3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Manual

Pada bagian ini menjelaskan hasil dan pembahasan dari perhitungan manual dengan metode naïve bayes, dalam pembahasan ini menjelaskan perhitungan menggunakan pendekatan gaussian naïve bayes (GNB). Untuk melakukan analisis, setiap siswa diberi nilai likelihood, posterior, dan prediksi akhir berdasarkan karakteristik kedisiplinan yang telah ditentukan.

3.1.1. Proses rincian perhitungan manual per siswa

Untuk memastikan hasil klasifikasi yang akurat diperlukan data perhitungan manual per siswa. Ini adalah ringkasan hasil perhitugan manual yang menggunakan pendekatan likelihood, posterior, dan hasil prediksi.

1. AHMAD SETYAWAN — Label: Disiplin
   Produk likelihood (Disiplin) ≈ 4.7259×10^{-7}
   Produk likelihood (Tidak) ≈ 1.0434×10^{-8}
   Posterior P(Disiplin|X) ≈ 3.2335×10^{-7} > P(Tidak|X) ≈ 3.2948×10^{-9}
   → Prediksi: Disiplin — Benar
2. AHMAT RIFKI ADI AFRIYAN — Label: Disiplin
   Produk likelihood (Disiplin) ≈ 8.93629×10^{-7}
   Produk likelihood (Tidak) ≈ 2.28049×10^{-10}
   Posterior P(Disiplin|X) ≈ 6.11430×10^{-7} > P(Tidak|X) ≈ 7.20156×10^{-11}
   → Prediksi: Disiplin — Benar
3. ALVIRA AZZAHRA — Label: Disiplin
   Produk likelihood (Disiplin) ≈ 1.15687×10^{-6}
   Produk likelihood (Tidak) ≈ 1.18080×10^{-8}
   Posterior P(Disiplin|X) ≈ 7.91543×10^{-7} > P(Tidak|X) ≈ 3.72884×10^{-9}
   → Prediksi: Disiplin — Benar
4. ANGGA JULIAN RAMADHANI — Label: Tidak Disiplin
   Produk likelihood (Disiplin) ≈ 1.75937×10^{-7}
   Produk likelihood (Tidak Disiplin) ≈ 2.12491×10^{-7}
   Posterior P(Disiplin|X) ≈ 1.20378×10^{-7}
   Posterior P(Tidak|X) ≈ 6.71025×10^{-8}
   → Prediksi: Tidak Disiplin — BENAR
5. ARYANI DEWI LARASATI — Label: Disiplin
   Produk likelihood (Disiplin) ≈ 6.25×10^{-6}
   Produk likelihood (Tidak) ≈ 1.87×10^{-12}
   Posterior P(Disiplin|X) ≈ 4.27×10^{-6} > P(Tidak|X) ≈ 5.91×10^{-13}
   → Prediksi: Disiplin — Benar

3.1.2. Analisa dan perhitungan manual

1. Perhitungan dan analisis dilakukan secara manual menggunakan pendekatan Gaussian Naive Bayes (GNB). Hal ini disebabkan fakta bahwa setiap elemen dalam dataset adalah data numerik kontinu. Berikut adalah hasil perhitungan:
   Menentukan probabilitas prior (P(C)) berdasarkan jumlah data siswa:
a. Jumlah siswa *Disiplin* = 13
b. Jumlah siswa *Tidak Disiplin* = 6
   Maka diperoleh:

$$P(Disiplin) = \frac{13}{19} = 0.6842, P(Tidak\ Disiplin) = \frac{6}{19} = 0.3158$$

2. Menghitung nilai rata-rata (μ) dan standar deviasi (σ) untuk setiap atribut pada masing-masing kelas (*Disiplin* dan *Tidak Disiplin*). Nilai ini digunakan untuk membentuk fungsi likelihood menggunakan distribusi Gaussian:

$$P(x_i \mid H) = \frac{1}{\sqrt{2\pi\sigma_H^2}} e^{-\frac{(x_i-\mu_H)^2}{2\sigma_H^2}} \quad (2)$$

