

BAB IV

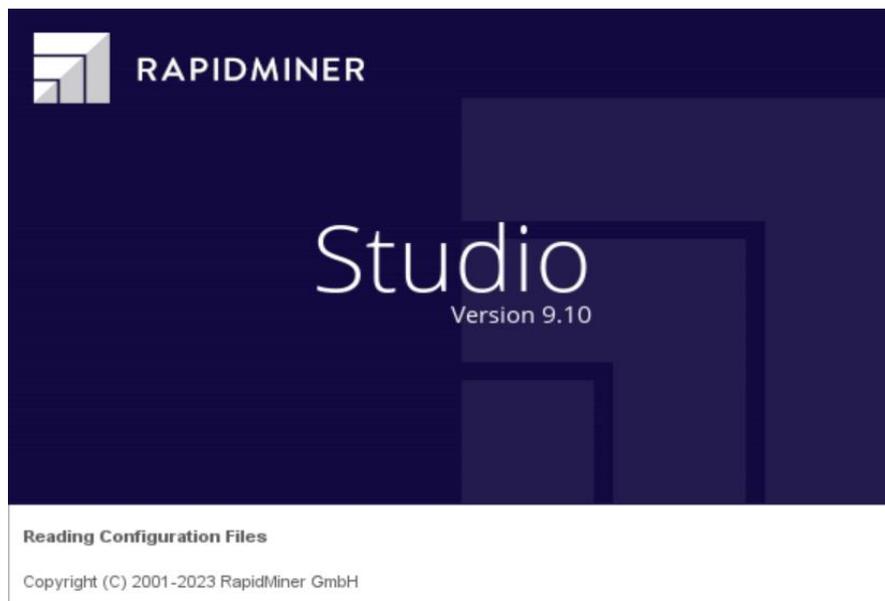
IMPLEMENTASI DAN PEMBAHASAN

3.1 Tahapan Implementasi dengan RapidMiner

Implementasi sistem prediksi dilakukan menggunakan platform RapidMiner Studio. Data kedisiplinan karyawan berdasarkan kedisiplinan PT. PP London Sumatera Indonesia Tbk. Sei Rumbiya Estate akan diimplementasikan ke dalam aplikasi RapidMiner 9.10 dengan menerapkan algoritma *Naïve Bayes*. Proses implementasi dilakukan dengan menjadikan variabel Kriteria sebagai atribut Label (target) untuk prediksi, dengan atribut-atribut lainnya sebagai atribut independen.

3.1.1 Import Data

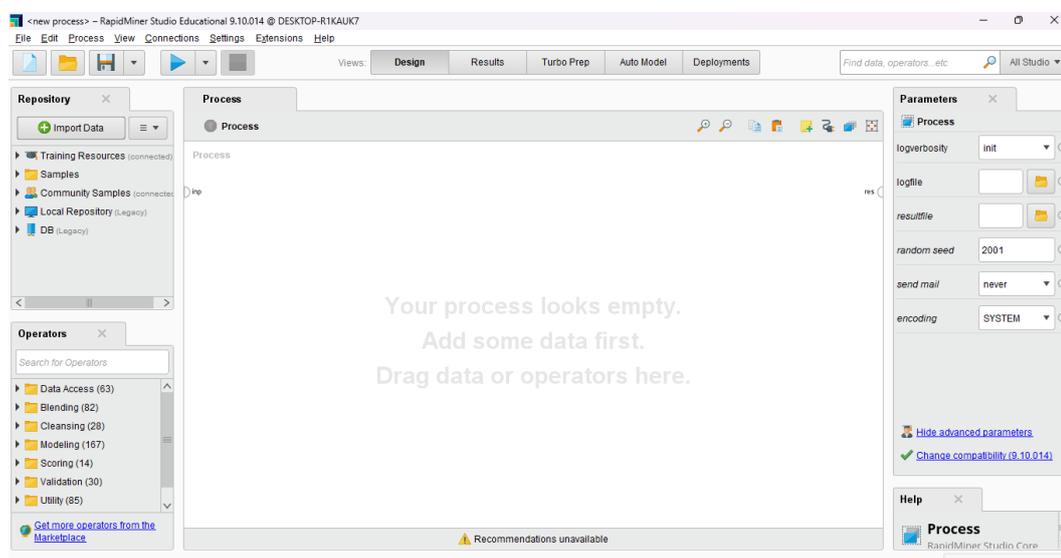
Tahap import data adalah proses memasukkan data dari sumber eksternal ke dalam lingkungan perangkat lunak atau sistem yang akan digunakan untuk analisis atau pemrosesan lebih lanjut.



Gambar 4. 1 Proses Membuka Aplikasi RapidMiner

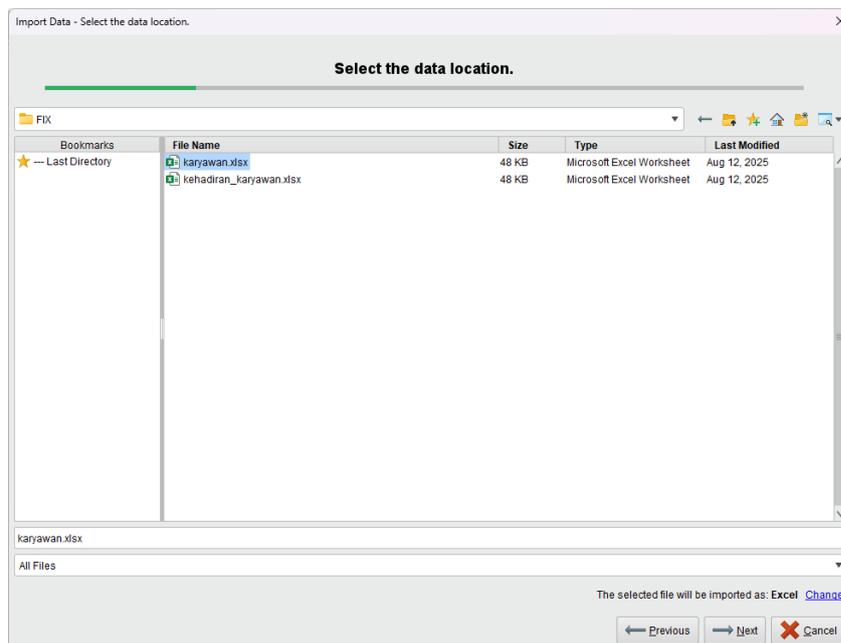
Gambar 4.1 menunjukkan tampilan awal (*splash screen*) saat membuka aplikasi RapidMiner Studio versi 9.10. Berikut penjelasan detailnya:

- Di bagian kiri atas terdapat logo RapidMiner yang merupakan identitas visual dari perangkat lunak ini.
- Terdapat tulisan besar "Studio" yang menandakan nama produk, yaitu RapidMiner Studio.
- Tertera versi aplikasi yaitu Version 9.10, menunjukkan bahwa pengguna menggunakan versi terbaru atau spesifik dari RapidMiner Studio.
- Tulisan "Reading Configuration Files" menandakan bahwa aplikasi sedang memuat file konfigurasi yang diperlukan untuk menjalankan program dengan benar.



