

BAB IV

HASIL DAN PEMBAHASAN

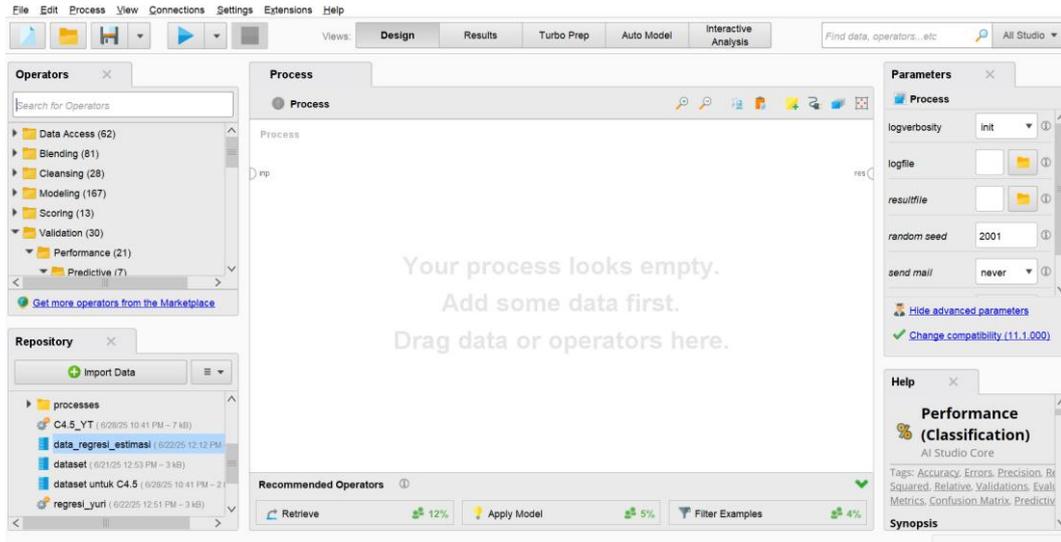
4.1 Pengujian Metode C4.5

Pengujian dalam penelitian ini dilakukan dengan menggunakan Microsoft Excel untuk mempermudah pemrosesan awal data, klasifikasi, perhitungan rumus algoritma C4.5. Penggunaan Excel bertujuan agar setiap tahapan dalam analisis data dapat dilakukan secara manual dan transparan, sehingga memudahkan dalam memahami logika di balik metode yang digunakan.

Data yang digunakan berasal dari laporan hasil panen bulanan PT. Sinar Pandawa selama tiga tahun, dengan tiga variabel utama: curah hujan (mm), jumlah pupuk (kg), dan hasil panen (ton).

4.1.1 Pengujian Menggunakan RapidMiner

Pengujian menggunakan RapidMiner bertujuan untuk mengimplementasikan metode algoritma C4.5 secara otomatis, serta untuk mengevaluasi performa model dengan lebih sistematis dan efisien. RapidMiner dipilih karena menyediakan antarmuka grafis (GUI) yang memudahkan pengguna dalam membangun workflow data mining tanpa perlu menulis kode pemrograman secara manual.



Gambar 4. 1 Aplikasi Altair RapidMiner 2025.1.0

Pada tahap ini dilakukan pengujian model prediksi hasil panen kelapa sawit di PT. Sinar Pandawa menggunakan perangkat lunak RapidMiner. Data yang telah dipraproses dimasukkan ke dalam RapidMiner dalam bentuk data tabular.

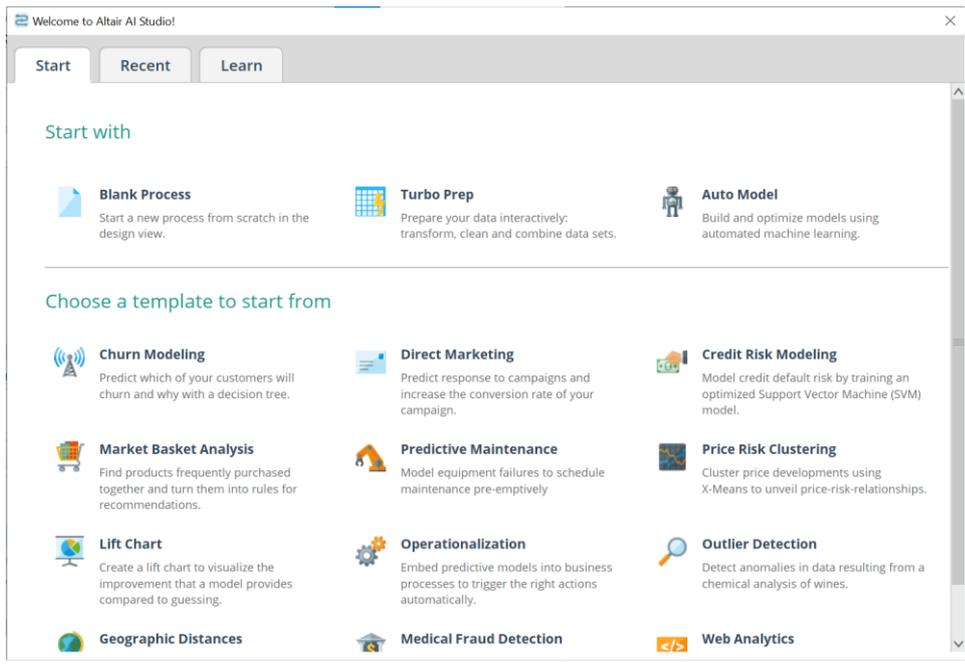
Tabel 4. 1 Data yang telah dipraproses

Curah hujan (mm)	Pupuk (kg)	Hasil Panen (ton)	Kategori Curah hujan
262	59000	2000	Rendah
375	50000	1900	Sedang
157	80000	2000	Rendah
530	24000	1300	Sedang
325	42000	1800	Rendah
621	27000	1300	Tinggi
312	49000	1800	Rendah
714	23000	1500	Tinggi
380	47000	1700	Sedang
215	30000	2000	Rendah
137	50000	2000	Rendah
521	35000	1500	Sedang
421	15000	1600	Sedang
147	60000	2000	Rendah
214	43000	2000	Rendah
363	40000	1800	Sedang
210	60000	1900	Rendah
530	23000	1300	Sedang
670	15000	1300	Tinggi