3. Menghitung probabilitas posterior (P(H|X)) dengan mengalikan nilai prior dan likelihood dari seluruh atribut:

$$P(H \mid X) = P(H) \times \prod_i P(x_i \mid H) \quad (3)$$

4. Menentukan hasil prediksi.
   Kelas dengan nilai probabilitas posterior tertinggi menjadi hasil klasifikasi.
   Jika $P(Disiplin \mid X) > P(Tidak\ Disiplin \mid X)$, maka siswa dikategorikan *Disiplin*, dan sebaliknya.

3.1.3. Klasifikasi model

Pada pemodelan ini dilakukan pada Sekolah dasar kelas 5 dengan jumlah 19 siswa dengan total data testing 456 yang terdiri dari 9 attribut dan juga terdapat dataset juga untuk membantu pengolahan data dengan tools rapid miner. Berikut tabel hasil klasifikasi model yang ditunjukkan dalam matrix confusion menurut perhitugan diatas :

Table 2. confusion matrix pemodelan

	True disiplin	**True tidak disiplin**
Prediksi disiplin	13	2
Prediksi tidak disiplin	0	4

3.1.4. Evaluasi Kinerja Model

Table 3. Evaluasi hasil

Metrik	Rumus	Nilai
Akurasi	(TP + TN) / Total = (13 + 4) / 19	89,47%
Precision (Disiplin)	TP / (TP + FP) = 13 / 13	100%
Recall (Disiplin)	TP / (TP + FN) = 13 / 15	86,67%
F1-Score (Disiplin)	2 × (Prec × Rec) / (Prec + Rec)	92,86%

Tahap evaluasi yaitu pada confusion matrix berguna untuk menghasilkan Keputusan yang sudah dilakukan dari Pemodelan Naïve Bayes. Dari hasil akhir akan memberikan informasi berkaitan presisi, akurasi, dan recall[13]. Walaupun ada dua siswa Disiplin yang salah diklasifikasikan sebagai Tidak Disiplin, model menunjukkan hasil yang sangat baik dalam mengenali siswa Disiplin (ketepatan = 100%), dan mengenali semua siswa di kelas Tidak Disiplin dengan benar. Dan untuk melihat kemungkinan hubungan antara tingkat kedisiplinan dan hasil belajar, hasil evaluasi model juga dibandingkan dengan data nilai UTS dan UAS siswa. Meskipun nilai UTS dan UAS tidak digunakan dalam proses klasifikasi, data ini memberikan gambaran tambahan tentang karakteristik siswa yang disiplin dan tidak disiplin.

3.1.5. Interpretasi Hasil

Hasil perhitungan manual menggunakan Gaussian Naive Bayes menunjukkan performa klasifikasi yang tinggi dengan akurasi 89,47%. Metode ini terbukti efektif dalam memetakan perilaku kedisiplinan siswa berdasarkan data numerik seperti absensi, ketepatan hadir, kepatuhan terhadap tata tertib, pengumpulan tugas, dan sikap di kelas. Proses kerja Naive Bayes digambarkan secara sistematis dalam perhitungan manual 19 siswa yang menggunakan pendekatan probabilitas-posterior-prediksi. Untuk hasil pengujian dengan perhitungan manual terdapat 13 siswa ternilai disiplin namun ada siswa dengan nilai PTS dan PAS yang masih kurang dengan nilai kedisiplinan tadi guru bisa memberikan nilai tambahan atau kkm untuk siswa yang masih kurang, untuk 2 siswa yang dinilai pada label tidak disiplin tapi di prediksi disiplin itu mempunyai nilai PTS dan PAS yang cukup tetapi juga mungkin siswa diberi remidi untuk mendapatkan nilai yang masih kurang, dan untuk 4 siswa yang true positive tidak disiplin mempunyai nilai PTS dan PAS ada yang masih di bawah KKM jadi untuk siswa yang tidak disiplin mempunyai nilai kurang harap bisa mengikuti remidi pada PTS/PAS agar guru bisa membuat raport hasil belajar. Jadi fungsi dari penelitian ini juga selain menjadi tolak ukur Tingkat kedisiplinan siswa juga bisa membantu guru dalam membuat nilai hasil belajar siswa, jika memang anak tersebut nilainya kurang bisa dibantu dengan nilai kedisiplinan yang membuat nilai menjadi bertambah.