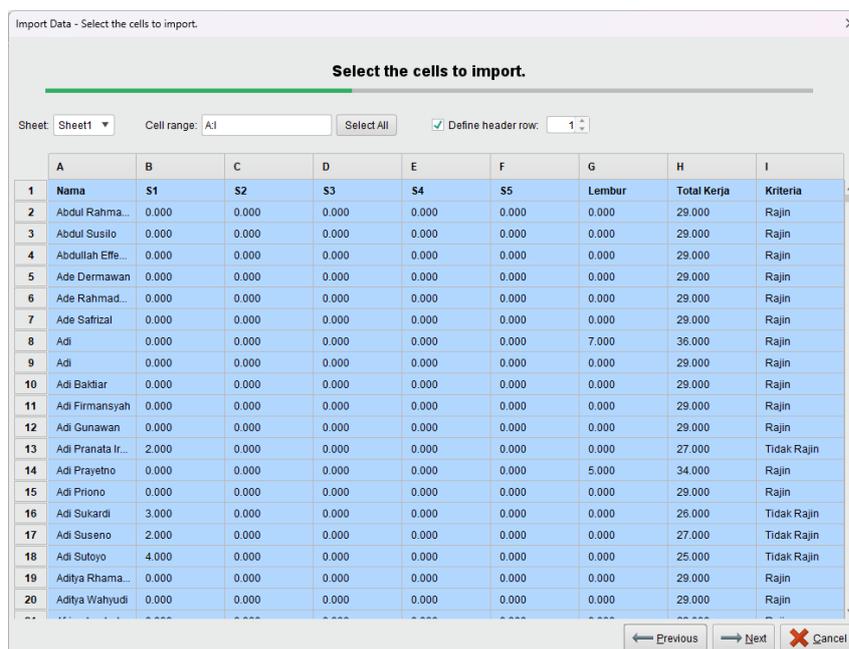
Gambar 4. 2 Lembar Kerja Utama Aplikasi RapidMiner

Gambar 4.2 menampilkan tampilan utama (main workspace) dari aplikasi RapidMiner Studio setelah aplikasi berhasil dibuka dan siap digunakan untuk melakukan proses analisis data.



Gambar 4.3 Direktori File Import Data

Gambar 4.3 menunjukkan tampilan jendela “Import Data - Select the data location” pada RapidMiner Studio. Tampilan ini muncul saat pengguna melakukan proses awal untuk mengimpor dataset ke dalam RapidMiner, khususnya dataset yang akan digunakan sebagai input untuk proses machine learning. Dalam gambar ini, terlihat bahwa sedang memilih file karyawan.xlsx, yang merupakan file Excel berisi data kedisiplinan karyawan PT. PP London Sumatera Indonesia Tbk. Sei Rumbiya Estate yang digunakan untuk membangun model prediksi kedisiplinan karyawan. File ini tersimpan dalam folder lokal yang ditandai dengan label "Last Directory" di bagian kiri tampilan. File karyawan.xlsx berukuran 48 KB, menunjukkan bahwa data yang dimuat cukup ringan dan tidak terlalu besar.



Gambar 4. 4 Select the Cells to Import

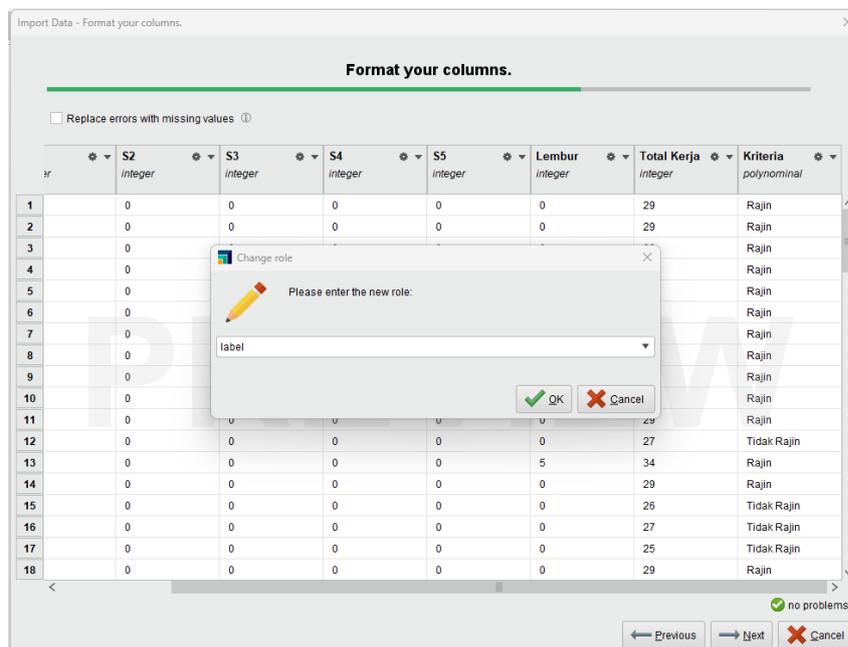
Gambar 4.4 menampilkan tahap kedua dari proses pengimporan data ke dalam RapidMiner Studio, yaitu langkah untuk memilih range data dari worksheet yang akan dimasukkan ke dalam proses analisis. Pada tampilan ini, pengguna dihadapkan pada jendela antarmuka dengan judul “Select the cells to import”, yang menunjukkan isi tabel dari Sheet1 file Excel karyawan.xlsx. Proses ini memungkinkan pengguna untuk memverifikasi struktur dataset sebelum melanjutkan ke tahap preprocessing.

Tabel 4. 1 Atribut Data

No	Atribut	Deskripsi
1	S1	Jumlah hari karyawan tidak hadir karena sakit biasa. Ini mencakup absensi yang tidak memerlukan perawatan lebih lanjut.

2	S2	Jumlah hari karyawan tidak hadir karena dirujuk ke fasilitas kesehatan lain. Ini menunjukkan bahwa karyawan memerlukan perawatan lebih lanjut.
3	S3	Jumlah hari karyawan tidak hadir karena dirawat di rumah sakit. Ini menunjukkan tingkat keparahan kondisi kesehatan.
4	S4	Jumlah hari karyawan tidak hadir karena sakit yang memerlukan perawatan di rumah sakit selama 4 bulan. Ini menunjukkan masalah kesehatan yang serius dan berkepanjangan.
5	S5	Jumlah hari karyawan tidak hadir karena sakit yang memerlukan perawatan di rumah sakit selama 5 bulan. Ini juga menunjukkan masalah kesehatan yang sangat serius dan dapat mempengaruhi produktivitas jangka panjang.
6	Lembur	Total jam lembur yang dilakukan oleh karyawan. Lembur dapat menjadi indikator komitmen dan beban kerja.
7	Total Kerja	$(\text{Jumlah Hari Kerja yaitu } 29 \text{ hari} + \text{Lembur}) - (S1+S2+S3+S4+S5)$.
8	Kriteria	Karyawan dikategorikan sebagai " Disiplin " jika jumlah total kerja mereka dalam sebulan mencapai atau melebihi 29 hari. Karyawan dikategorikan sebagai "Tidak Disiplin " jika jumlah total kerja mereka kurang dari 29 hari.

