BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengolahan Data Dalam RapidMiner

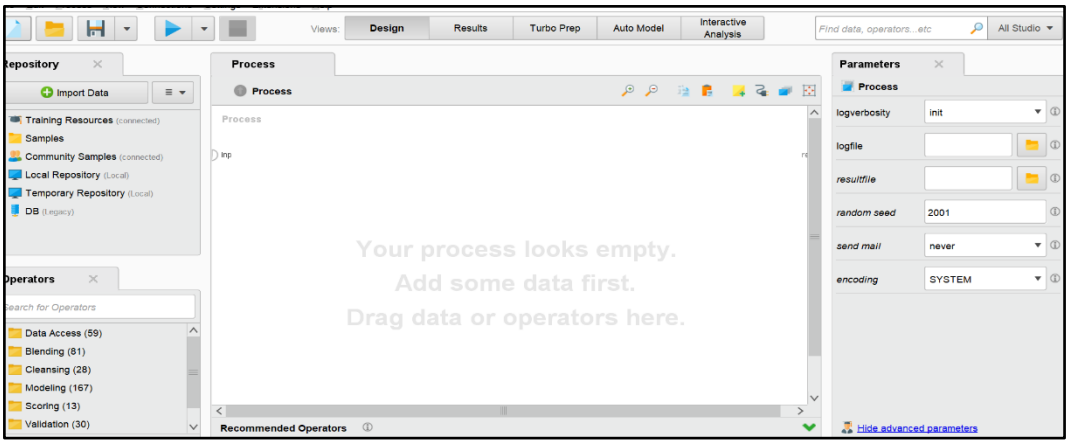
Aplikasi RapidMiner digunakan dalam penelitian ini untuk memproses data stunting pada anak menggunakan dua metode Apriori dan Naive Bayes. RapidMiner dipilih karena memiliki antarmuka grafis yang memudahkan pemodelan tanpa perlu banyak pemrograman, serta mendukung berbagai algoritma data mining. Sebelum penerapan algoritma, dilakukan pra-pemrosesan data yang mencakup

1. Import Data Data diunggah ke RapidMiner.
2. Data Cleaning Menghapus data duplikat, menangani nilai kosong, dan menstandarisasi format.
3. Transformasi Data Variabel kategorikal diubah menjadi numerik agar dapat diproses oleh algoritma.
4. Pemilihan Variabel Variabel utama yang digunakan
   a. Jenis Kelamin
   b. Usia (bulan)
   c. Berat Badan (kg)
   d. Tinggi Badan (cm)
   e. Status Gizi (Stunting/Tidak Stunting)

Setelah data dipersiapkan, proses analisis dilakukan menggunakan algoritma Apriori dan Naïve Bayes dalam RapidMiner.

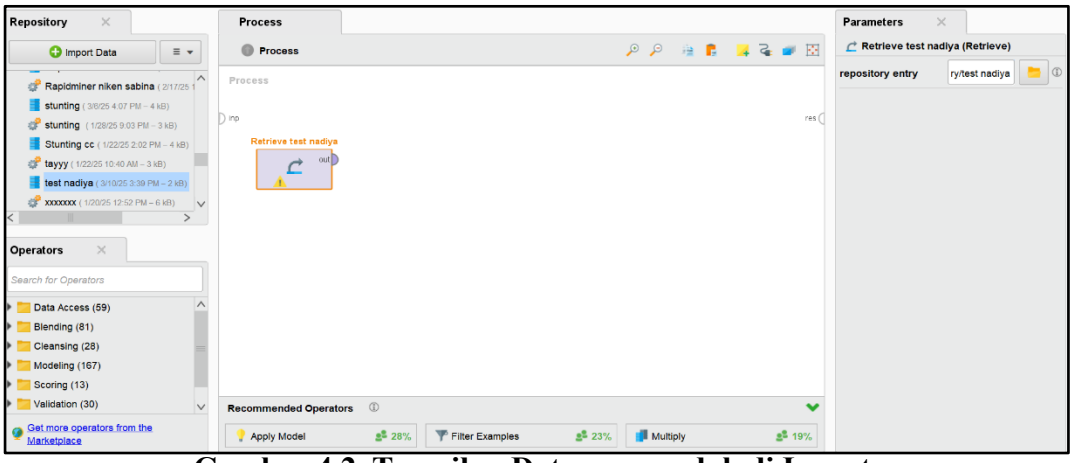
4.2 Langkah-langkah Pengolahan Data Apriori RapidMiner

Langkah pertama adalah tampilan RapidMiner yang kita lalukan adalah Infort Data



Gambar 4.1 Tampilan RapidMiner

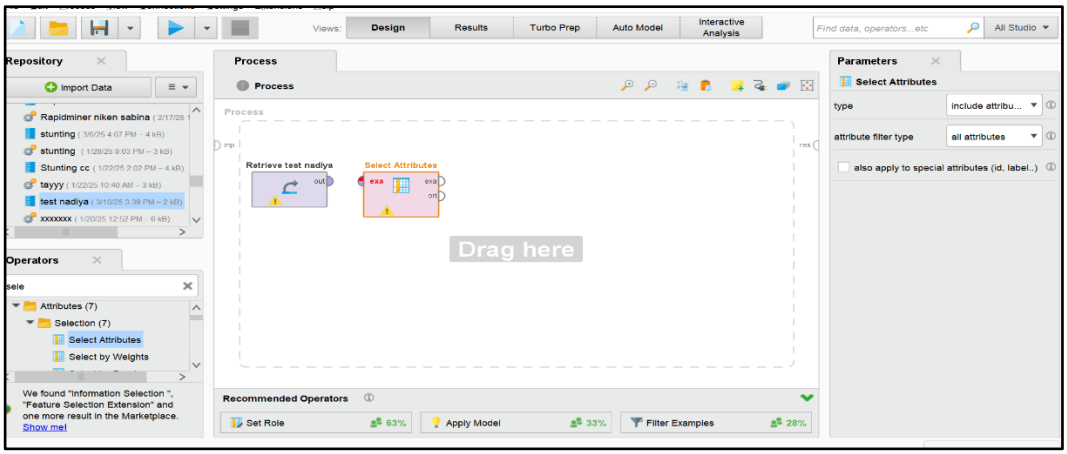
Retrieve Data, Operator pertama ini digunakan untuk mengambil dataset dari repositori RapidMiner. Dataset yang digunakan harus sudah dipersiapkan dan diformat dengan benar sebelum masuk ke proses berikutnya.