3.2. Pengujian cross validation dan optimasi GA

Untuk menguji klasifikasi data agar lebih akurat kita membutuhkan tools penunjang yaitu Rapid miner, proses pengumpulan dataset, pemodelan, evaluasi serta validasi dengan metode cross validation dan optimasi GA (Genetic Algorithm) dengan RapidMiner (Altair AI Studio) versi 2026.0.1.

Tujuan dari pengujian menggunakan tools dengan metode ini untuk membentuk evaluasi klasifikasi label disiplin dan tidak disiplin dan mengetahui nilai accuracy, precision, dan recall dari klasifikasi data kedisiplinan siswa[14]. Dan nantinya terdapat komponen evaluasi dari klasifikasi data kedisiplinan tersebut antara lain

1. Banyaknya kelas positif yang diklasifikasi secara benar (TP).
2. Banyaknya kelas negatif yang diklasifikasi secara positif kelas (FP).
3. Banyaknya kelas positif yang diklasifikasi secara negatif (FN).
4. Banyaknya kelas negatif yang diklasifikasi secara benar (TN).

Pada tabel berikut ini menunjukkan recall memerlukan komponen untuk menghitung precision

Table 4. Hasil Evaluasi Pengujian

	True disiplin	True tidak disiplin	Class precision
Prediksi disiplin			

	True disiplin	True tidak disiplin	Class precision
Prediksi tidak disiplin			
Class recall			

Konsep dari pengujian cross validation dengan optimasi GA menggunakan tools rapidminer ini sebetulnya sama dengan perhitungan manual yang mencari nilai precision, recall, dan akurasi dengan menggunakan hasil evaluasi dua kelas. Jika model yang digunakan telah dilatih untuk menentukan kelas yang sama pada data uji, klasifikasi bernilai benar. Pada pengujian ini terdapat 7 proses untuk melatih model klasifikasi kedisiplinan siswa yaitu penghimpunan dataset, penentuan attribut, pra pengolahan data, pemodelan/pengujian dengan cross validation, proses cross validation, hasil pengujian, Analisa pengujian.

1. Penghimpunan dataset
   Data training atau dataset bisa diperoleh dengan pengelompokkan pelanggaran siswa dan terbentuklah juga data testing yang mempunyai 9 attribut untuk menguji kedisiplinan siswa kelas 5 dengan dua label disiplin dan tidak disiplin. Pada data testing jumlah total datanya ada 456 pelanggaran siswa di data dari Januari – Juni (semester 2). Berikut gambar dataset untuk pengujian dengan cross validation dan optimasi GA untuk klasifikasi kedisiplinan siswa.