Tabel 4.1 yang ditampilkan terdiri dari 8 atribut (kolom) dan beberapa baris data kedisiplinan karyawan. Setiap kolom mewakili fitur (feature) atau variabel yang akan digunakan dalam proses klasifikasi.



Gambar 4.5 Set Role Variabel Target

Gambar 4.5 menampilkan tahap formatting kolom dalam proses impor data menggunakan RapidMiner Studio. Pada tahap ini, pengguna diminta untuk menentukan peran (*role*) dari masing-masing atribut atau kolom yang telah dipilih sebelumnya dalam dataset. Fokus utama gambar ini adalah penetapan variabel target (label) untuk proses klasifikasi. Variabel target dalam penelitian ini adalah “Kriteria”, yang akan digunakan oleh algoritma *Naïve Bayes* untuk membedakan antara karyawan yang “Disiplin” dan “Tidak Disiplin” berdasarkan atribut-atribut yang digunakan.

Penetapan peran atribut sebagai label merupakan langkah sangat penting dalam proses pembelajaran mesin. Tanpa penetapan ini, RapidMiner tidak akan mengetahui atribut mana yang menjadi target prediksi. Dalam konteks penelitian ini: Label: Kriteria (Disiplin / Tidak Disiplin). Fitur input: S1, S2, S3, S4, S5, Lembur, dan Total Kerja. Dengan set role yang tepat, algoritma *Naïve Bayes* dapat

menghitung probabilitas kemunculan label berdasarkan kombinasi atribut kedisiplinan karyawan.

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

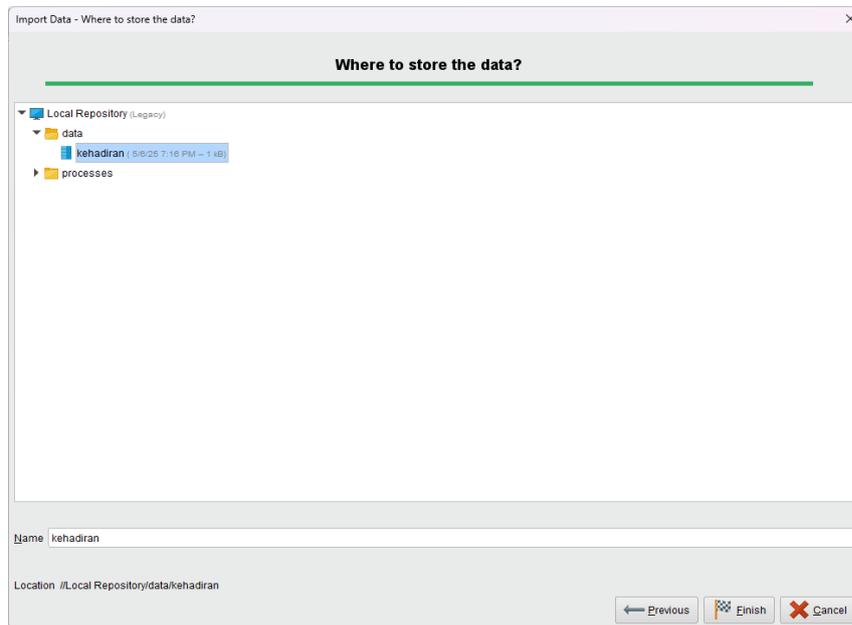
	Nama	S1	S2	S3	S4	S5	Lembur	Total Kerja
	<i>polynomial</i>	<i>integer</i>						
1	Abdul Rahman H...	0	0	0	0	0	0	29
2	Abdul Susilo	0	0	0	0	0	0	29
3	Abdullah Effendi	0	0	0	0	0	0	29
4	Ade Dermawan	0	0	0	0	0	0	29
5	Ade Rahmadani	0	0	0	0	0	0	29
6	Ade Safrizal	0	0	0	0	0	0	29
7	Adi	0	0	0	0	0	7	36
8	Adi	0	0	0	0	0	0	29
9	Adi Bakliar	0	0	0	0	0	0	29
10	Adi Firmansyah	0	0	0	0	0	0	29
11	Adi Gunawan	0	0	0	0	0	0	29
12	Adi Pranata Irawan	2	0	0	0	0	0	27
13	Adi Prayetno	0	0	0	0	0	5	34
14	Adi Priono	0	0	0	0	0	0	29
15	Adi Sukardi	3	0	0	0	0	0	26
16	Adi Suseno	2	0	0	0	0	0	27
17	Adi Sutoyo	4	0	0	0	0	0	25
18	Aditya Rhamadan	0	0	0	0	0	0	29

no problems.

← Previous Next → ✖ Cancel

Gambar 4.6 Format dan Tipe Data

Gambar 4.6 memperlihatkan tampilan akhir dari proses penyesuaian format dan tipe data sebelum data digunakan dalam proses machine learning di RapidMiner Studio. Tahapan ini merupakan bagian dari proses impor data yang dikenal dengan jendela “Format your columns”. Tujuan utama dari langkah ini adalah untuk memastikan bahwa setiap atribut (kolom) dalam dataset memiliki tipe data dan peran (role) yang sesuai dengan karakteristik datanya dan kebutuhan algoritma klasifikasi yang akan digunakan.



Gambar 4. 7 Direktori Penyimpanan File

Gambar 4.7 memperlihatkan tampilan jendela RapidMiner pada tahap akhir proses import dataset, tepatnya di bagian “Where to store the data?”. Pada tahap ini, pengguna menentukan lokasi penyimpanan internal untuk menyimpan dataset yang telah berhasil diimpor, agar dapat digunakan kembali dalam proses analisis selanjutnya tanpa perlu mengimpor ulang.

Row No.	Kriteria	Nama	S1	S2	S3	S4	S5	Lembur	Total Kerja
1	Rajin	Abdul Rahma...	0	0	0	0	0	0	29
2	Rajin	Abdul Susilo	0	0	0	0	0	0	29
3	Rajin	Abdullah Eff...	0	0	0	0	0	0	29
4	Rajin	Adi Dermaw...	0	0	0	0	0	0	29
5	Rajin	Adi Rahmad...	0	0	0	0	0	0	29
6	Rajin	Adi Sahtal	0	0	0	0	0	0	29
7	Rajin	Adi	0	0	0	0	0	7	36
8	Rajin	Adi	0	0	0	0	0	0	29
9	Rajin	Adi Baktar	0	0	0	0	0	0	29
10	Rajin	Adi Firmansy...	0	0	0	0	0	0	29
11	Rajin	Adi Gunawan	0	0	0	0	0	0	29
12	Tidak Rajin	Adi Pranata K...	2	0	0	0	0	0	27
13	Rajin	Adi Prayitno	0	0	0	0	0	5	34
14	Rajin	Adi Priono	0	0	0	0	0	0	29
15	Tidak Rajin	Adi Sukardi	3	0	0	0	0	0	26
16	Tidak Rajin	Adi Suseno	2	0	0	0	0	0	27
17	Tidak Rajin	Adi Suloyo	4	0	0	0	0	0	25
18	Rajin	Aditya Rham...	0	0	0	0	0	0	29

