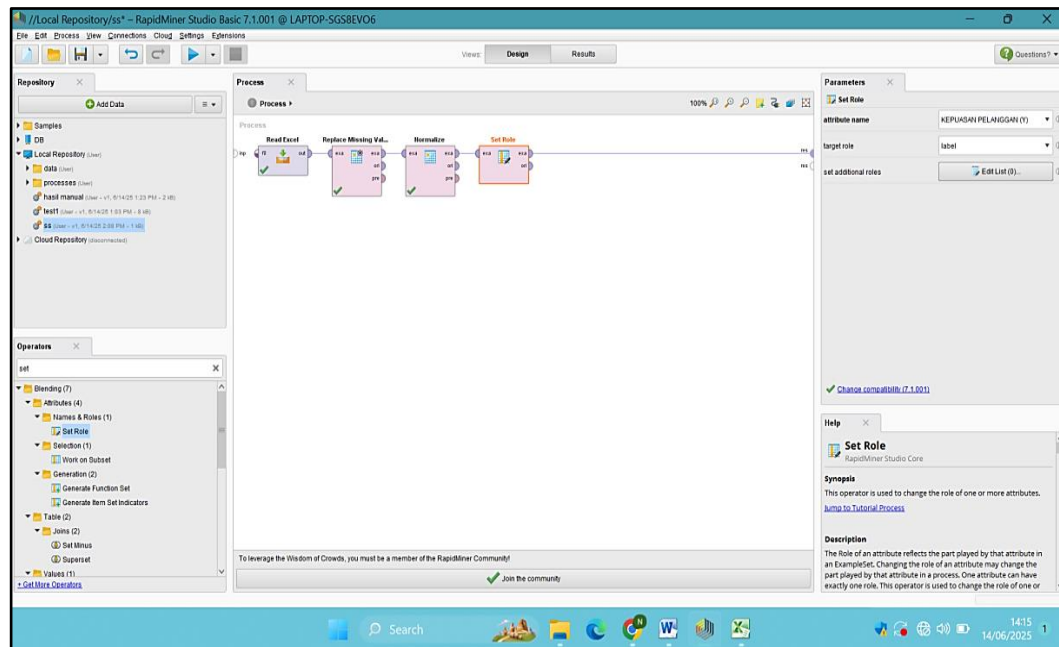
Gambar 4.11 Operator Replace Missing Values

Proses penanganan nilai hilang (*missing value*) digunakan untuk menggantikan data yang hilang dalam dataset dengan nilai tertentu, seperti rata-rata, median, atau modus, agar dataset menjadi lebih lengkap dan siap dianalisis. Langkah ini penting untuk memastikan bahwa analisis atau model tidak terganggu oleh data yang hilang. Setelah proses penggantian nilai hilang, tahap berikutnya adalah normalisasi seperti yang terlihat pada gambar 4.12 di bawah. bertujuan untuk mencegah variabel dengan rentang nilai yang besar mendominasi analisis.



Gambar 4.12 Operator *Normalization Data*

Operator *Normalization* digunakan untuk menormalkan data, mengubah skala nilai dalam dataset agar berada dalam rentang yang sama, biasanya antara 0 dan 1. Salah satu metode yang digunakan adalah *Z-Transformation*, yang mengubah data sehingga memiliki rata-rata 0 dan deviasi standar 1. Proses normalisasi penting untuk memastikan setiap fitur memiliki skala yang sebanding, yang mempengaruhi hasil analisis atau model.