452	30000	1500	Sedang
320	45000	1400	Rendah
140	50000	2000	Rendah
432	30000	1500	Sedang
221	52000	2000	Rendah
690	27000	1500	Tinggi
150	83000	2000	Rendah
289	59000	1900	Rendah
227	73000	1900	Rendah
785	25000	1300	Tinggi
343	51000	1700	Sedang
333	57000	1800	Rendah
130	87000	2000	Rendah
143	80000	2000	Rendah
530	25000	1300	Sedang
152	85000	2000	Rendah
380	45000	1500	Sedang
150	80000	2000	Rendah
542	32000	1300	Sedang
250	70000	1800	Rendah
321	50000	1600	Rendah
130	75000	2000	Rendah
240	50000	1800	Rendah

Berikut langkah-langkah pengujian yang dilakukan:

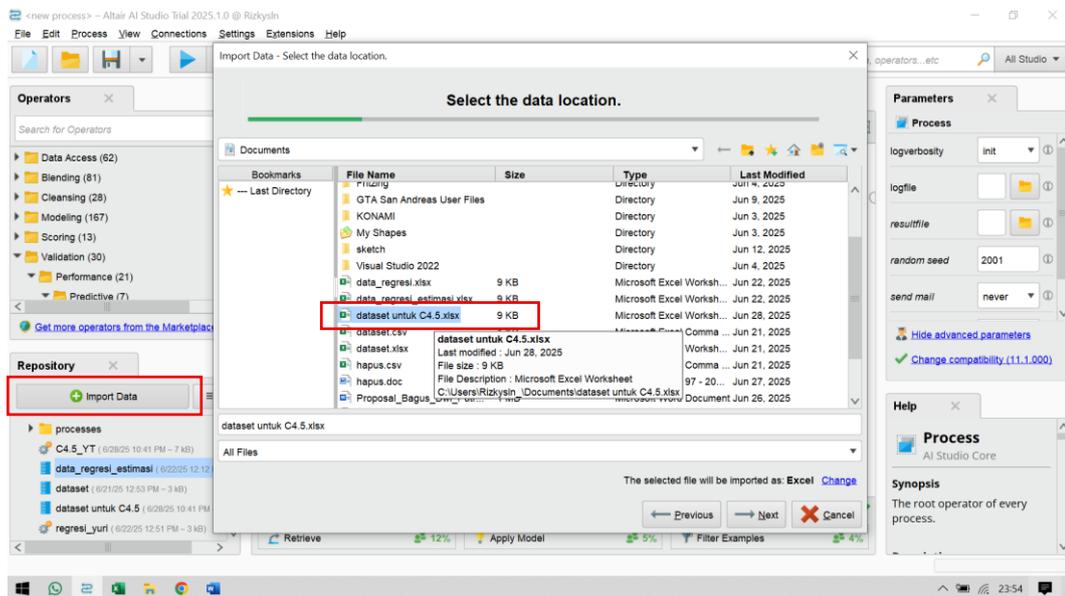
Langkah awal, buka perangkat lunak RapidMiner. Klik menu File → New Process, lalu pilih submenu Blank (Start with a blank process). Tampilan awal dapat dilihat pada Gambar 4.2.



Gambar 4. 2 Tampilan Awal Altair RapidMiner AI Studio

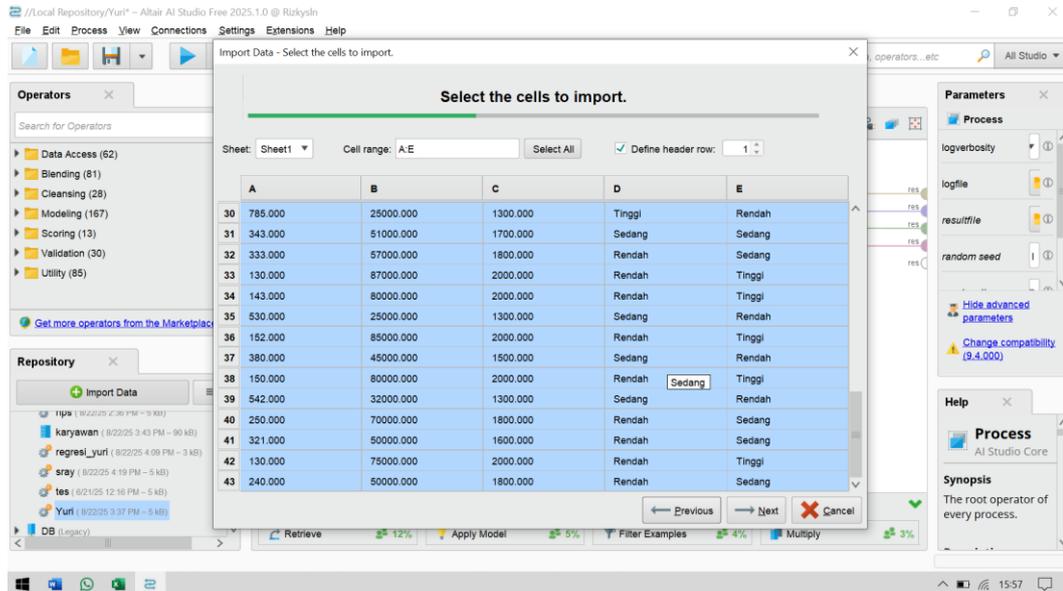
4.1.2 Langkah Pengujian

Klik menu Repository di bagian kiri, lalu klik Import Data seperti pada Gambar 4.3.