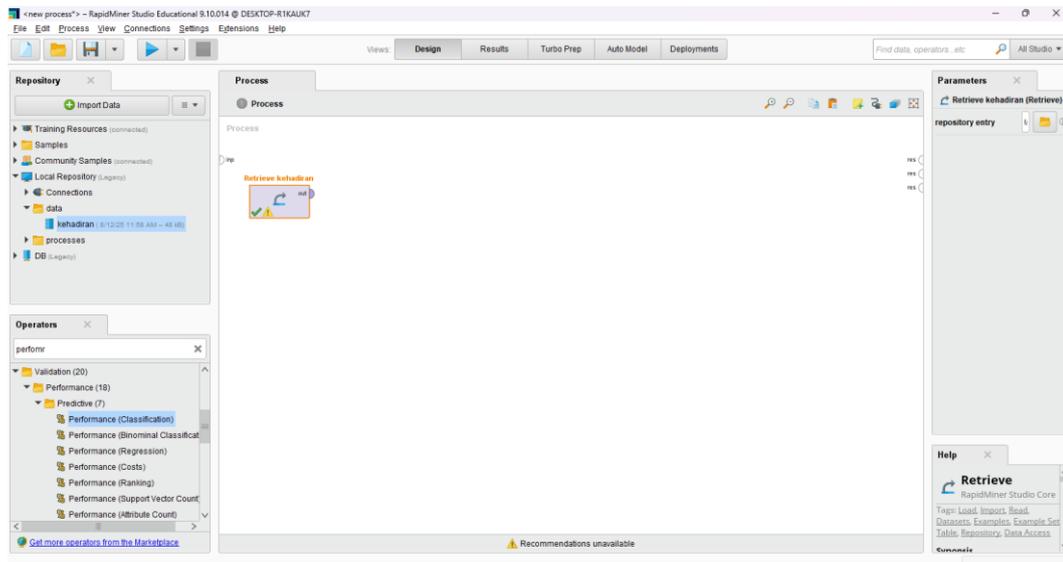
Gambar 4. 8 Preview Dataset

Gambar 4.8 menampilkan tampilan jendela RapidMiner Studio saat pengguna sedang melakukan preview dataset kedisiplinan yang telah berhasil diimpor ke dalam sistem. Dataset ini merupakan kumpulan data kedisiplinan karyawan yang digunakan dalam penelitian untuk memprediksi kedisiplinan dengan algoritma *Naïve Bayes*. Gambar ini menunjukkan tampilan aktif pada tab Results, yaitu bagian dari RapidMiner yang menampilkan hasil eksekusi dari operator seperti Read Excel atau Retrieve. Tab ini memungkinkan pengguna melihat isi dataset sebelum melakukan proses lebih lanjut seperti preprocessing, training model, dan evaluasi.

Nama Dataset ditampilkan pada tab “ExampleSet (Local Repository/data/karyawan)”, menandakan bahwa dataset yang sedang ditinjau adalah kedisiplinan yang tersimpan di repositori lokal dalam folder data. Jumlah Data menyebutkan: “ExampleSet (1.022 examples, 1 special attribute, 8 regular attributes)”. Artinya: Total 1.022 baris data (instances). Terdapat 1 atribut khusus (special attribute), yaitu atribut Kriteria yang telah diset sebagai label (target). Dan terdapat 8 atribut regular (fitur prediktor).

3.1.2 Pra-pemrosesan Data

Setelah data berhasil diimpor ke dalam RapidMiner, langkah selanjutnya adalah melakukan pra-pemrosesan data. Tahap ini sangat penting untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam proses analisis atau pemodelan machine learning.



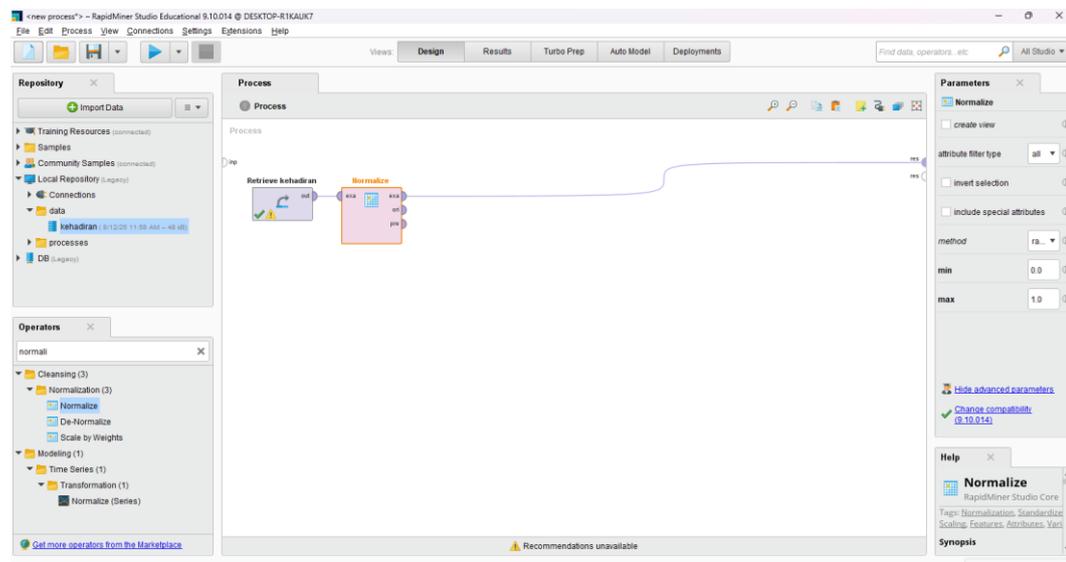
Gambar 4. 9 Retrieve Data

Gambar 4.9 menunjukkan tampilan proses awal dalam RapidMiner Studio, yaitu langkah Retrieve “kedisiplinan” yang berfungsi untuk mengambil dataset yang sudah disimpan di repository lokal agar dapat digunakan dalam pra-pemrosesan data.