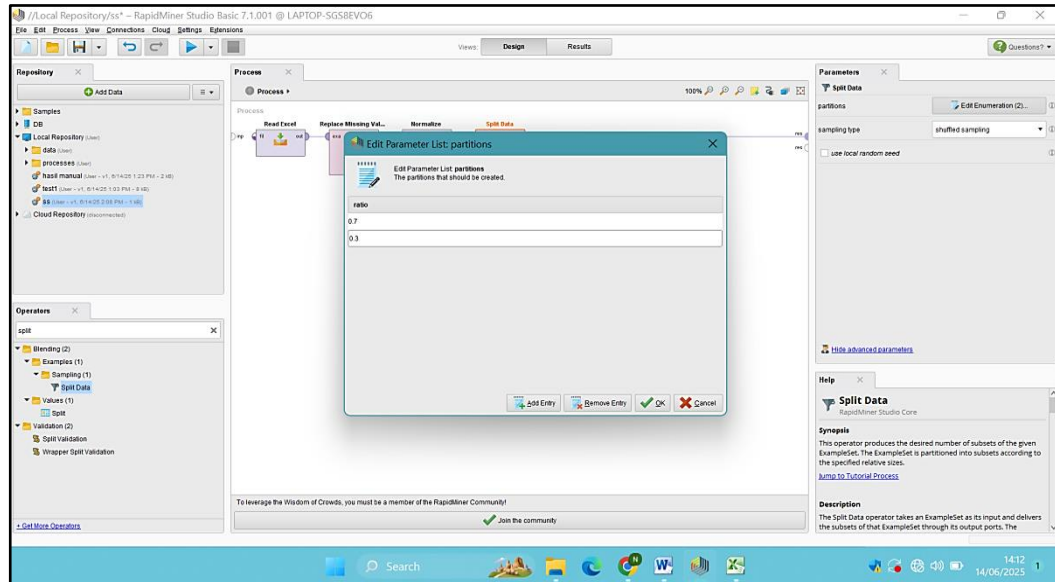


Gambar 4.13 Operator Set Role

Proses *Set Role* digunakan untuk menetapkan peran pada atribut dalam dataset yang akan dianalisis. Pada parameter *Attribute Name*, kita memilih atribut yang akan diberi peran, seperti *KEPUASAN PELANGGAN (Y)*. Di *Target Role*, kita menentukan peran yang akan diberikan, misalnya *label*, yang menunjukkan atribut tersebut sebagai target prediksi dalam model. Pada *Add Additional Roles*, kita juga dapat menambahkan peran lain pada atribut yang sama, memberikan fleksibilitas dalam pemanfaatan atribut untuk berbagai fungsi analisis.

4.3.3 Proses Split Data

Sebelum melakukan analisis, langkah pertama yang dilakukan adalah membagi dataset menjadi dua bagian, yaitu data latih (*training*) dan data uji (*testing*). Pada gambar ini, operator *Split Data* digunakan untuk membagi dataset sesuai dengan kebutuhan. Pembagian data umumnya dilakukan dengan rasio tertentu, seperti 70% untuk data latih dan 30% untuk data uji, guna memastikan bahwa model dapat dievaluasi menggunakan data yang belum pernah ditemui sebelumnya terlihat pada gambar 4.14 di bawah ini.