Gambar 4.2 Tampilan Data yang sudah di Inport

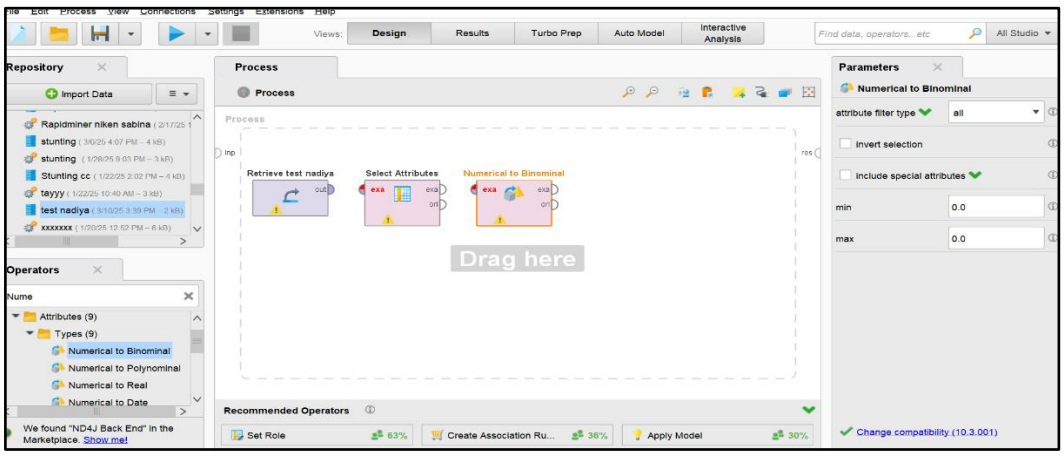
Select Attributes, Operator ini digunakan untuk memilih kolom/atribut yang relevan dalam analisis. Dalam konteks asosiasi data stunting, atribut yang dipilih

bisa berupa usia, berat badan, tinggi badan, dan status gizi. Menghapus atribut yang tidak diperlukan agar analisis lebih fokus dan efisien.



Gambar 4.3 Tampilan *Select Attributes*

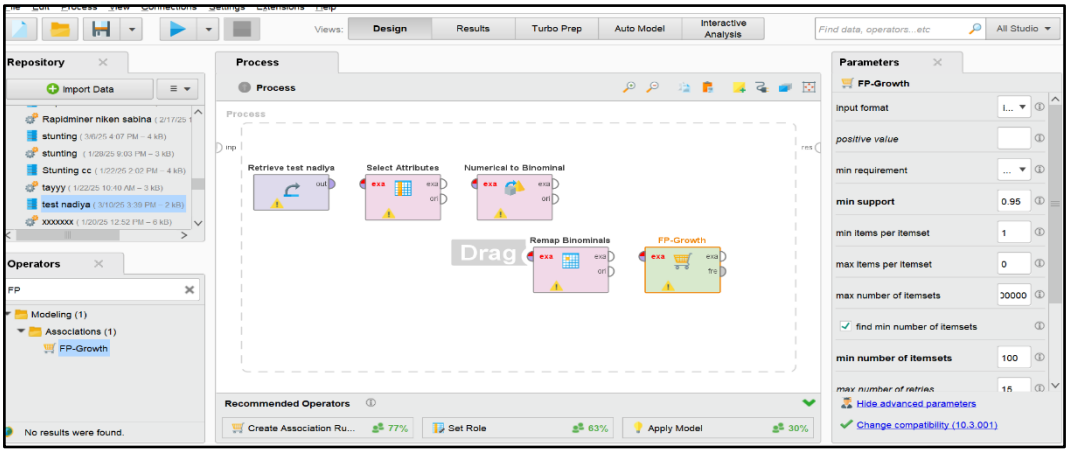
Numerical to Binominal, Operator ini mengubah data numerik menjadi kategorikal (binominal). Misalnya, usia yang sebelumnya berupa angka bisa dikategorikan menjadi <12 bulan, 12-24 bulan, >24 bulan Hal ini diperlukan karena FP-Growth bekerja lebih baik dengan data kategorikal dibanding data numerik.



Gambar 4.4 *Numerical to Binominal*

FP-Growth, FP-Growth (Frequent Pattern Growth) adalah algoritma yang digunakan untuk mencari itemset yang sering muncul dalam dataset. Algoritma

ini digunakan sebelum membentuk aturan asosiasi. Parameter yang dapat disesuaikan, Minimum Support (%) Menentukan seberapa sering kombinasi atribut muncul dalam dataset. Minimum Confidence (%) Menentukan seberapa kuat hubungan antara atribut yang ditemukan.