Name	Type	Missing	Statistics
Kriteria	Polynomial	0	Level: Tidak Rajin (201), Rajin (731) Values: Rajin (731), Tidak Rajin (201)
Nama	Polynomial	0	Level: Zunaidi Nasution (1), Must Junaidi (6) Values: Junaidi (6), Sugianto (6), ... [928 more]
S1	Integer	0	Min: 0, Max: 17, Average: 0.659
S2	Integer	0	Min: 0, Max: 11, Average: 0.050
S3	Integer	0	Min: 0, Max: 3, Average: 0.006
S4	Integer	0	Min: 0, Max: 1, Average: 0.003
S5	Integer	0	Min: 0, Max: 1, Average: 0.004
Lembur	Integer	0	Min: 0, Max: 7, Average: 0.607
Total Kerja	Integer	0	Min: 7, Max: 36, Average: 29.885

Gambar 4. 10 Handle Missing Value

Gambar 4.10 ini menampilkan antarmuka RapidMiner Studio saat pengguna meninjau statistik ringkasan dari dataset 'kedisiplinan' setelah proses import selesai. Tampilan ini muncul di tab Results, bagian dari operator 'Read Excel' atau 'Retrieve', yang digunakan untuk meninjau kualitas data sebelum memasuki tahap preprocessing atau pemodelan. Kolom Missing menunjukkan nilai 0 untuk semua atribut, yang berarti tidak ada data yang hilang dalam dataset ini. Ini menunjukkan bahwa dataset telah siap digunakan langsung tanpa perlu proses imputasi atau pembersihan data lanjutan.



Gambar 4.11 Normalisasi Data

Gambar 4.11 merupakan tahapan penting dalam preprocessing sebelum dilakukan pemodelan di RapidMiner yaitu tahap Normalisasi Data. Gambar ini menunjukkan tampilan antarmuka RapidMiner Studio pada tab Design, ketika pengguna sedang membangun alur proses untuk proyek klasifikasi prediksi kedisiplinan karyawan. Fokus proses ini adalah pada penerapan proses normalisasi

data menggunakan operator `Normalize`. Operator ini berfungsi untuk melakukan normalisasi skala nilai pada atribut numerik.

Normalisasi berguna untuk menstandarkan rentang nilai sehingga metode klasifikasi seperti *Naive Bayes* dapat bekerja secara optimal dan tidak berat sebelah. Normalisasi memastikan bahwa semua atribut numerik memberikan kontribusi yang seimbang. Dengan nilai input dalam rentang standar (0–1), model dapat belajar pola data dengan lebih efisien. Normalisasi membantu mempercepat proses pelatihan karena menghindari nilai ekstrim atau *outlier* yang besar.

Row No.	Kriteria	S1	S2	S3	S4	S5	Lembur	Total Kerja	Nama
1	Rajin	0	0	0	0	0	0	0.759	Abdul Rahma...
2	Rajin	0	0	0	0	0	0	0.759	Abdul Susilo
3	Rajin	0	0	0	0	0	0	0.759	Abdullah Effe...
4	Rajin	0	0	0	0	0	0	0.759	Ade Dermaw...
5	Rajin	0	0	0	0	0	0	0.759	Ade Rahmad...
6	Rajin	0	0	0	0	0	0	0.759	Ade Saftzal
7	Rajin	0	0	0	0	0	1	1	Adi
8	Rajin	0	0	0	0	0	0	0.759	Adi
9	Rajin	0	0	0	0	0	0	0.759	Adi Bakhtar
10	Rajin	0	0	0	0	0	0	0.759	Adi Firmansy...
11	Rajin	0	0	0	0	0	0	0.759	Adi Gunawan
12	Tidak Rajin	0.118	0	0	0	0	0	0.690	Adi Pranata Ir...
13	Rajin	0	0	0	0	0	0.714	0.931	Adi Prayetho
14	Rajin	0	0	0	0	0	0	0.759	Adi Priyono
15	Tidak Rajin	0.176	0	0	0	0	0	0.655	Adi Sukardi
16	Tidak Rajin	0.118	0	0	0	0	0	0.690	Adi Suseno
17	Tidak Rajin	0.235	0	0	0	0	0	0.621	Adi Sutoyo
18	Rajin	0	0	0	0	0	0	0.759	Aditya Rham...

Gambar 4. 12 Hasil Normalisasi Data

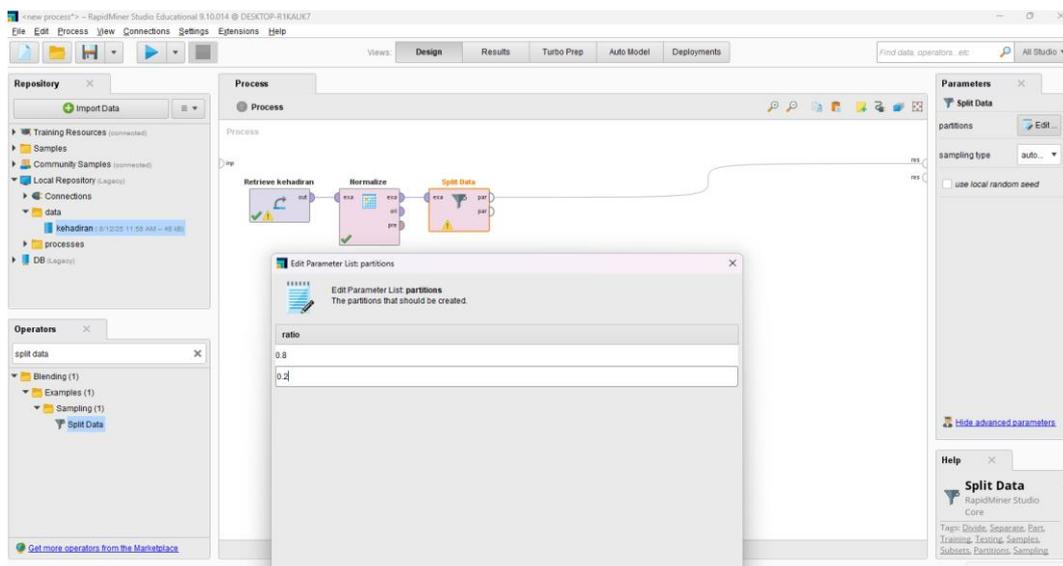
Gambar 4.12 menampilkan hasil proses normalisasi data pada dataset karyawan di PT. PP London Sumatra Indonesia Tbk, yang dilakukan menggunakan perangkat lunak RapidMiner. Normalisasi ini bertujuan untuk mengubah nilai atribut ke dalam skala yang seragam (umumnya antara 0 hingga 1) sehingga meminimalkan bias akibat perbedaan skala antar variabel saat digunakan dalam pemodelan algoritma *machine learning*.

Kolom "Lembur" berisi data jumlah lembur yang telah dinormalisasi, di mana angka 0 menunjukkan tidak ada lembur dan angka >0 menunjukkan ada aktivitas lembur dengan skala yang disesuaikan. Kolom "Total Kerja" menunjukkan persentase atau rasio total Kedisiplinan karyawan dibandingkan total hari kerja, setelah dinormalisasi. Misalnya, nilai 10.759 berarti tingkat Kedisiplinan karyawan tersebut sekitar 75,9%. Dataset berisi 1.022 baris data (examples), di mana setiap baris merepresentasikan satu karyawan dengan atribut yang sudah dinormalisasi.

Gambar ini menunjukkan bahwa seluruh variabel numerik telah melalui proses normalisasi, sehingga siap digunakan untuk proses pelatihan dan pengujian model prediksi menggunakan algoritma machine learning seperti *Naïve Bayes*. Normalisasi ini penting untuk menghindari dominasi variabel dengan nilai skala besar terhadap hasil prediksi.