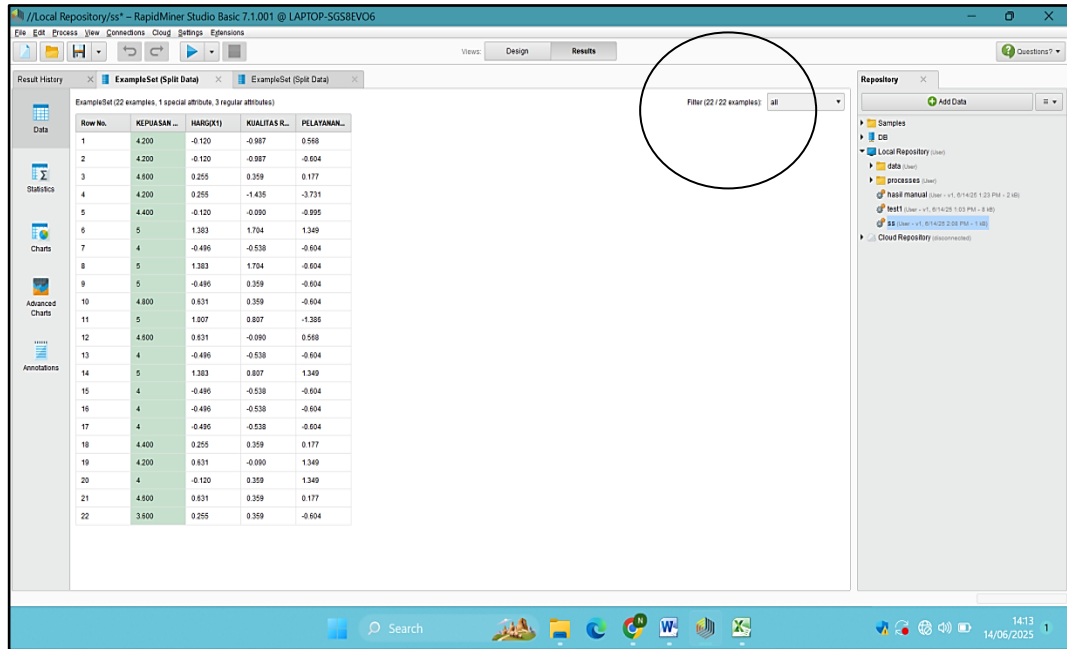


Gambar 4.14 Operator *Split Data*

Setelah melakukan pembagian data dengan rasio 0.7 untuk data latihan (*training*) dan 0.3 untuk data uji (*testing*), hasilnya adalah 53 data latihan yang digunakan untuk melatih model dan 22 data uji yang digunakan untuk menguji kinerja model. Pembagian ini memastikan bahwa model dapat dilatih dengan cukup data dan diuji dengan data yang belum pernah dilihat sebelumnya. Hasil pembagian data ini dapat dilihat pada gambar 4.15 dan 4.16 di bawah.

Row No.	KEMASAN	NARGOKT1	KUALITAS R.	PELAYANAN
1	4.400	-0.495	-1.883	-2.949
2	4.800	-0.120	0.359	0.959
3	4	1.007	0.359	0.177
4	5	-3.501	-0.090	0.959
5	4	-0.120	0.359	-0.214
6	5	1.383	1.704	1.349
7	4.900	0.831	0.807	0.588
8	4	-0.495	-0.987	-0.604
9	3.800	-1.247	-0.538	-0.604
10	4	-0.495	-0.538	-0.604
11	4	-0.495	-0.538	-0.604
12	3.200	-1.623	-3.229	-0.995
13	3.600	-3.126	-1.883	-1.369
14	4	-1.623	-1.883	-1.369
15	3.600	-0.495	-1.435	-0.995
16	4	-0.495	-0.538	-0.604
17	5	0.255	0.359	-0.214
18	5	0.255	-0.538	-0.214
19	4	1.007	0.359	0.588
20	4.900	-0.120	-0.538	1.349
21	4.900	0.255	0.359	-0.604
22	5	-0.120	-0.538	1.349
23	5	1.383	1.704	1.349
24	5	1.383	1.704	1.349
25	5	1.383	1.256	1.349
26	4	1.383	-0.538	0.588