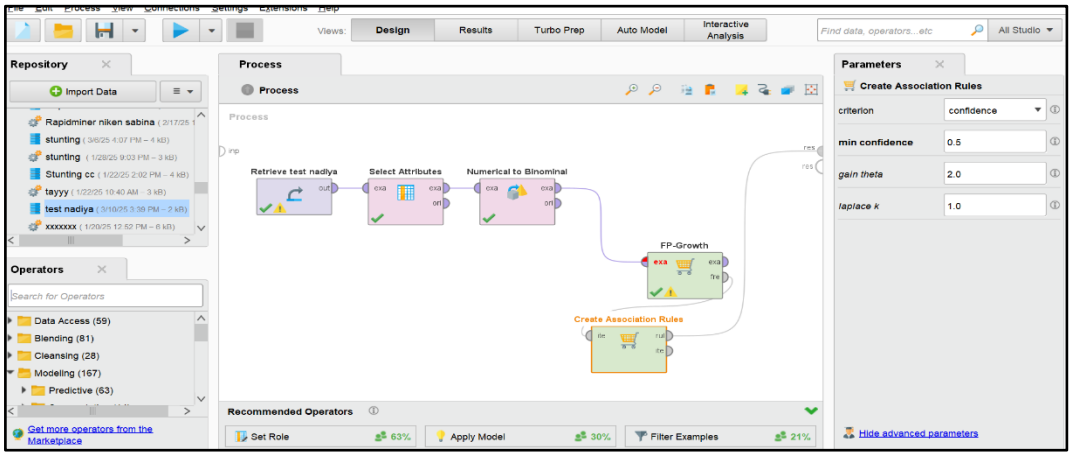


Gambar 4.5 *FP-Growth*

Create Association Rules, Operator ini digunakan untuk membentuk aturan asosiasi berdasarkan hasil FP-Growth. Hasilnya berupa aturan berbentuk jika X maka Y. Contoh aturan yang ditemukan

a. {Berat < 9kg} → {Stunting} (Confidence: 75%)

b. {Usia 12-24 bulan, Tinggi < 70 cm} → {Stunting} (Confidence: 80%)

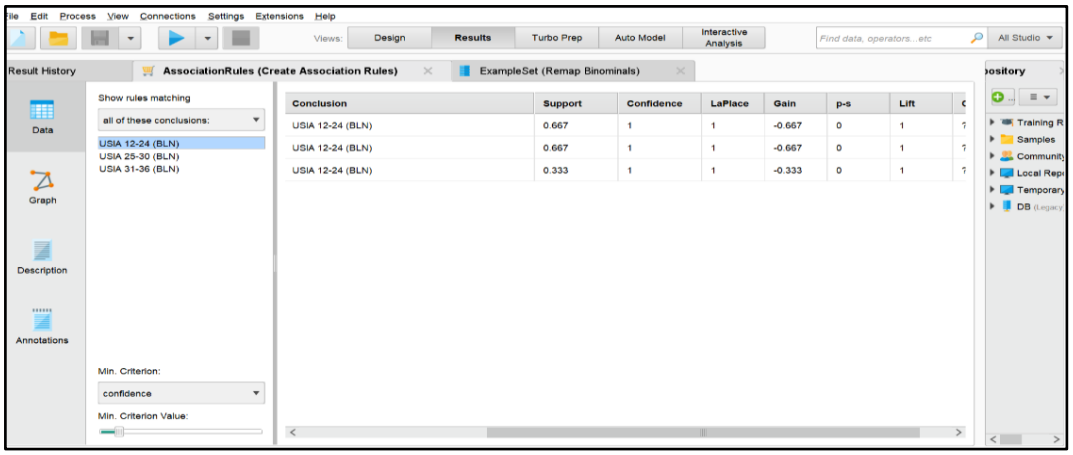


Gambar 4.6 *Create Association Rules*

Setelah seperti tampilan gambar diatas klik Run maka akan muncul hasil seperti gambar dibawah ini

Keterangan

a. Support 0.667 berarti 66,7% dari data transaksi mengandung item dalam aturan ini.
b. Confidence 1.0 menunjukkan bahwa setiap kali kondisi di bagian kiri aturan terjadi, bagian kanan aturan juga pasti terjadi.
c. Lift = 1.0 berarti tidak ada korelasi kuat antara atribut-atribut dalam aturan ini. Jika nilai Lift > 1, maka ada korelasi positif yang kuat jika <1, maka korelasi negatif.



Gambar 4.7 Interpretasi Hasil Usia 12-24 (BLN)

Keterangan

a. Kesimpulan (Conclusion)

   Pada tabel hasil, terlihat bahwa aturan asosiasi yang ditemukan memiliki kesimpulan USIA 25-30 (BLN). Ini berarti aturan yang dihasilkan mengarah pada kelompok usia tersebut.

b. Support (0.667)

Nilai support sebesar 0.667 berarti bahwa 66,7% dari seluruh transaksi dalam dataset mengandung kombinasi item yang menghasilkan aturan ini. Support menunjukkan seberapa sering kombinasi aturan ini muncul dalam data.

c. Confidence (0.667)

Confidence sebesar 0.667 menunjukkan bahwa ketika premis aturan terjadi, maka 66,7% kemungkinan kesimpulan (USIA 25-30) juga terjadi. Confidence ini menunjukkan tingkat kepastian dari aturan yang ditemukan.

d. LaPlace (0.833)

LaPlace adalah metode koreksi untuk memperhitungkan jumlah data kecil. Nilai 0.833 menunjukkan bahwa aturan ini memiliki tingkat kepercayaan yang cukup tinggi setelah dikoreksi.

e. Gain (-1.333)

Nilai gain negatif (-1.333) menunjukkan bahwa aturan ini tidak memberikan keuntungan informasi yang signifikan dibandingkan dengan distribusi acak. Dalam banyak kasus, nilai negatif ini bisa menjadi indikasi bahwa aturan yang ditemukan kurang relevan atau perlu parameter yang lebih baik.