3.1.3 Pembagian Data

Setelah data berhasil diimpor dan dilakukan pra-pemrosesan seperti normalisasi, langkah selanjutnya adalah pembagian data (data splitting). Proses ini sangat penting dalam machine learning untuk memastikan model yang dibangun dapat diuji secara objektif.



Gambar 4.13 Pembagian Data

Gambar 4.13 memperlihatkan proses pembagian data (*Split Data*) pada perangkat lunak RapidMiner Studio, yang dilakukan setelah tahap normalisasi data. Tujuan dari pembagian data ini adalah untuk memisahkan dataset menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*) agar model prediksi dapat dibangun dan dievaluasi secara objektif. Proses ini memperlihatkan bahwa 80% dari total data akan digunakan sebagai data latih, yaitu data yang digunakan untuk membangun dan melatih model Algoritma *Naive Bayes*. Sedangkan 20% dari total data akan digunakan sebagai data uji, yaitu data yang digunakan untuk mengukur kinerja model yang telah dilatih. Data latih (80%) digunakan untuk mempelajari pola dan hubungan antar variabel dalam dataset, sehingga model dapat melakukan prediksi. Data uji (20%) digunakan untuk menguji sejauh mana model dapat melakukan prediksi terhadap data baru yang belum pernah dilihat sebelumnya, sehingga mengukur kemampuan generalisasi model.

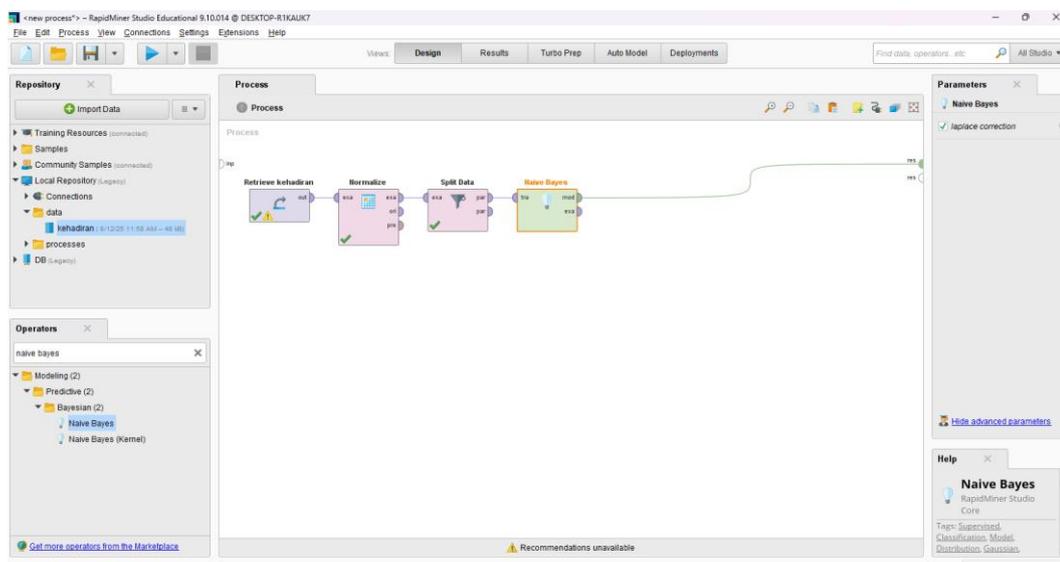
Row No.	Kriteria	S1	S2	S3	S4	S5	Lembur	Total Kerja	Nama
1	Rajin	0	0	0	0	0	0	0.759	Abdul Rahma...
2	Rajin	0	0	0	0	0	0	0.759	Abdul Susilo
3	Rajin	0	0	0	0	0	0	0.759	Abdullah Effe...
4	Rajin	0	0	0	0	0	0	0.759	Ade Dermaw...
5	Rajin	0	0	0	0	0	0	0.759	Ade Saftizal
6	Rajin	0	0	0	0	0	0	0.759	Adi
7	Rajin	0	0	0	0	0	0	0.759	Adi Bakhtar
8	Rajin	0	0	0	0	0	0	0.759	Adi Firmansy...
9	Rajin	0	0	0	0	0	0	0.759	Adi Gunawan
10	Tidak Rajin	0.118	0	0	0	0	0	0.690	Adi Pranata Ir...
11	Rajin	0	0	0	0	0	0.714	0.931	Adi Prayelno
12	Tidak Rajin	0.176	0	0	0	0	0	0.655	Adi Sukardi
13	Tidak Rajin	0.118	0	0	0	0	0	0.690	Adi Suseno
14	Tidak Rajin	0.235	0	0	0	0	0	0.621	Adi Sutoyo
15	Rajin	0	0	0	0	0	0	0.759	Aditya Wahyudi
16	Rajin	0	0	0	0	0	0	0.759	Ageng Siswa...
17	Rajin	0	0	0	0	0	0.857	0.966	Agus
18	Rajin	0	0	0	0	0	0	0.759	Agus Aditya P...

Gambar 4. 14 Hasil Pembagian Data

Gambar 4.14 menunjukkan tampilan hasil pembagian data (split data) pada perangkat lunak RapidMiner Studio setelah dilakukan proses normalisasi data. Pada tahap ini, dataset Kedisiplinan karyawan telah dipisahkan menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*), sesuai dengan rasio yang telah ditentukan sebelumnya yaitu 80% untuk data latih dan 20% untuk data uji. Berdasarkan hasil pembagian data (split data) dapat dijelaskan: Jumlah data seluruhnya = 1.022 contoh (examples). Jumlah data latih (training data) = 818 contoh, yang merupakan 80% dari total data. Jumlah data uji (testing data) = 204 contoh, yang merupakan 20% dari total data.

3.1.4 Implementasi Metode *Naïve Bayes*

Bagian ini menjelaskan proses Implementasi Metode *Naïve Bayes*. Ini merupakan tahap di mana algoritma *Naïve Bayes* diterapkan pada data pelatihan yang telah disiapkan sebelumnya.