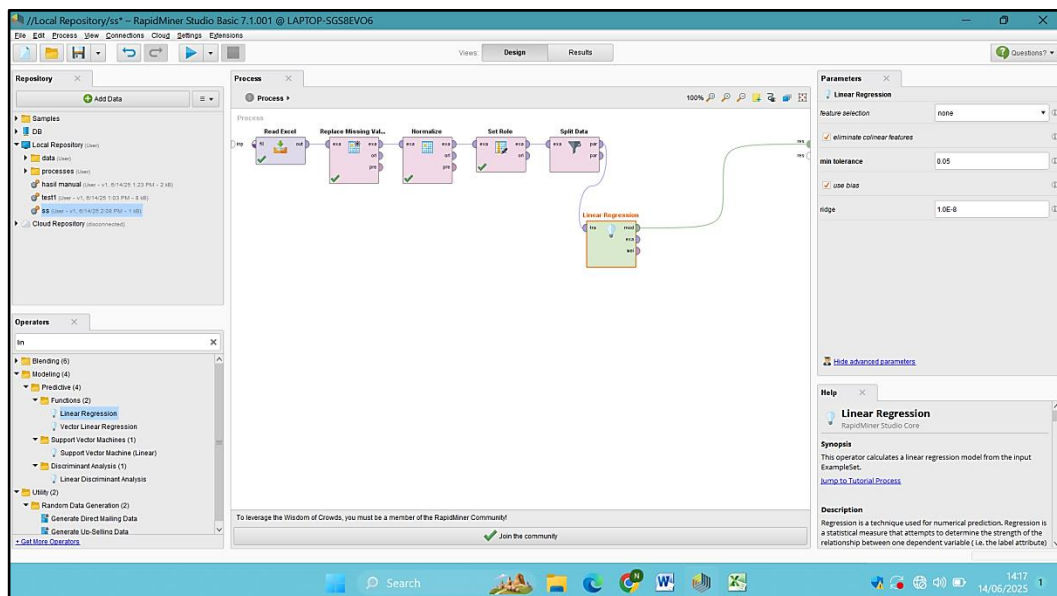
Gambar 4.15 Data *Training*



Gambar 4.16 Data Testing

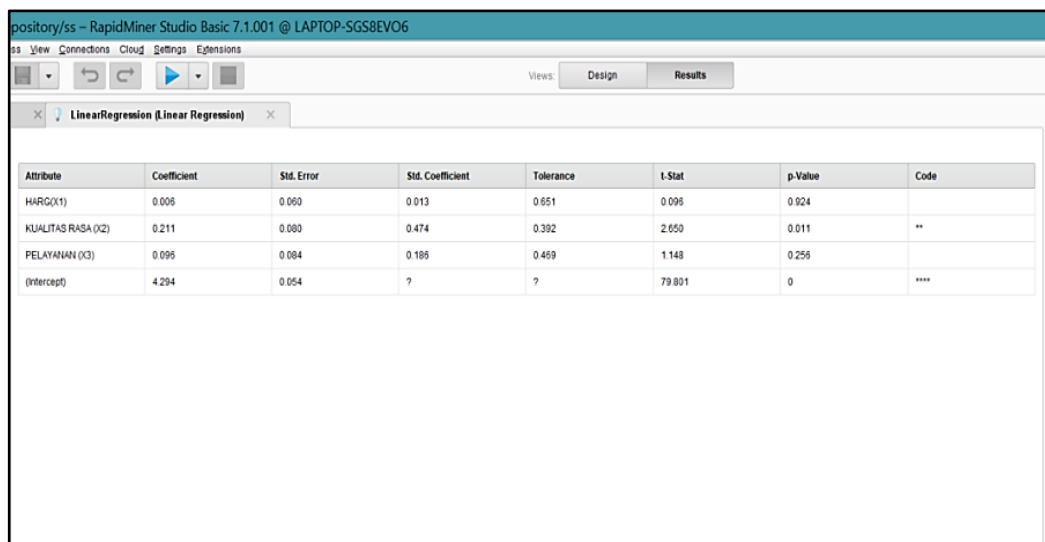
4.3.4 Penerapan Metode *Linear Regression*

Setelah data dibersihkan dan dipersiapkan, operator *Linear Regression* diterapkan untuk membangun model regresi linier. Proses ini melibatkan pelatihan model untuk memprediksi keuntungan berdasarkan fitur yang tersedia. Algoritma regresi linier akan digunakan untuk menganalisis hubungan antara variabel-variabel yang ada dalam dataset.



Gambar 4.17 Pemodelan Operator *Linear Regression*

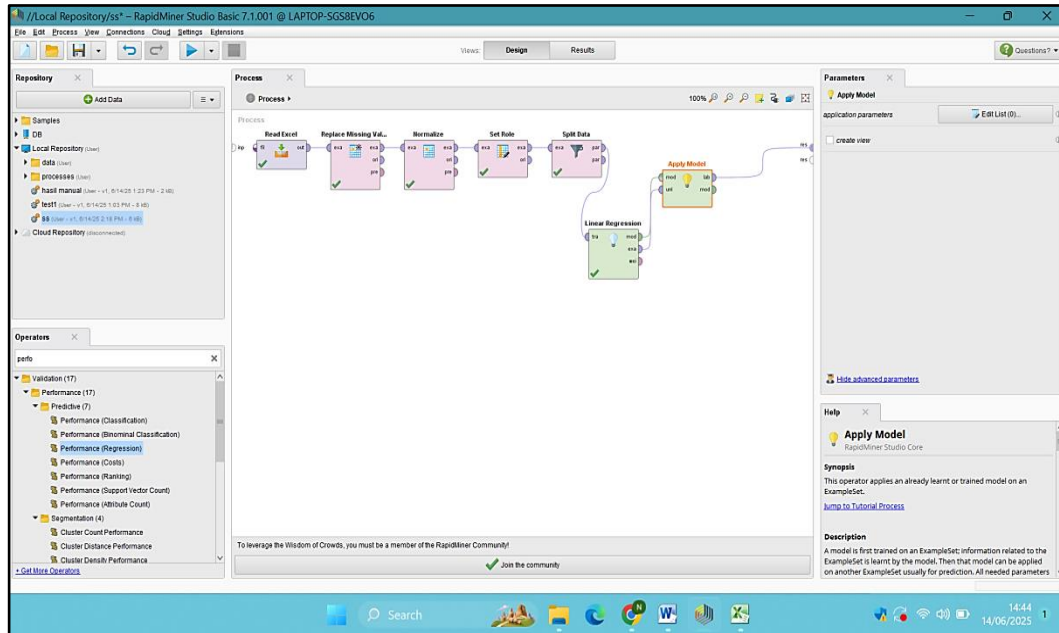
Proses ini melibatkan penggunaan algoritma regresi linier pada data latih (*training data*) untuk membangun model. Regresi linier adalah metode statistik yang digunakan untuk mengidentifikasi hubungan antara variabel dependen (*target*) dengan satu atau lebih variabel independen (*fitur*). Operator ini mengatur parameter yang telah ditentukan, seperti pemilihan fitur yang relevan (*feature selection*), serta menyesuaikan nilai toleransi dan bias sesuai dengan konfigurasi yang diinginkan.



Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
HARGO(X1)	0.006	0.060	0.013	0.651	0.096	0.924	
KUALITAS RASA (X2)	0.211	0.080	0.474	0.392	2.650	0.011	**
PELAYANAN (X3)	0.096	0.084	0.186	0.469	1.148	0.256	
(Intercept)	4.294	0.054	?	?	79.801	0	****

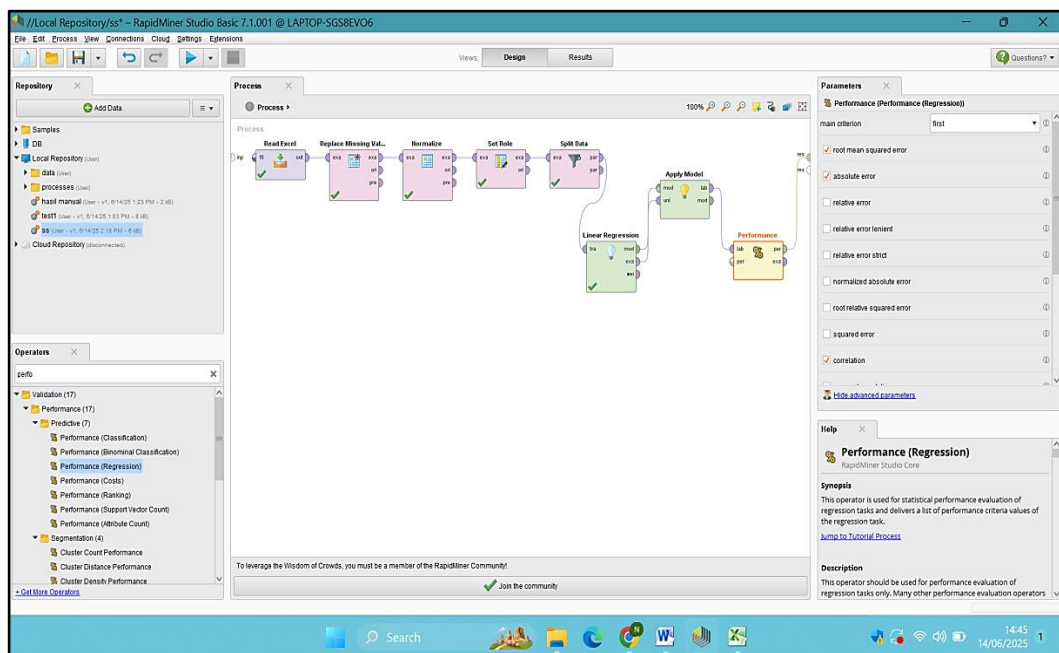
Gambar 4.18 Hasil Regresi Linier

Hasil regresi linier yang ditampilkan pada gambar ini memberikan informasi terkait pengaruh masing-masing atribut terhadap model yang dibangun. Atribut seperti harga (X_1), kualitas rasa (X_2), dan pelayanan (X_3) memiliki koefisien yang mencerminkan kontribusi masing-masing variabel terhadap prediksi target. Nilai *Std. Error* mengindikasikan tingkat ketelitian estimasi koefisien, sementara *Std. Coefficient* menunjukkan besarnya pengaruh relatif setiap atribut setelah memperhitungkan perbedaan skala variabel. *Tolerance* menggambarkan potensi multikolinearitas antar atribut, sedangkan *T-Stat* dan *P-Value* digunakan untuk menguji signifikansi statistik dari koefisien yang dihasilkan. Nilai *p-value* yang lebih kecil dari 0.05 menunjukkan bahwa atribut tersebut memiliki pengaruh signifikan terhadap model yang sedang dikembangkan.