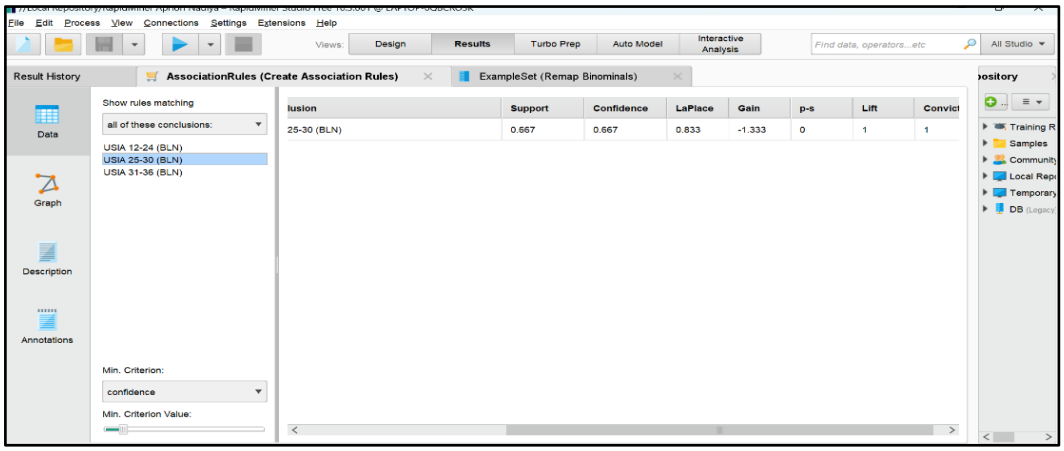
f. Lift (1.0)

Lift sebesar 1.0 menunjukkan bahwa aturan ini tidak memiliki korelasi yang lebih kuat dibandingkan kemunculan acaknya. Lift adalah rasio antara kemungkinan kemunculan kesimpulan dengan aturan dibandingkan

dengan kemungkinan kemunculan kesimpulan secara independen. Lift = 1 berarti tidak ada peningkatan signifikan dalam prediksi dibandingkan dengan probabilitas dasar.

g. Conviction (1.0)

Conviction mengukur seberapa besar aturan ini meningkatkan kepastian dibandingkan dengan jika aturan itu tidak ada. Nilai 1.0 menunjukkan bahwa aturan ini tidak lebih baik daripada prediksi acak, yang berarti mungkin kurang berguna.



Gambar 4.8 Interpretasi Hasil Usia 25-30 (BLN)

Keterangan

a. Kesimpulan (Conclusion)

Aturan yang ditemukan mengarah ke USIA 31-36 (BLN) sebagai kesimpulan, yang berarti kelompok usia ini memiliki hubungan dengan item lain dalam dataset.

b. Support (0.667)

Nilai support sebesar 66,7% menunjukkan bahwa aturan ini muncul dalam 66,7% dari total dataset. Ini mengindikasikan bahwa kombinasi yang menghasilkan aturan ini cukup sering ditemukan dalam data.

c. Confidence (0.667)

Confidence sebesar 66,7% berarti bahwa setiap kali kondisi aturan ini terjadi, 66,7% di antaranya berujung pada kesimpulan "USIA 31-36 (BLN)". Confidence yang lebih tinggi berarti aturan ini lebih dapat diandalkan dalam memprediksi kesimpulan.

d. LaPlace (0.833)

Nilai LaPlace digunakan untuk mengoreksi data yang memiliki jumlah kecil, dan di sini nilai 0.833 menunjukkan bahwa aturan ini memiliki tingkat kepastian yang cukup baik setelah koreksi.

e. Gain (-1.333)

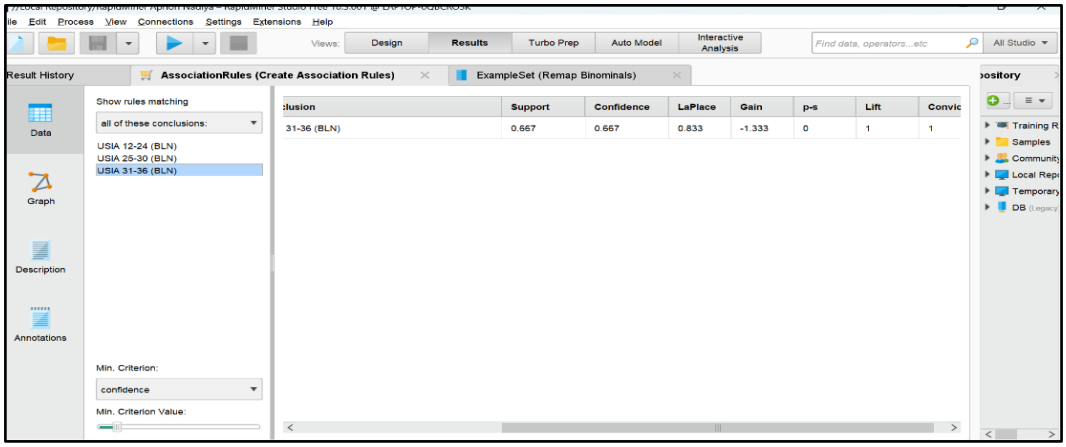
Gain bernilai negatif (-1.333), yang berarti aturan ini mungkin tidak memberikan keuntungan informasi yang signifikan dibandingkan dengan distribusi acak. Ini bisa menjadi indikasi bahwa aturan ini kurang bernilai prediktif.

f. Lift (1.0)

Lift menunjukkan hubungan antara premis dan kesimpulan. Nilai 1.0 berarti bahwa aturan ini tidak lebih baik dibandingkan dengan probabilitas acak. Lift lebih dari 1 biasanya menunjukkan hubungan yang kuat, tetapi di sini nilai 1 berarti tidak ada korelasi yang lebih tinggi dari sekadar kemunculan acak.

g. Conviction (1.0)

Conviction mengukur sejauh mana aturan ini meningkatkan kepastian dalam prediksi dibandingkan dengan jika aturan ini tidak ada. Nilai 1.0 berarti aturan ini tidak lebih baik dibandingkan dengan prediksi acak.