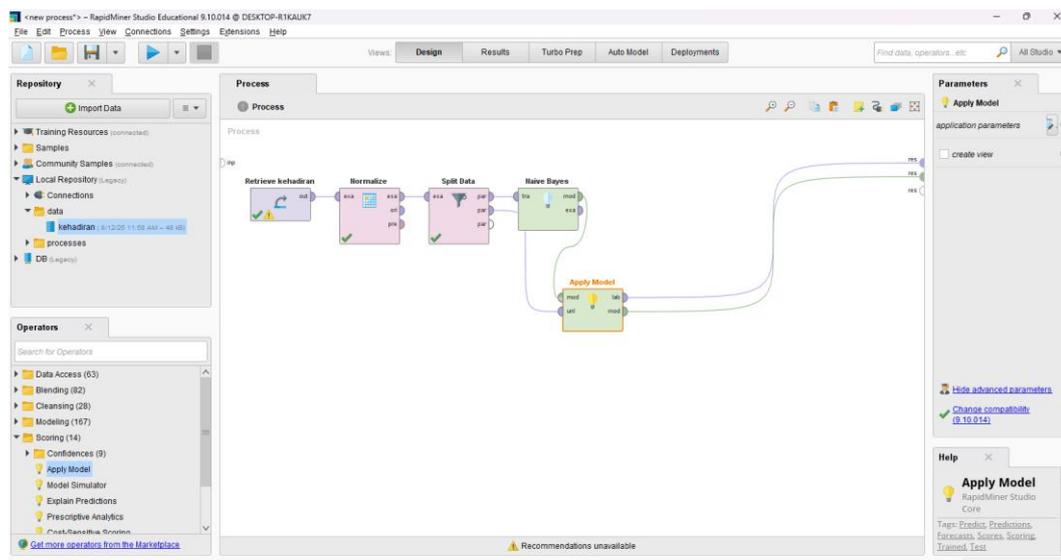
Gambar 4. 15 Implementasi *Naïve Bayes*

Gambar 4.15 menunjukkan proses implementasi algoritma *Naïve Bayes* dalam RapidMiner Studio yang digunakan dalam penelitian ini. Proses ini merupakan salah satu tahapan utama dalam pembangunan model klasifikasi untuk prediksi Kedisiplinan karyawan pada PT. PP London Sumatra Indonesia Tbk. Sei Rumbiya Estate. Implementasi ini dilakukan dengan pendekatan visual workflow yang memanfaatkan serangkaian operator yang saling terhubung dalam tab Design.

Operator utama dalam implementasi ini adalah *Naïve Bayes*. Operator ini menerima input data latih dari Split Data dan menghasilkan model klasifikasi berbasis probabilistik. Parameter Laplace correction diaktifkan untuk menghindari probabilitas nol pada atribut kategori yang tidak muncul dalam data latih. Model yang dihasilkan selanjutnya akan digunakan untuk melakukan klasifikasi terhadap data uji pada tahap berikutnya.

Gambar ini menunjukkan tahapan penting dalam membangun model klasifikasi menggunakan *Naïve Bayes* dalam RapidMiner. Dimulai dari

pengambilan data, normalisasi, pembagian data, hingga pelatihan model. Implementasi ini mencerminkan pendekatan sistematis dan terstruktur dalam membangun pipeline machine learning, khususnya dalam penelitian yang berkaitan dengan klasifikasi prediksi Kedisiplinan karyawan.

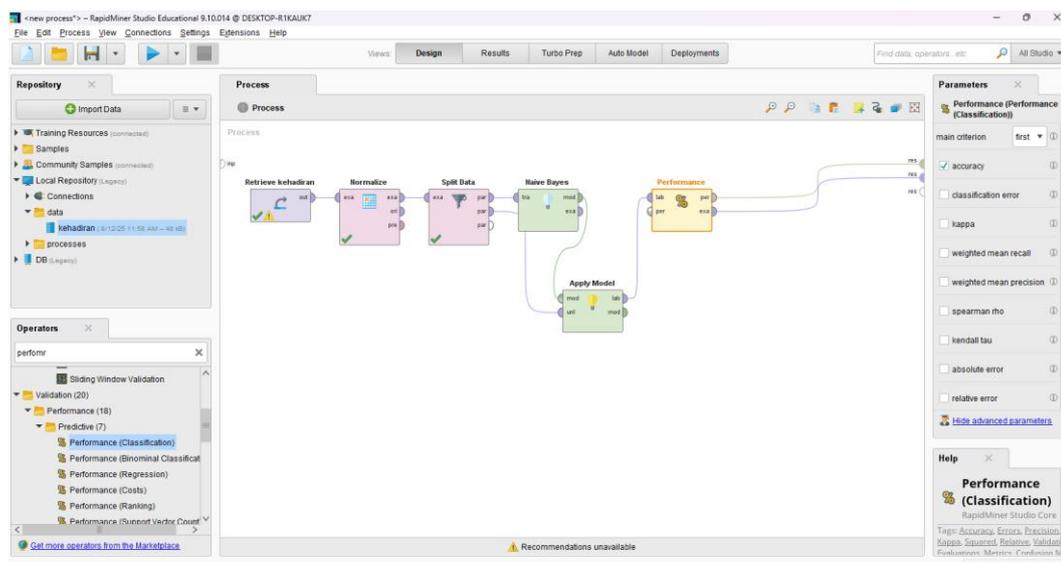


Gambar 4. 16 Apply Model

Gambar 4.16 menunjukkan proses penerapan model (*Apply Model*) terhadap data pengujian yang telah dibagi sebelumnya. Proses ini merupakan lanjutan dari hasil pelatihan model *Naïve Bayes* pada tahap sebelumnya. Tujuan utama dari langkah ini adalah untuk menguji kemampuan model dalam melakukan prediksi terhadap data baru yang belum pernah dilihat oleh model selama proses pelatihan.

Adapun operator utama yang digunakan dalam proses ini adalah: *Apply Model* dan *Output Apply Model*. Operator *Apply Model* berfungsi untuk menerapkan model klasifikasi *Naïve Bayes* yang telah dibangun terhadap subset data uji. Operator ini menerima dua input, yaitu: *mod* (model) dari output operator *Naïve Bayes*, dan *unl* (*unlabeled data*) yang berasal dari output “testing data”

operator *Split Data*. Hasil dari operator ini adalah data yang telah diprediksi kelasnya oleh model *Naïve Bayes* berdasarkan atribut-atribut input seperti S-S5, Lembur, dan Total Kerja. Sedangkan Output dari operator Apply Model menghasilkan sebuah dataset baru yang terdiri dari data uji disertai kolom label hasil prediksi (prediction) dari model. Kolom ini nantinya akan dibandingkan dengan label aktual untuk mengetahui seberapa akurat model dalam memprediksi Kedisiplinan karyawan. Gambar 4.16 dengan demikian merepresentasikan proses pengujian prediktif yang krusial dalam proses data mining. Tanpa tahap ini, tidak akan diketahui apakah model mampu melakukan klasifikasi dengan baik atau tidak.



Gambar 4. 17 Pengukuran Performance

Gambar 4.17 memperlihatkan proses evaluasi performa model klasifikasi *Naïve Bayes*. Evaluasi dilakukan dengan menggunakan operator Performance yang mengukur seberapa baik hasil prediksi yang dihasilkan oleh model terhadap data uji, berdasarkan metrik evaluasi klasifikasi. Operator Performance digunakan untuk mengevaluasi hasil prediksi yang telah dihasilkan oleh operator Apply Model. Input

dari operator ini adalah hasil prediksi (*example set*) yang telah dilengkapi dengan kolom label aktual dan kolom prediksi. Operator ini menghitung metrik-metrik penting yang mencakup: Akurasi (*Accuracy*): Persentase prediksi yang benar dari keseluruhan data uji. *Precision*: Kemampuan model dalam mengklasifikasikan karyawan yang benar-benar Disiplin secara tepat. *Recall (Sensitivity)*: Seberapa baik model mendeteksi seluruh karyawan yang benar-benar Disiplin. *F1-Score*: Rata-rata harmonis dari *precision* dan *recall*.