Gambar 4.19 Operator *Apply Model* Regresi Linier

Setelah model regresi linier dilatih menggunakan data latih, langkah selanjutnya adalah menerapkan model yang telah terlatih pada data uji untuk mengevaluasi seberapa efektif model dalam memprediksi target berdasarkan fitur yang ada. Pada tahap ini, operator akan menggunakan model yang telah dibangun untuk menghasilkan prediksi pada data uji.



Gambar 4.20 Operator *Performance* Regresi Linier

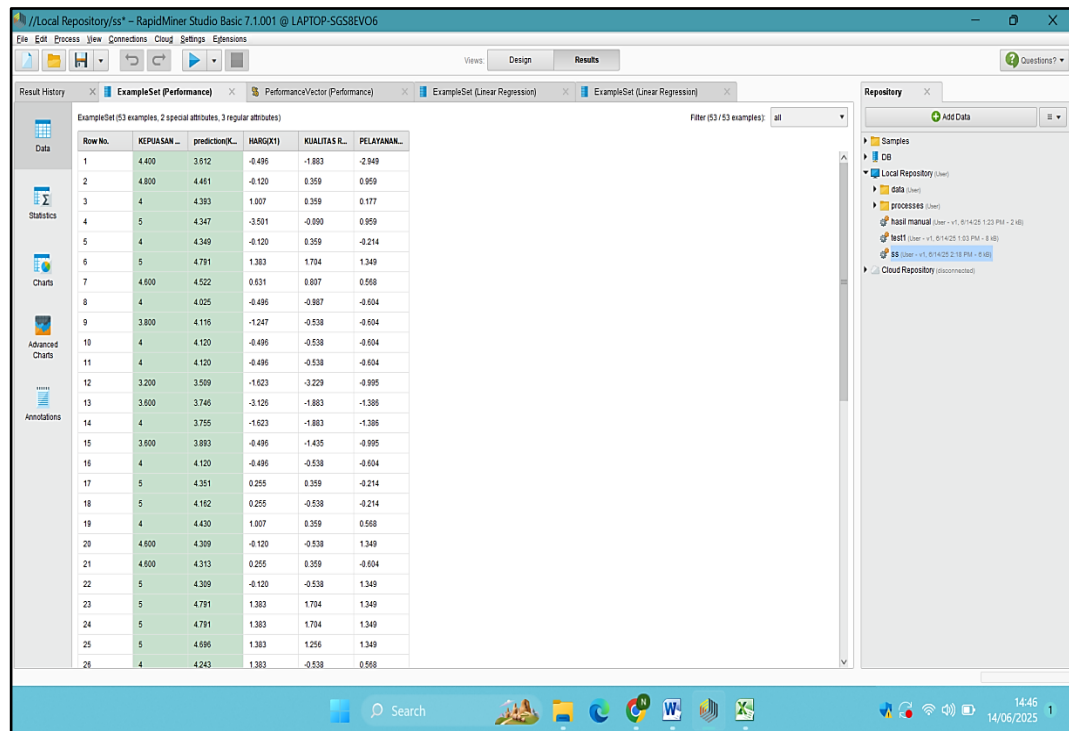
Setelah model diterapkan pada data uji, langkah berikutnya adalah mengevaluasi kinerjanya. Pada tahap ini, kita mengukur sejauh mana model dapat memprediksi hasil dengan akurat menggunakan data yang belum pernah diproses sebelumnya. Untuk menilai kualitas prediksi yang dihasilkan oleh model regresi linier, digunakan beberapa metrik kinerja, seperti RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*), dan koefisien korelasi. Koefisien korelasi digunakan untuk mengukur sejauh mana hubungan antara prediksi model dan nilai sebenarnya. Hasil evaluasi ini memberikan gambaran mengenai akurasi model, yang dapat dilihat pada Gambar 4.21 di bawah ini.