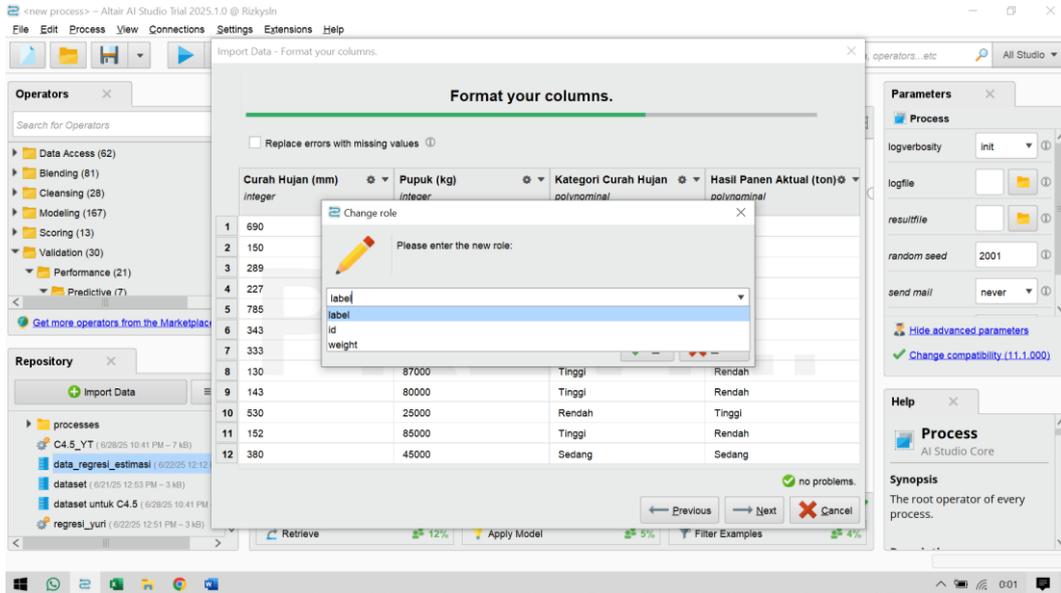
Gambar 4. 3 Import dan Pilih Data Excel

Setelah melakukan pencarian data yang akan diinput maka akan terlihat pada kolom data yang sudah disiapkan di file *Excel* dengan format *.xlsx* yang sudah berhasil dibaca oleh perangkat lunak Rapidminer. Selanjutnya, memilih lokasi data yang akan diolah berdasarkan *Sheet* apabila data *Excel* yang diperoleh memiliki *Sheet* yang banyak seperti Gambar 4.4 lalu klik *Next*



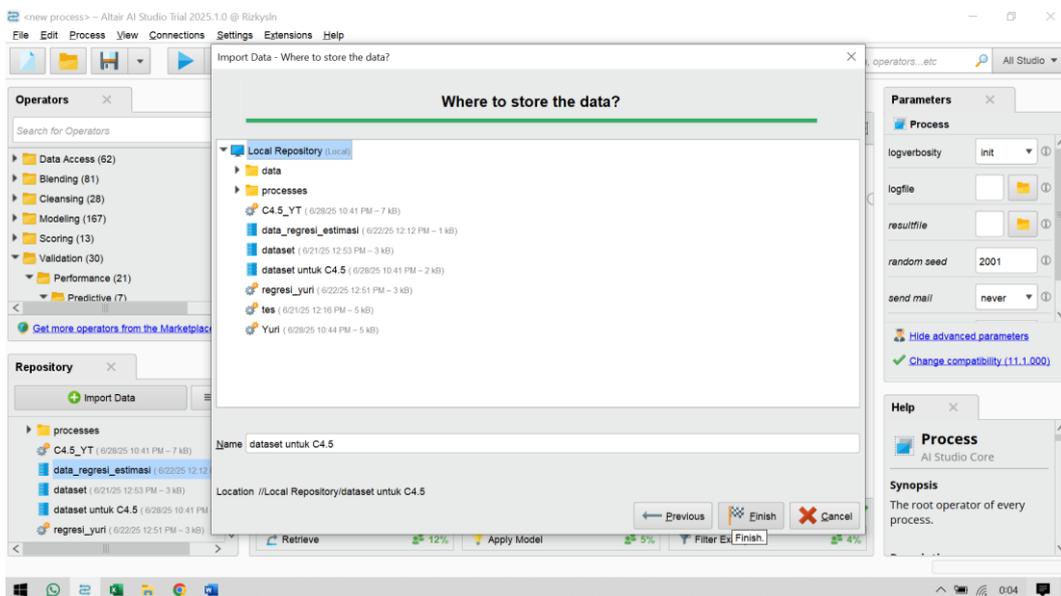
Gambar 4. 4 Pilih Rentang Kolom yang Digunakan

Setelah memilih lokasi *Sheet* yang akan diolah, langkah selanjutnya klik segitiga pada *Transaksi* lalu pilih *Change Type* dan *Polynomial*. Berikutnya klik lagi segitiga pada *Transaksi*, lalu pilih *Change Rule* dan pilih label lalu klik *OK* dan *Next*. Seperti pada Gambar 4.5.