Hasil evaluasi akan ditampilkan dalam bentuk tabel atau ringkasan angka performa yang ditampilkan secara otomatis oleh RapidMiner di *tab Results*. Nilai-nilai metrik ini akan dijadikan dasar dalam menilai apakah model layak digunakan untuk proses prediksi di lingkungan nyata, khususnya untuk mendukung memprediksi Kedisiplinan karyawan.

Dengan adanya tahapan *Apply Model* dan *Performance* seperti ditunjukkan dalam Gambar 4.16 dan Gambar 4.17, maka keseluruhan proses klasifikasi *Naïve Bayes* telah dijalankan secara lengkap. Selanjutnya, hasil evaluasi yang dihasilkan akan dibahas secara lebih rinci dalam subbab Pembahasan, yang akan mencakup interpretasi nilai akurasi dan metrik lainnya serta relevansinya dalam konteks penelitian ini.

3.2 Pembahasan

Hasil dari masing-masing algoritma akan menunjukkan seberapa baik mereka dapat memprediksi Kedisiplinan karyawan berdasarkan Kehadiran dengan fitur-fitur yang telah ditentukan.

	true Rajin	true Tidak Rajin	class precision
accuracy: 100.00%			
pred Rajin	145	0	100.00%
pred Tidak Rajin	0	58	100.00%
class recall	100.00%	100.00%	

Gambar 4. 18 Hasil Performance *Naïve Bayes*

Gambar 4.18 menunjukkan hasil evaluasi kinerja model *Naïve Bayes* yang digunakan untuk memprediksi kriteria Kedisiplinan karyawan (Disiplin atau Tidak Disiplin) pada data uji. Dari tabel performa yang ditampilkan oleh RapidMiner, dapat dijelaskan secara detail sebagai berikut:

1. Akurasi (Accuracy)

Model mencapai 100,00% akurasi, artinya seluruh prediksi yang dilakukan model sesuai dengan label aktual pada data uji. Tidak ada satupun kesalahan prediksi yang terjadi pada pengujian ini.

2. Presisi (Precision)

Presisi untuk kelas Disiplin adalah 100,00%. Ini berarti semua prediksi Disiplin benar-benar Disiplin. Presisi untuk kelas Tidak Disiplin juga 100,00%, artinya semua prediksi Tidak Disiplin benar-benar sesuai.

3. Recall (Sensitivitas)

Recall untuk kelas Disiplin adalah 100,00%, yang berarti seluruh data Disiplin berhasil teridentifikasi oleh model. Recall untuk kelas Tidak Disiplin juga 100,00%, menandakan semua data Tidak Disiplin berhasil diklasifikasikan dengan benar.

4. F1-Score (tidak ditampilkan secara eksplisit dalam gambar): dengan precision 100% dan recall 100%, nilai F1-score dapat dihitung sebagai berikut:

$$\begin{aligned} \text{F1-score} &= 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \\ &= 2 \times (1 \times 1) / (1 + 1) \\ &\approx 1 \text{ atau } 100\% \end{aligned}$$

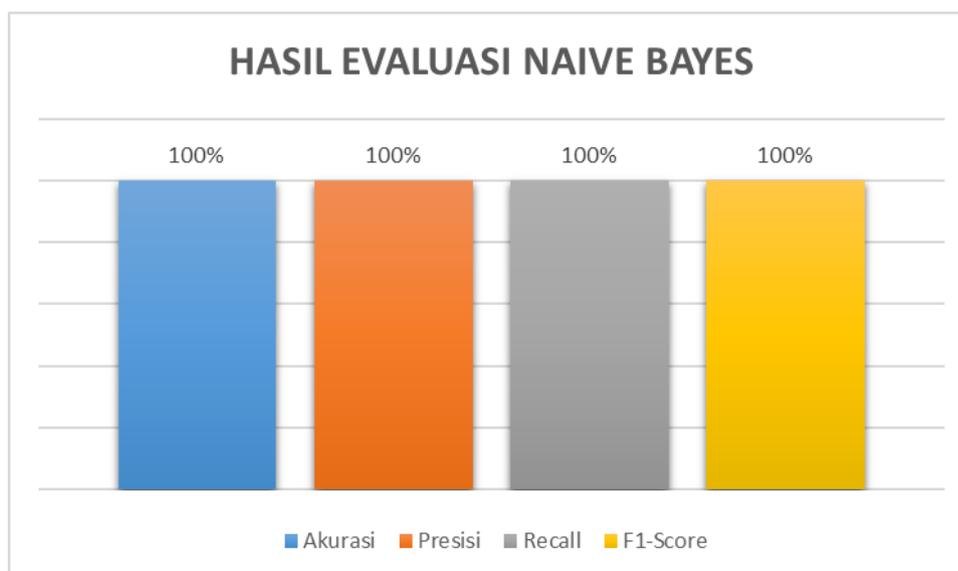
Berdasarkan data di atas, dapat disusun confusion matrix seperti diperlihatkan pada tabel 4.2

Tabel 4. 2 Confusion Matrix

	Aktual Positif	Aktual Negatif
Prediksi Positif	146 (TP)	0 (FP)
Prediksi Negatif	0 (FN)	58 (TN)

1. Dari total 204 data uji, model berhasil memprediksi keseluruhan data dengan benar, terdiri dari:
 - o 146 data positif yang diklasifikasikan secara benar (True Positif).
 - o 58 data negatif yang diklasifikasikan secara benar (True negatif).

2. Tidak terdapat data aktual positif yang diprediksi sebagai negatif (False Negative) dan data aktual negatif yang diprediksi sebagai positif (False Positif).



Gambar 4. 19 Metrik Evaluasi Kinerja *Naïve Bayes*

Berdasarkan hasil evaluasi kinerja model *Naïve Bayes* yang ditunjukkan pada Gambar 4.19, diperoleh nilai akurasi sebesar 100%, yang berarti seluruh prediksi model sesuai dengan data aktual tanpa adanya kesalahan klasifikasi. Nilai precision untuk kedua kelas, yaitu Disiplin dan Tidak Disiplin, masing-masing mencapai 100%, menunjukkan bahwa semua prediksi positif yang diberikan model benar adanya. Nilai recall untuk kedua kelas juga sebesar 100%, menandakan bahwa model berhasil mengidentifikasi seluruh data positif yang ada di dalam dataset uji. Selanjutnya, nilai F1-Score yang merupakan gabungan dari precision dan recall juga mencapai 100%, yang mengindikasikan keseimbangan sempurna antara kemampuan model dalam menghindari false positives maupun false negatives. Hasil ini menunjukkan bahwa model *Naïve Bayes* yang dibangun mampu

memberikan performa klasifikasi yang sangat baik dan konsisten dalam memprediksi Kedisiplinan karyawan berdasarkan variabel-variabel yang digunakan.