Gambar 4.21 Hasil *Performance* Regresi Linier

Evaluasi yang digunakan untuk menilai akurasi dan kualitas prediksi model. *Root Mean Squared Error* (RMSE) sebesar 0.369 mencerminkan tingkat kesalahan rata-rata model dalam memprediksi nilai, di mana nilai yang lebih rendah menandakan akurasi model yang lebih tinggi. *Absolute Error* (MAE) yang bernilai 0.209 menunjukkan rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Semakin kecil nilai MAE, semakin baik model dalam menghasilkan prediksi yang mendekati nilai sebenarnya.

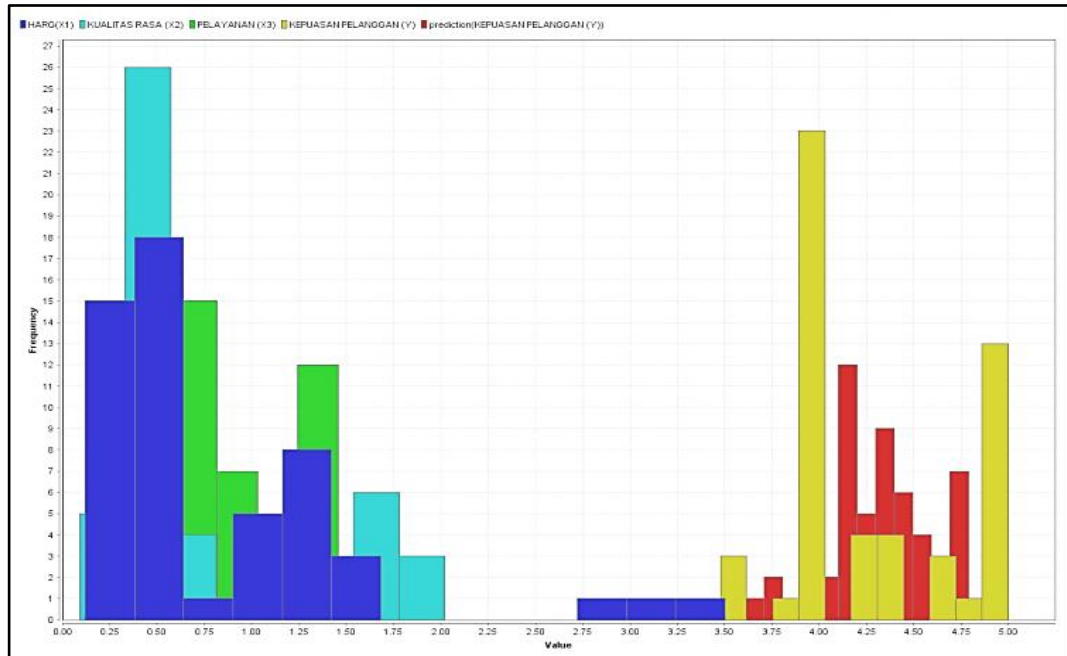
Selain itu, nilai korelasi sebesar 0.630 menggambarkan kekuatan hubungan linear antara nilai prediksi dan nilai aktual. Nilai korelasi yang lebih tinggi menunjukkan bahwa model berhasil menangkap hubungan antara variable dengan lebih baik, dengan nilai yang semakin mendekati 1 menunjukkan kecocokan yang lebih kuat antara hasil prediksi dan data aktual.



Gambar 4.22 Hasil Performance

Hasil penerapan model regresi linier pada dataset yang mencakup atribut seperti Kepuasan Pelanggan, Harga, Kualitas Rasa, dan Pelayanan menunjukkan perbandingan antara nilai prediksi yang dihasilkan oleh model dan nilai aktual untuk masing-masing atribut. Tabel ini memberikan gambaran yang jelas mengenai seberapa baik model dapat memprediksi nilai yang diharapkan berdasarkan fitur yang tersedia dalam dataset.