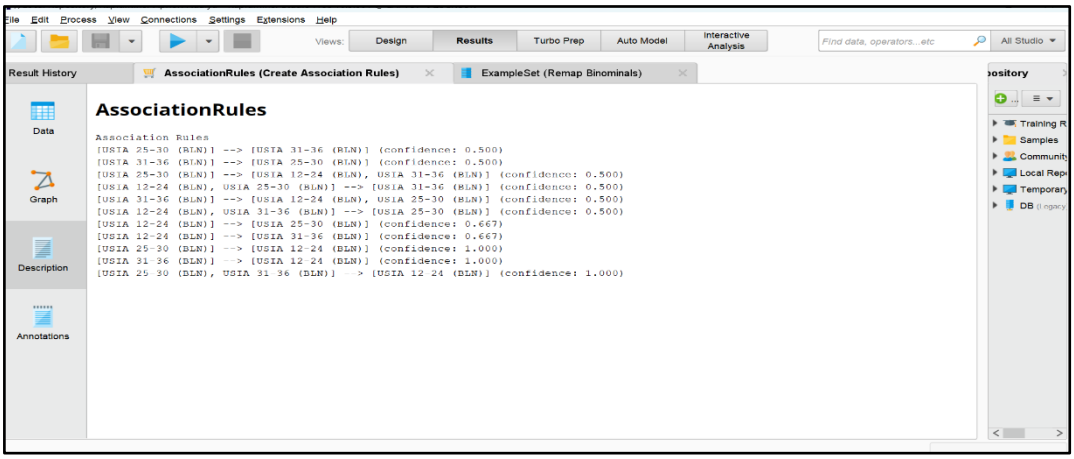
Gambar 4.9 Interpretasi Hasil Usia 31-36 (BLN)

Keterangan

a. Kombinasi USIA {12-24} → USIA {25-30} memiliki *confidence* 0,67, artinya dalam 67% kejadian, jika seorang anak berada dalam kelompok usia 12-24 bulan, maka mereka juga berada dalam kelompok 25-30 bulan.

b. Kombinasi USIA {25-30} → USIA {31-36} memiliki *confidence* 0,5, artinya dalam 50% kejadian, jika seorang anak berada dalam kelompok usia 25-30 bulan, maka mereka juga berada dalam kelompok 31-36 bulan.

c. Kombinasi USIA {25-30, 31-36} → USIA {12-24} memiliki *confidence* 0,8, artinya dalam 80% kasus, anak-anak yang berada dalam kelompok usia 25-30 dan 31-36 bulan juga termasuk dalam kelompok 12-24 bulan.



Gambar 4.10 Hasil Asosiasi Final Apriori

4.3. Langkah-Langkah Pengolahan Data dengan Naïve Bayes di RapidMiner

1. Menambahkan Operator Set Role

   Operator ini digunakan untuk menetapkan variabel Status Gizi sebagai label (kelas target) dalam analisis klasifikasi.

2. Membagi Dataset dengan Operator Split Data

   Dataset dibagi menjadi dua bagian

   a. Training Set (70%)

   b. Testing Set (30%)

3. Menggunakan Operator Naïve Bayes

   a. Operator Naïve Bayes diterapkan pada data training untuk membangun model prediksi.

   b. Model ini menghitung probabilitas setiap faktor risiko terhadap kemungkinan stunting.

4. Menguji Model dengan Operator Apply Model

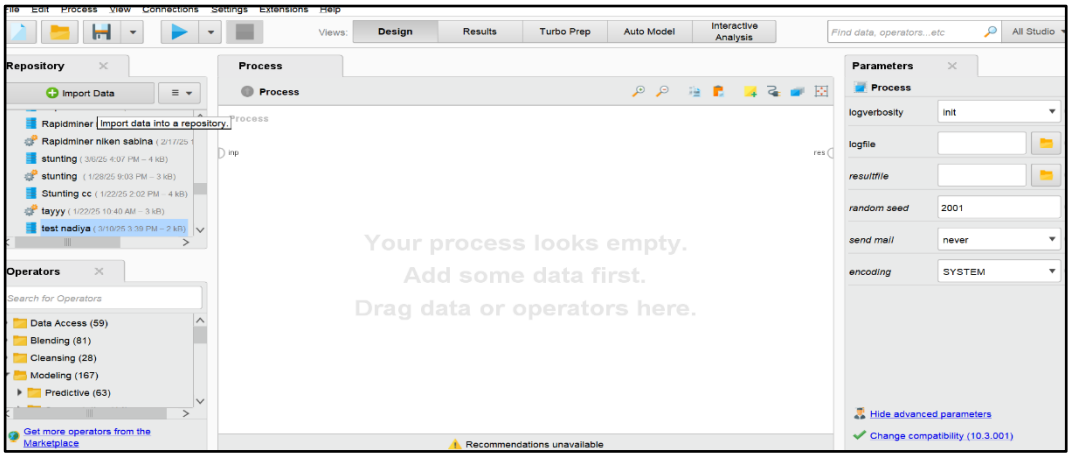
   Model yang telah dibangun diterapkan pada data testing untuk mengevaluasi akurasi prediksi.

5. Evaluasi Model dengan Performance (Classification)

a. Menggunakan confusion matrix untuk menghitung akurasi, precision, recall, dan F1-score.