Gambar 4.5 Tampilan *Change Rule*

Selanjutnya memberi nama dan lokasi penyimpanan data yang akan diproses lalu klik *Finish* seperti Gambar 4.6.



Gambar 4.6 Lokasi Penyimpanan Data Yang Akan Diproses

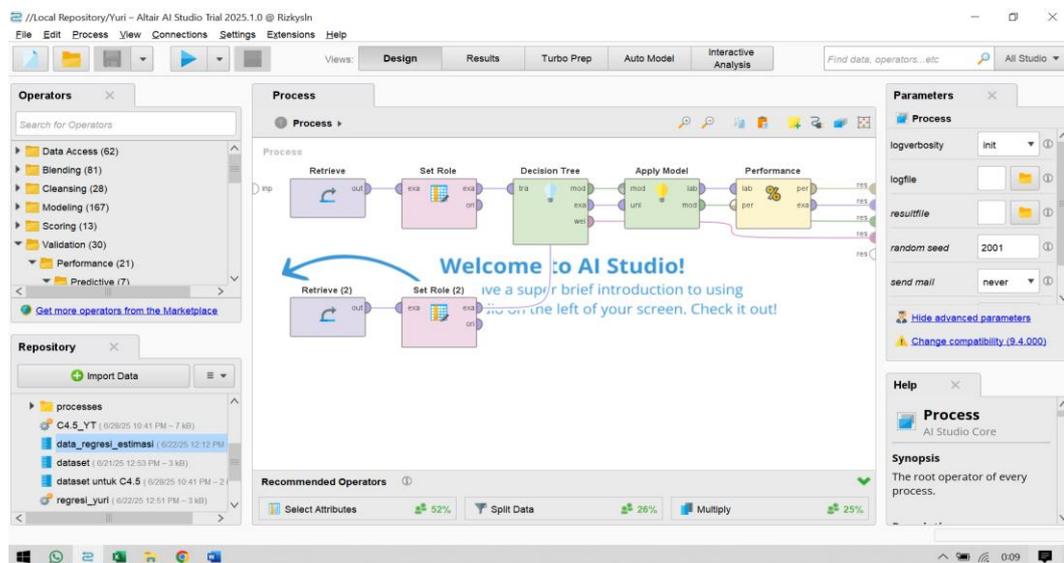
Setelah data berhasil diinputkan, langkah selanjutnya adalah membuka tampilan Design View pada RapidMiner. Pada tampilan ini, pengguna akan

diarahkan ke lembar kerja utama (Main Process) yang berfungsi sebagai area visual untuk membangun alur proses analisis data.

Seperti terlihat pada Gambar 4.7, dalam workspace ini pengguna dapat:

1. Menambahkan data (melalui operator *Retrieve*)
2. Menetapkan atribut target (*Set Role*)
3. Membuat model klasifikasi menggunakan *Decision Tree*
4. Menerapkan model ke data uji dengan *Apply Model*
5. Mengevaluasi performa model melalui *Performance*

Setiap operator disambungkan secara berurutan untuk membentuk alur kerja yang logis. Proses ini memungkinkan pengguna melakukan pengolahan data, pelatihan model, dan evaluasi dalam satu alur kerja yang terstruktur dan otomatis.



Gambar 4. 7 Tampilan Alur Proses Lengkap di RapidMiner (Decision Tree)

4.1.3 Hasil Pengujian

Setelah seluruh tahapan proses pengujian dilakukan menggunakan algoritma Decision Tree (C4.5) di RapidMiner, maka diperoleh hasil prediksi klasifikasi hasil panen kelapa sawit. Hasil tersebut dapat dilihat pada Gambar 4.8.

Row No.	Kategori Cu...	prediction(K...	confidence(...	confidence(...	confidence(...	Curah hujan...	Pupuk (kg)	Hasil Panen...	Kategori Ha...
1	Rendah	Rendah	1	0	0	262	59000	2000	Tinggi
2	Sedang	Sedang	0	1	0	375	50000	1900	Sedang
3	Rendah	Rendah	1	0	0	157	80000	2000	Tinggi
4	Sedang	Sedang	0	1	0	530	24000	1300	Rendah
5	Rendah	Rendah	1	0	0	325	42000	1800	Sedang
6	Tinggi	Tinggi	0	0	1	621	27000	1300	Rendah
7	Rendah	Rendah	1	0	0	312	49000	1800	Sedang
8	Tinggi	Tinggi	0	0	1	714	23000	1500	Rendah
9	Sedang	Sedang	0	1	0	380	47000	1700	Sedang
10	Rendah	Rendah	1	0	0	215	30000	2000	Tinggi
11	Rendah	Rendah	1	0	0	137	50000	2000	Tinggi
12	Sedang	Sedang	0	1	0	521	35000	1500	Rendah
13	Sedang	Sedang	0	1	0	421	15000	1600	Sedang
14	Rendah	Rendah	1	0	0	147	60000	2000	Tinggi