Dengan membandingkan prediksi dan nilai aktual, kita dapat mengevaluasi akurasi model dalam menggeneralisasi pola yang ditemukan pada data latih ke data uji. Selain itu, proses ini juga memungkinkan untuk menilai apakah model mengalami bias atau tidak, serta mengukur sejauh mana kesalahan prediksi terjadi, yang sangat penting untuk mengevaluasi kinerja model dan memastikan bahwa model berfungsi secara optimal dalam memprediksi data yang belum pernah dilihat sebelumnya.



Gambar 4.23 Histogram

Histogram ini menggambarkan distribusi nilai prediksi model regresi linier dibandingkan dengan nilai aktual dari data uji. Setiap warna mewakili distribusi atribut berbeda, seperti Kepuasan Pelanggan, Harga, Kualitas Rasa, dan Pelayanan. Histogram ini memberikan gambaran visual tentang seberapa baik model memprediksi nilai yang diharapkan, serta memperlihatkan apakah model menghasilkan prediksi yang konsisten atau menyimpang. Selain itu, visualisasi ini membantu mendeteksi adanya outlier atau pola yang memerlukan perhatian lebih.

4.4 Pembahasan

Penelitian ini bertujuan untuk memprediksi tingkat kepuasan pelanggan pada Sambal Sinyar-Nyar menggunakan metode Regresi Linier. Model regresi linier ini diterapkan untuk menganalisis hubungan antara beberapa faktor yang mempengaruhi kepuasan pelanggan, yaitu harga, kualitas rasa, dan pelayanan. Tujuan utama dari pemodelan ini adalah untuk mengukur seberapa akurat metode regresi linier dalam memprediksi tingkat kepuasan pelanggan berdasarkan data yang telah dikumpulkan.

Setelah proses pemodelan selesai, evaluasi kinerja model dilakukan dengan menggunakan beberapa metrik kesalahan, yaitu *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). Metrik ini digunakan untuk menilai akurasi prediksi yang dihasilkan oleh model, dengan tujuan memastikan bahwa model dapat memprediksi dengan baik berdasarkan data uji yang belum pernah dilihat sebelumnya.

Berdasarkan hasil evaluasi yang diperoleh, model Regresi Linier menunjukkan kinerja yang lebih baik dibandingkan dengan metode lainnya, karena memiliki nilai kesalahan yang lebih rendah. Nilai RMSE pada model regresi linier tercatat sebesar 0.369. Nilai RMSE yang lebih rendah ini menunjukkan bahwa model regresi linier memiliki tingkat kesalahan yang lebih kecil dan lebih akurat dalam memprediksi kepuasan pelanggan. Selain itu, nilai MAE pada model regresi linier juga lebih kecil, yang semakin memperkuat keakuratan prediksi model ini.

Hasil evaluasi lebih lanjut menunjukkan bahwa nilai koefisien korelasi sebesar 0.630 menggambarkan hubungan linear yang signifikan antara nilai prediksi dan nilai aktual. Semakin mendekati angka 1, semakin kuat hubungan antara variabel dependen dan independen dalam model, yang menunjukkan bahwa model regresi linier ini dapat diandalkan untuk memprediksi tingkat kepuasan pelanggan secara akurat. Koefisien-koefisien regresi untuk atribut yang digunakan dalam model adalah sebagai berikut, Harga (X_1) nilai koefisien 0.005, Kualitas Rasa (X_2) nilai koefisien 0.211, Pelayanan (X_3) nilai koefisien 0.096, dan Intercept nilai koefisien 4.234.

Berdasarkan hasil koefisien dan p-value yang ada, dapat disimpulkan bahwa harga dan kualitas rasa memiliki pengaruh signifikan terhadap kepuasan pelanggan, sementara pelayanan tidak menunjukkan pengaruh signifikan terhadap kepuasan pelanggan dalam model ini. Oleh karena itu, metode Regresi Linier merupakan pilihan yang tepat dan lebih efektif untuk memprediksi kepuasan pelanggan pada Sambal Sinyar-Nyar. Model regresi linier terbukti memberikan hasil yang lebih akurat dan dapat diandalkan untuk prediksi kepuasan pelanggan dalam penelitian.