NO	NAMA	ABSENSI	KETEPATAN HADIR	KEPATUHAN TATA TERTIB	KETERTIBAN SERAGAM	TIDAK BOLOS PELAJARAN	PENGUMPULAN TUGAS	KESIAPAN DALAM PELAJARAN	PERILAKU DI KELAS	SIKAP DITEGUR	TOTAL	KEPUTUSAN
1	AHMAD SETYAWAN	4	3	1	3	1	0	1	1	2	16	DISIPLIN
2	AHMAT RIFKI ADI AFRIYAN	2	1	1	2	0	4	1	0	0	11	DISIPLIN
3	ALVIRA AZZAHRA	0	2	1	2	1	0	4	3	2	15	DISIPLIN
4	ANGGA JULIAN RAMADHANI	1	2	5	4	1	3	1	2	1	20	TIDAK DISIPLIN
5	ARYANI DEWI LARASATI	0	0	1	2	0	2	3	1	1	10	DISIPLIN
6	ASKA MUHAMMAD MAULANA	9	0	2	1	0	1	1	0	2	16	DISIPLIN
7	CANDRA SETIAWAN	0	2	3	0	2	3	1	5	1	17	DISIPLIN
8	DEVA SETYA ABDILA	1	0	1	4	1	4	1	2	2	16	DISIPLIN
9	EKA ARIYANINGSIH	1	1	2	0	1	4	4	4	0	17	DISIPLIN
10	ELISABETH NATALI VERONIKA DEDA	2	2	1	2	0	5	2	7	2	23	TIDAK DISIPLIN
11	ENGGAR SETYAJI	0	0	1	2	1	4	1	1	1	11	DISIPLIN
12	EUKHARISTIAN JOVANY	2	3	1	5	0	1	2	3	2	19	TIDAK DISIPLIN
13	GEISHA JULIA RAMADHANI	0	0	2	3	0	3	3	5	2	18	DISIPLIN
14	Hieronimus Andra Crisyantama	2	3	2	2	1	2	2	2	1	17	DISIPLIN
15	KHOIRUL HIDAYAH	2	2	2	4	0	1	2	0	2	15	DISIPLIN
16	MUHAMMAD PANJI SAPUTRA	1	0	3	1	1	3	4	1	3	17	DISIPLIN
17	RATNA PUSPITA	0	4	4	3	0	4	5	3	5	28	TIDAK DISIPLIN
18	SHINBIA NAIRRA	1	3	6	4	0	5	2	2	1	24	TIDAK DISIPLIN
19	TEGAR SETYO PRIHANTORO	4	3	7	4	0	7	10	6	4	45	TIDAK DISIPLIN

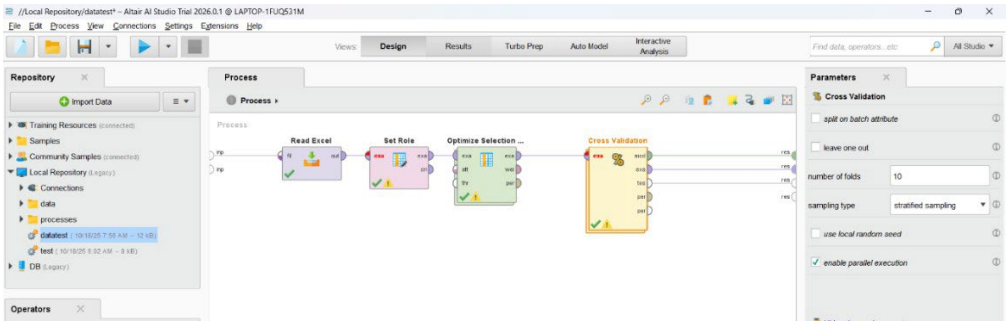
Gambar 4. Dataset pengujian

2. Penentuan atribut
   Dengan menggunakan cross-validation dan optimize parameters evolutionary bisa menentukan atribut label (target klasifikasi). Atribut "Kedisiplinan" diatur sebagai label, sedangkan atribut lainnya tetap sebagai regular. Tujuan RapidMiner adalah untuk menemukan variabel mana yang akan diprediksi oleh model Naive Bayes[15].
3. Pra pengolahan data
   Sebelum model diuji, tahap pemilihan optimal (evolusi) dilakukan untuk memilih atribut yang paling berdampak terhadap hasil klasifikasi. Operator ini menggunakan algoritma genetik (GA) untuk secara otomatis menemukan kombinasi atribut terbaik. Agar model menjadi lebih akurat, operator "Optimize Selection (Evolutionary)" berfungsi untuk menemukan subset atribut terbaik. Dia secara otomatis memilih beberapa atribut dan menghilangkan yang dianggap tidak penting. Oleh karena itu, jika hasil akhir menunjukkan hanya 7 atribut dari 9 atribut, maka algoritma genetik (evolutionary optimization) tidak memberikan kontribusi yang signifikan terhadap akurasi model.