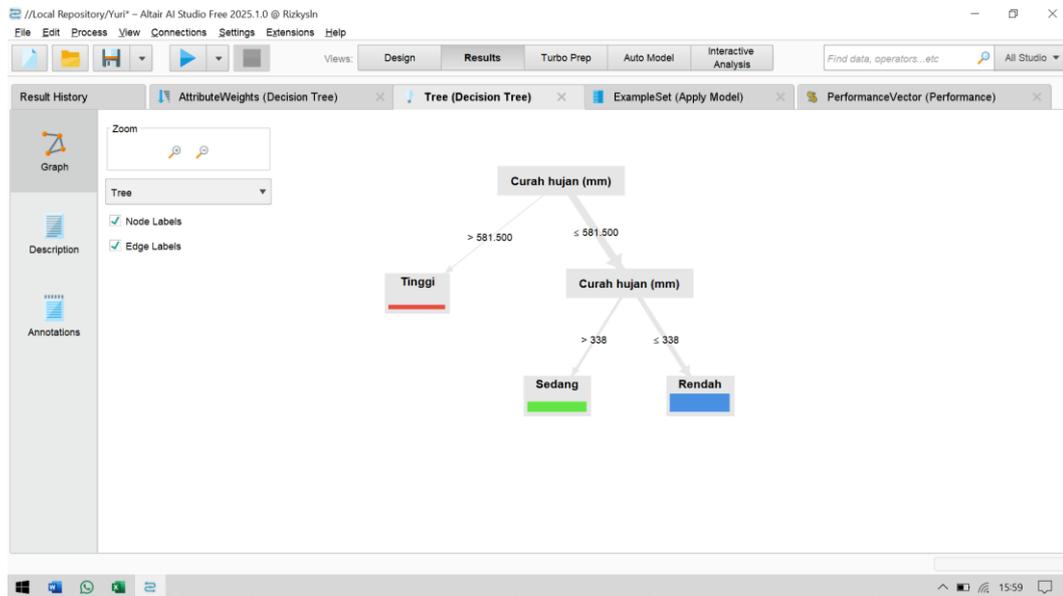
Gambar 4. 8 Hasil Prediksi dan Confidence Menggunakan Decision Tree

Gambar tersebut menampilkan data aktual hasil panen dalam kategori (Tinggi, Sedang, Rendah) yang dibandingkan dengan hasil prediksi oleh sistem. Selain itu juga disertai nilai confidence untuk masing-masing prediksi kelas, serta data pendukung seperti curah hujan dan jumlah pupuk yang digunakan.

Berdasarkan hasil prediksi:

1. Model berhasil memprediksi hasil panen dengan cukup akurat.
2. Mayoritas prediksi memiliki confidence tinggi terhadap kelas prediksinya (nilai 1).

Selanjutnya, struktur pohon keputusan yang dihasilkan oleh model dapat dilihat pada Gambar 4.9.



Gambar 4. 9 Visualisasi Pohon Keputusan (Decision Tree)

Pohon keputusan menunjukkan bahwa atribut “Curah Hujan (mm)” merupakan atribut paling berpengaruh terhadap klasifikasi hasil panen. Model membagi data berdasarkan nilai curah hujan:

1. Jika curah hujan ≤ 338 mm, maka hasil panen diklasifikasikan sebagai Rendah.
2. Jika curah hujan $> 581,5$ mm dan kategori curah hujan adalah Rendah, maka hasil panen diprediksi Tinggi.
3. Jika curah hujan > 338 mm dan kategori curah hujan adalah Sedang, maka hasil panen diprediksi Sedang.

Model ini secara logis membentuk aturan klasifikasi berbasis nilai dan kategori curah hujan, yang relevan terhadap hasil panen di lapangan.

4.2 Pengujian Metode Regresi Linear

Pengujian metode regresi linear dilakukan menggunakan RapidMiner dengan dua jenis data utama, yaitu data pelatihan (training) dan data pengujian

(testing) untuk memprediksi hasil panen kelapa sawit berdasarkan curah hujan dan jumlah pupuk.

4.2.1 Data Pelatihan (Training Data)

Data pelatihan merupakan data historis hasil panen yang telah lengkap, mencakup:

1. Curah hujan (mm)
2. Jumlah pupuk (kg)
3. Hasil panen aktual (ton)

Data ini digunakan untuk melatih model regresi linear agar mampu mempelajari hubungan antara variabel input dan output. Tampilan data pelatihan dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Dataset Pelatihan Regresi Linear

Curah Hujan (mm)	Pupuk (kg)	Hasil Panen (ton)
292	59000	2000
375	50000	1900
157	80000	2000
530	24000	1300
325	42000	1800
621	27000	1200
312	49000	1800
714	23000	1500
380	47000	1700
215	30000	2000
137	50000	2000
521	35000	1500
421	15000	1600
147	60000	2000
214	43000	2000
363	40000	1800
210	60000	1900
530	23000	1300
670	15000	1300
452	30000	1500
320	45000	1400
140	50000	2000
432	30000	1500

221	52000	2000
690	27000	1500
150	83000	2000
289	59000	1900
227	73000	1900
785	25000	1300
343	51000	1700
333	57000	1800
130	87000	2000
143	80000	2000
530	25000	1300
152	85000	2000
380	45000	1500
150	80000	2000
542	32000	1300
250	70000	1800
321	50000	1600
130	75000	2000
240	50000	1800

4.2.2 Data Pengujian (Testing Data)

Setelah model selesai dilatih, data pengujian dimasukkan untuk dilakukan proses prediksi. Data ini memiliki struktur yang sama dengan data pelatihan, namun kolom hasil panen dikosongkan agar model dapat memprediksi nilainya berdasarkan variabel curah hujan dan pupuk.