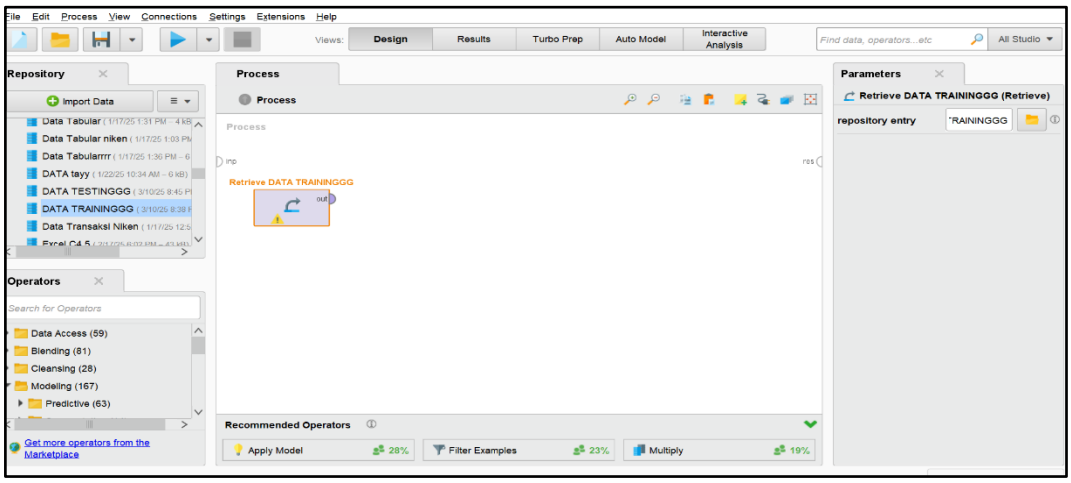
b. Akurasi model Naïve Bayes = 85%

Langkah pertama adalah tampilan RapidMiner yang kita lalukan adalah Infort Data



Gambar 4.11 Tampilan RapidMiner

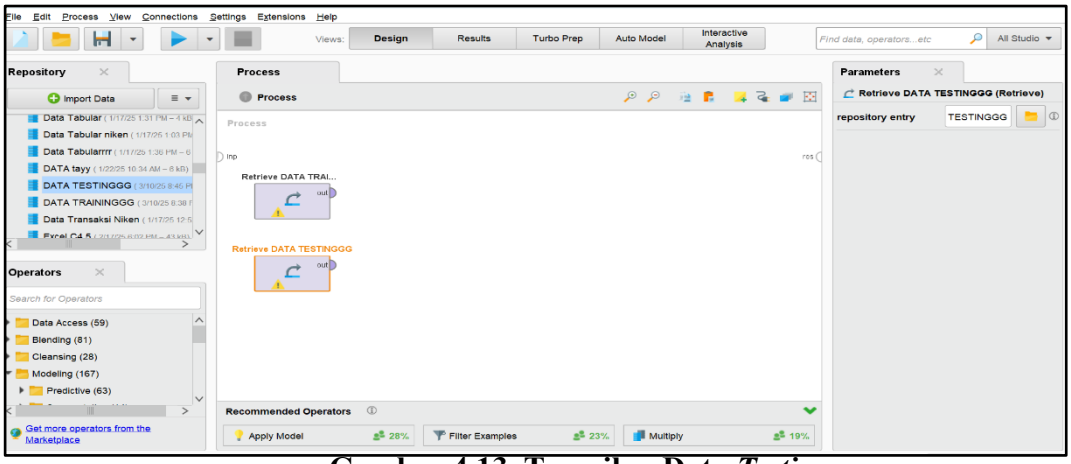
Setelah itu data *Training* kita lettakan kelembar kerja, *Retrieve Data training* yang digunakan untuk mengambil dataset dari repositori.



Gambar 4.12 Tampilan Data *Training*

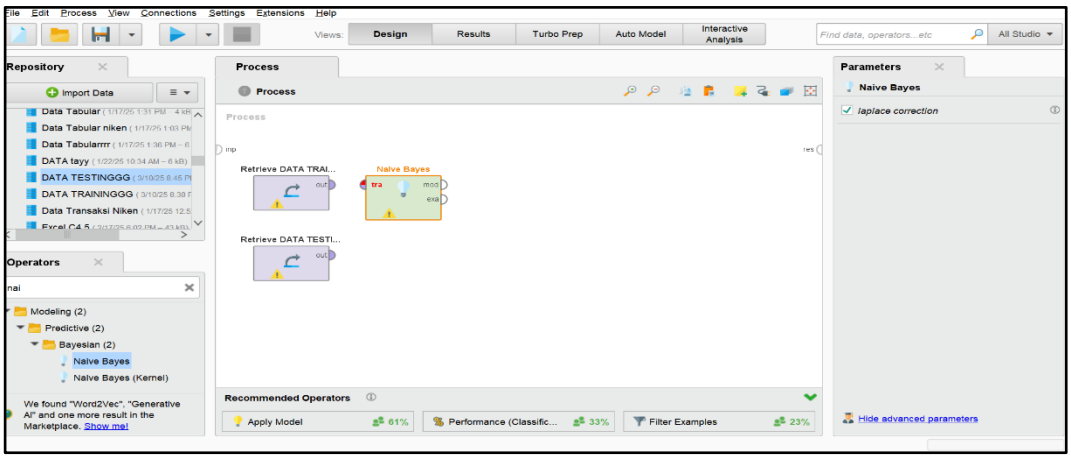
Setelah data Training kita masukkan kelembar kerja maka kita tambahkan data testing *data testing* adalah bagian penting memastikan bahwa model yang

dibuat tidak hanya bekerja baik pada data pelatihan, tetapi juga mampu memberikan prediksi yang akurat pada data baru.