Tabel 4. 3 Dataset Prediksi (Kolom Hasil Panen Kosong)

Curah hujan (mm)	Pupuk (kg)	Hasil panen (ton)
300	60000	
450	40000	
210	75000	
600	30000	
120	85000	
360	50000	
255	65000	
700	28000	
170	82000	
390	47000	

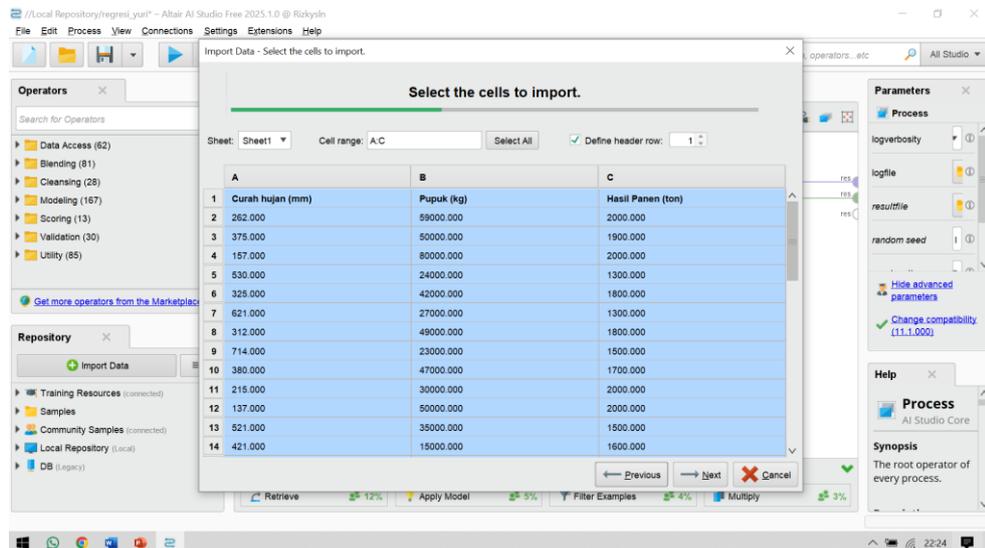
Setelah kedua data tersebut dimasukkan, dilakukan proses modelling dan evaluasi menggunakan operator Linear Regression. Hasil prediksi model dapat dilihat dan dibandingkan dengan data aktual, serta dievaluasi menggunakan metrik MAE, RMSE, dan R^2 .

4.2.3 Langkah Pengujian

Pengujian metode regresi linear dilakukan menggunakan aplikasi RapidMiner dengan beberapa tahapan sebagai berikut:

1. Import Data Training

Data historis hasil panen yang sudah lengkap (curah hujan, pupuk, hasil panen) diinput ke dalam RapidMiner sebagai *training data*. Data ini digunakan untuk melatih model agar sistem dapat mempelajari hubungan antar variabel.

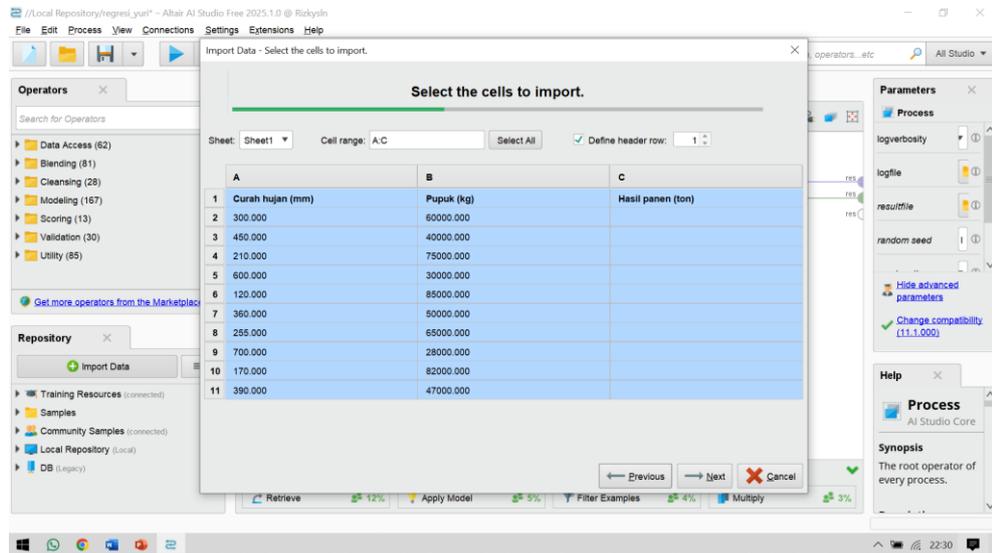


Gambar 4. 10 *Import Data Training*

2. Import Data Testing

Selanjutnya dilakukan input *testing data*, yaitu data dengan variabel curah hujan dan pupuk, tetapi kolom hasil panen dikosongkan. Data ini dipakai

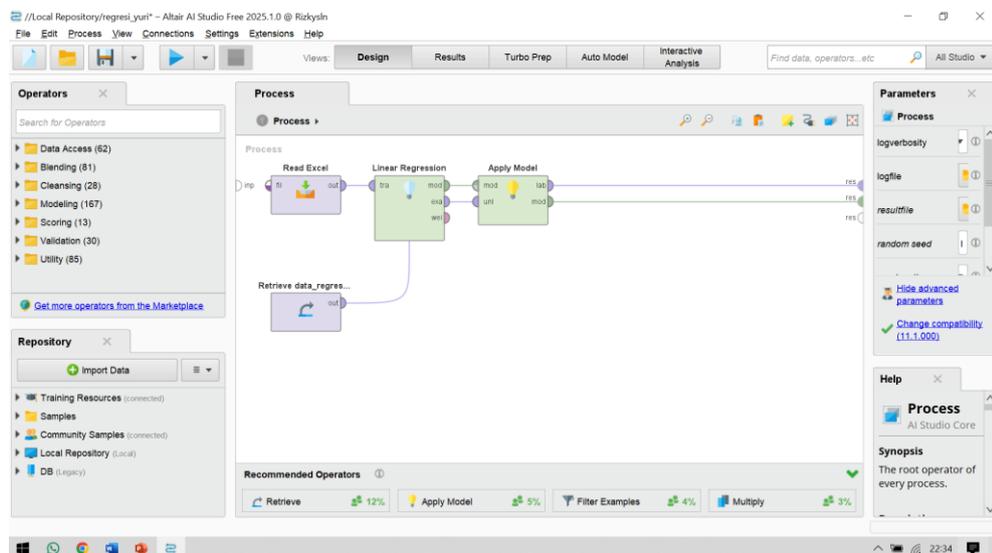
untuk menguji kemampuan model dalam memprediksi nilai hasil panen berdasarkan persamaan regresi yang terbentuk.



Gambar 4. 11 Import Data Testing

3. Perancangan Model (*Design Process*)

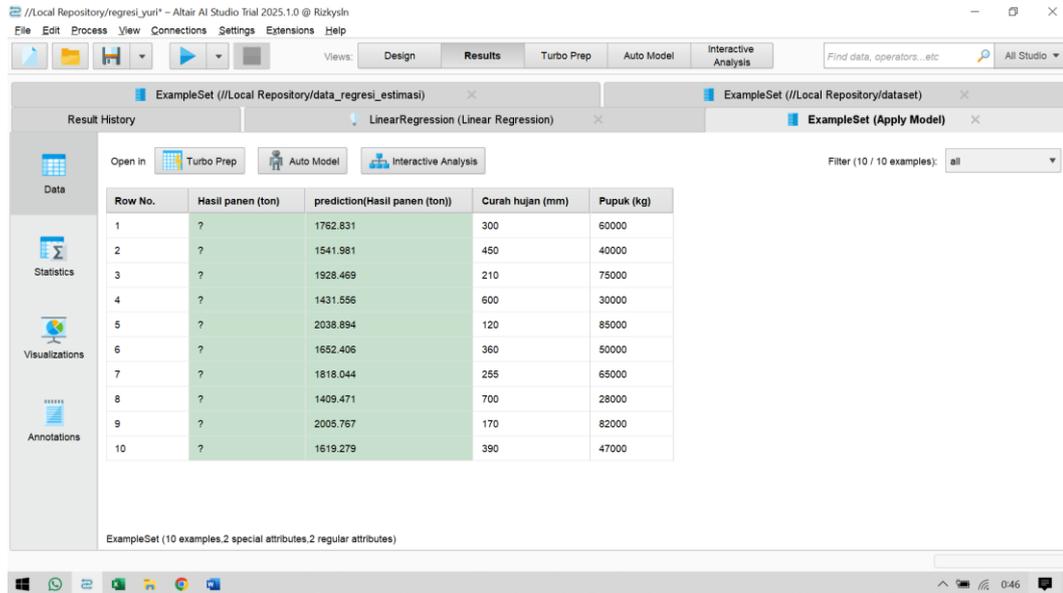
Pada tahap ini dilakukan penyusunan *workflow* di RapidMiner seperti pada Gambar 4.12 berikut ini.



Gambar 4. 12 Tampilan *Design* di RapidMiner (Regresi Linear)

4.2.4 Hasil Prediksi Regresi Linear

Setelah model dilatih menggunakan data historis, tahap selanjutnya adalah menguji kemampuannya untuk memprediksi hasil panen menggunakan data baru yang belum diketahui nilai hasil panennya. RapidMiner menghasilkan prediksi pada kolom “prediction (Hasil Panen)” seperti ditunjukkan pada Gambar 4.13.



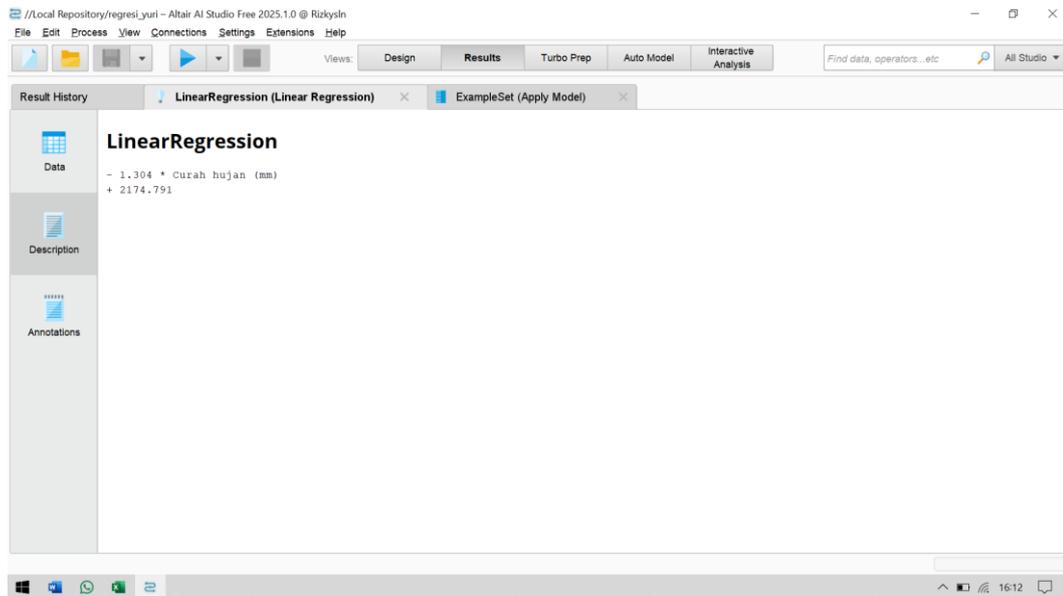
Row No.	Hasil panen (ton)	prediction(Hasil panen (ton))	Curah hujan (mm)	Pupuk (kg)
1	?	1762.831	300	60000
2	?	1541.981	450	40000
3	?	1928.469	210	75000
4	?	1431.556	600	30000
5	?	2038.894	120	85000
6	?	1652.406	360	50000
7	?	1818.044	255	65000
8	?	1409.471	700	28000
9	?	2005.767	170	82000
10	?	1619.279	390	47000

Gambar 4. 13 Hasil Prediksi Regresi Linear dengan Data Uji

Data pengujian ini hanya memiliki input berupa curah hujan dan jumlah pupuk, tanpa mencantumkan hasil panen aktual (disimbolkan dengan tanda “?”). Berdasarkan nilai-nilai tersebut, sistem berhasil memprediksi hasil panen untuk setiap baris data menggunakan model regresi linear yang telah dibentuk sebelumnya.