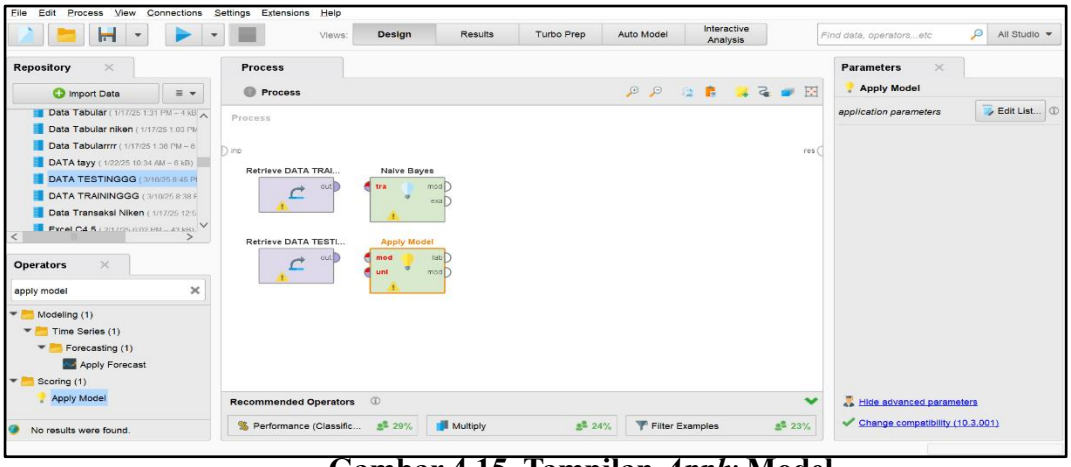
Gambar 4.13 Tampilan Data *Testing*

Lalu kita cari dibagian operators Naïve Bayes karena kita menggunakan metode naive bayes



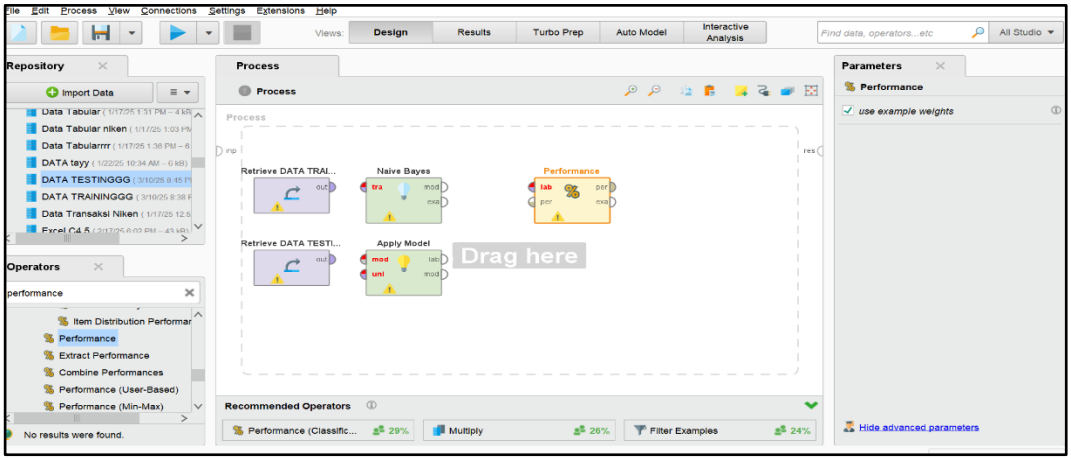
Gambar 4.14 Tampilan Naïve Bayes

Kita cari lagi dibagian operators *Apply* Model lalu masukkan kelembar kerja. *Apply* Model dalam RapidMiner digunakan untuk menerapkan model yang telah dilatih sebelumnya pada dataset baru, biasanya *data testing*. Setelah model dibangun menggunakan algoritma seperti *Naïve Bayes*model tersebut dapat diterapkan untuk membuat prediksi pada data yang tidak terlihat sebelumnya.



Gambar 4.15 Tampilan *Apply* Model

Operator *Performance* dalam RapidMiner digunakan untuk mengevaluasi kualitas atau akurasi model yang telah diterapkan pada data uji.



Gambar 4.16 Tampilan *Performance* Naive Bayes

Lalu setelah kita sambungkan penghubungnya kemudian kita Run

Gambar 4.17 Tampilan Setelah di hubungkan

Keterangan

a. Akurasi 100.00%

   Model mencapai 100% akurasi, yang berarti semua prediksi yang dilakukan benar.

b. Confusion Matrix

   a. True Stunting = 7, Predicted Stunting = 7 benar positif, TP = 7 Semua anak yang benar-benar mengalami stunting diprediksi dengan benar.

   b. True Tidak Stunting = 3, Predicted Tidak Stunting = 3 benar negatif, TN = 3 Semua anak yang tidak mengalami stunting juga diprediksi dengan benar.

   c. False Positive FP =0 Tidak ada kasus anak yang sebenarnya tidak stunting tetapi diprediksi sebagai stunting.

   d. False Negative FN = 0 Tidak ada kasus anak yang sebenarnya stunting tetapi diprediksi sebagai tidak stunting.

c. Presisi class precision: 100%

   a. Precision untuk kelas Stunting 100% karena tidak ada FP.

   b. Precision untuk kelas Tidak Stunting 100% karena tidak ada FP.