4.2.5 Persamaan Model Regresi

RapidMiner juga secara otomatis menghasilkan persamaan regresi yang menggambarkan hubungan matematis antara variabel input dan output. Persamaan regresi linear yang diperoleh dapat dilihat pada Gambar 4.14.



Gambar 4. 14 Persamaan Regresi Linear yang Dibentuk oleh Sistem

Persamaan yang dihasilkan adalah sebagai berikut:

$$\text{Hasil Panen} = - 1.304 \times \text{Curah Hujan (mm)} + 2174.791$$

Persamaan ini menunjukkan bahwa setiap kenaikan 1 mm curah hujan diperkirakan akan menurunkan hasil panen sebesar 1.304 (sesuai satuan hasil panen yang digunakan dalam data). Nilai konstanta (intersep) sebesar 2174.791 menunjukkan perkiraan hasil panen ketika curah hujan bernilai nol. Nilai konstanta yang relatif besar dapat dipengaruhi oleh skala dan satuan data yang digunakan, sehingga interpretasi perlu disesuaikan dengan konteks lapangan.

4.3 Perbandingan Hasil Prediksi

Setelah dilakukan pengujian terhadap kedua metode yaitu Decision Tree (C4.5) dan Regresi Linear, maka pada bagian ini akan dilakukan analisis perbandingan dari segi output hasil prediksi, bentuk hasil, serta evaluasi performa dari masing-masing model.

4.3.1 Perbandingan Jenis Output

Jenis output dari kedua metode yang digunakan dalam penelitian ini berbeda. Decision Tree menghasilkan output dalam bentuk kelas atau kategori (klasifikasi), sedangkan regresi linear menghasilkan output dalam bentuk angka atau nilai prediksi. Perbedaan ini dapat dilihat pada Tabel 4.4 berikut:

Tabel 4. 4 Perbandingan Jenis Output

Kriteria	Decision Tree (C4.5)	Regresi Linear
Jenis Output	Kategori (Rendah/Sedang/Tinggi)	Angka numerik (dalam ton)
Format Prediksi	Kelas (klasifikasi)	Nilai (prediksi kuantitatif)
Interpretasi	Mudah dipahami dan logis	Akurat secara numerik

Berdasarkan Tabel 4.4, dapat disimpulkan bahwa Regresi Linear lebih cocok digunakan apabila pengguna membutuhkan hasil dalam bentuk nilai kuantitatif. Sementara itu, C4.5 lebih mudah diinterpretasikan karena memberikan hasil klasifikasi yang langsung terbaca oleh pengguna awam.

4.3.2 Perbandingan Hasil Evaluasi

Selanjutnya dilakukan evaluasi performa terhadap kedua metode menggunakan beberapa metrik pengujian seperti MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), dan R-Squared (R^2). Nilai-nilai evaluasi ini ditampilkan dalam Tabel 4.5 berikut:

Tabel 4. 5 Perbandingan Hasil Evaluasi

Metrik Evaluasi	Decision Tree (C4.5)	Regresi Linear
MAE (Mean Absolute Error)	52.300	980.000
RMSE	57.680	3.380
R-Squared (R^2)	–	0.525

Dari hasil evaluasi tersebut:

1. Regresi Linear memiliki nilai RMSE dan MAE yang lebih kecil, menandakan bahwa model ini lebih akurat dalam memprediksi hasil panen dalam bentuk angka.
2. Namun, C4.5 lebih mudah diterapkan dalam sistem pakar atau klasifikasi cepat, karena bentuk keluarannya adalah kategori.

4.3.3 Kelebihan dan Kekurangan Masing-Masing Metode

Selain dari hasil evaluasi, penting juga untuk mengetahui kelebihan dan kekurangan masing-masing metode agar dapat menentukan metode yang paling sesuai dengan kebutuhan. Rangkuman kelebihan dan kekurangan metode C4.5 dan regresi linear dapat dilihat pada Tabel 4.6 berikut:

Tabel 4. 6 Kelebihan dan Kekurangan Metode

Aspek	C4.5	Regresi Linear
Kelebihan	Mudah dipahami (berbasis aturan), cocok untuk klasifikasi	Prediksi kuantitatif lebih presisi
Kekurangan	Tidak memberikan nilai numerik, hanya kelas	Kurang intuitif, lebih rumit jika outlier banyak

Tabel 4.6 menunjukkan bahwa C4.5 lebih mudah digunakan untuk interpretasi berbasis logika, sedangkan regresi linear lebih tepat untuk analisis prediksi berbasis angka. Oleh karena itu, pemilihan metode harus disesuaikan dengan kebutuhan pengguna akhir atau sistem yang akan dibangun.

4.3.4 Kesimpulan Perbandingan

Berdasarkan ketiga aspek yang telah dibahas, dapat disimpulkan bahwa:

1. Metode regresi linear lebih unggul dari segi akurasi prediksi kuantitatif.
2. Metode C4.5 lebih cocok digunakan untuk sistem klasifikasi berbasis kategori.

3. Pemilihan metode terbaik tergantung pada kebutuhan pengguna: jika ingin estimasi angka hasil panen → gunakan regresi; jika ingin klasifikasi sederhana → gunakan C4.5.