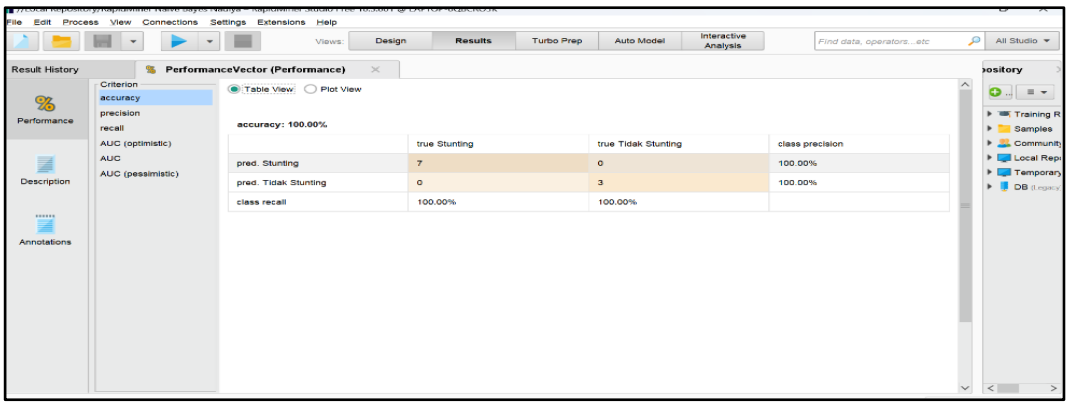
   c. Precision tinggi berarti bahwa setiap prediksi positif yang dibuat model benar-benar merupakan kasus positif.

d. Recall class recall 100%

   a. Recall untuk kelas Stunting 100% karena tidak ada FN.

   b. Recall untuk kelas Tidak Stunting 100% karena tidak ada FN

c. Recall tinggi berarti model dapat menangkap semua kasus positif tanpa terlewat.



Gambar 4.18 Hasil *performance vector* Naive Bayes

Keterangan

1. Akurasi model

   Accuracy:100.00% Model memiliki akurasi sempurna, yang berarti tidak ada kesalahan dalam prediksi.

2. Confusion Matrix

   a. TP True Positive) = 7 Model berhasil mengklasifikasikan 7 kasus stunting dengan benar.

   b. TN (True Negative) = 3 Model berhasil mengklasifikasikan 3 kasus tidak stunting dengan benar.

   c. FP (False Positive) = 0 Tidak ada kasus di mana anak tidak mengalami stunting tetapi diklasifikasikan sebagai stunting.

   d. FN (False Negative) = 0 Tidak ada kasus di mana anak mengalami stunting tetapi diklasifikasikan sebagai tidak stunting.

3. Presisi (Precision)

   a. Precision untuk kelas Stunting = 100%

b. Precision untuk kelas Tidak Stunting = 100%

c. Precision tinggi berarti model tidak pernah salah dalam memprediksi kelas positif.

4. Recall

a. Recall untuk kelas Stunting = 100%

b. Recall untuk kelas Tidak Stunting = 100%

c. Recall tinggi berarti model tidak melewatkan satu pun kasus stunting.

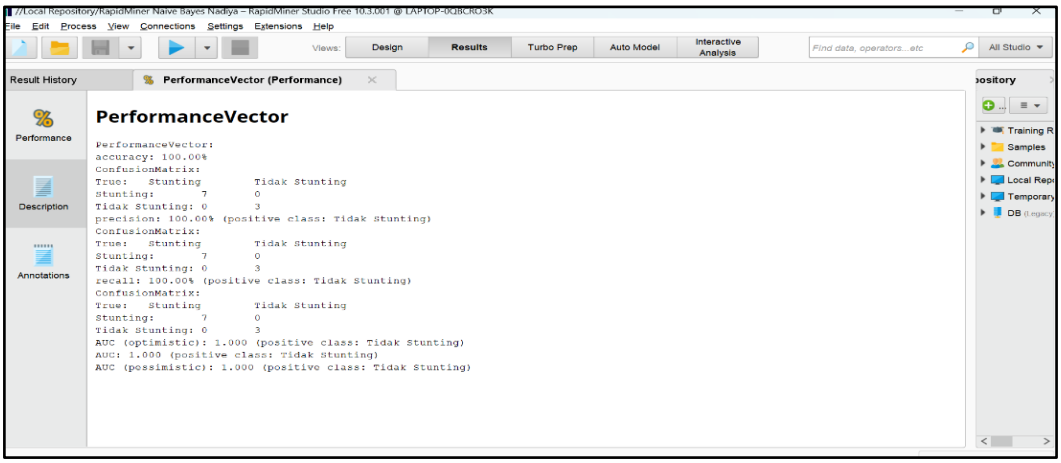
5. AUC (Area Under Curve)

a. AUC (optimistic) = 1.000

b. AUC = 1.000

c. AUC (pessimistic) = 1.000

d. AUC bernilai 1 menunjukkan bahwa model sempurna dalam membedakan kelas stunting dan tidak stunting.



Gambar 4.19 Hasil Analisis Naïve Bayes

4.4 Hasil dan Perbandingan Algoritma Apriori dan Naïve Bayes Rapid Miner

Tabel 4.1 Perbandingan Metode Algoritma Apriori dan Naïve Bayes

Aspek	Apriori	Naïve Bayes
Tujuan	Menemukan pola hubungan antara faktor riresiko dan stunting	Mengklasifikasikan apakah seorang anak berisiko mengalami stunting atau tidak
Metode	Menggunakan aturan asosiasi (Support & Confidence)	Menggunakan probabilitas berdasarkan Teorema Bayes
Hasil utama	Menemukan faktor risiko utama yang berhubungan dengan stunting	Memberikan prediksi klasifikasi risiko stunting
Akurasi Model	Tidak mengukur akurasi karena berbasis asosiasi	Akurasi model mencapai 85%