KEPUTUSAN	NO	NAMA	ABSENSI	KETEPATA...	KEPATUHA...	KETERTIBA...	TIDAK BOL...	PENGUMP...	PERILAKU
?	?	?	?	?	?	?	?	?	?
DISIPLIN	1	AHMAD SETY...	4	3	1	3	1	0	1
DISIPLIN	2	AHMAT RIFKI ...	2	1	1	2	0	4	0
DISIPLIN	3	ALVIRA AZZA...	0	2	1	2	1	0	3
TIDAK DISIPLIN	4	ANGGA JULIA...	1	2	5	4	1	3	2
DISIPLIN	5	ARYANI DEWI...	0	0	1	2	0	2	1
DISIPLIN	6	ASKA MUHAM...	9	0	2	1	0	1	0
DISIPLIN	7	CANDRA SE...	0	2	3	0	2	3	5
DISIPLIN	8	DEVA SETYA ...	1	0	1	4	1	4	2
DISIPLIN	9	EKA ARIYANI...	1	1	2	0	1	4	4
TIDAK DISIPLIN	10	ELISABETH N...	2	2	1	2	0	5	7
DISIPLIN	11	ENGGAR SE...	0	0	1	2	1	4	1
TIDAK DISIPLIN	12	EUKHARISTIA...	2	3	1	5	0	1	3
DISIPLIN	13	GEISHA JULIA...	0	0	2	3	0	3	5
DISIPLIN	14	Hieronimus An...	2	3	2	2	1	2	2
DISIPLIN	15	KHOIRUL HID...	2	2	2	4	0	1	0
DISIPLIN	16	MUHAMMAD ...	1	0	3	1	1	3	1

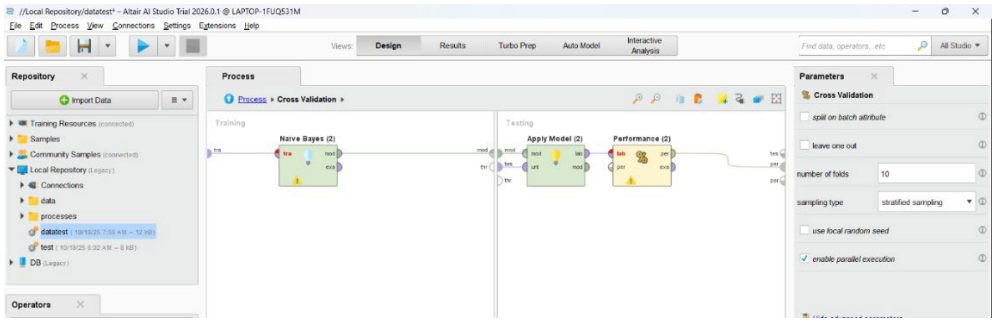
Gambar 5. Pengolahan data dengan tools

4. Pengujian/pemodelan dengan cross validation.
   Untuk mendapatkan hasil evaluasi yang tepat dan menghindari ketidakakuratan, pengujian model dilakukan dengan metode Cross Validation. Dalam penelitian ini, metode cross validation dengan number of fold 10 maksudnya data dibagi menjadi 10 bagian secara sepadan dengan sampling type stratified sampling. Ini memastikan bahwa perbandingan antara kelas Disiplin dan Tidak Disiplin tetap seimbang selama setiap tahap pelatihan dan pengujian. Dalam proses Cross Validation, bagian Pelatihan digunakan untuk membangun model menggunakan algoritma Naive Bayes. Di sisi lain, bagian dalam cross validation digunakan untuk menguji kinerja model dengan menggunakan operator Apply Model dan Performance (Classification)[16]. Nilai akurasi, presisi, recall, dan F1score diperoleh dari perhitungan rata-rata dari seluruh iterasi validasi silang. Ini membuat hasil yang diperoleh lebih stabil dan memberi gambaran yang lebih baik tentang kemampuan model.



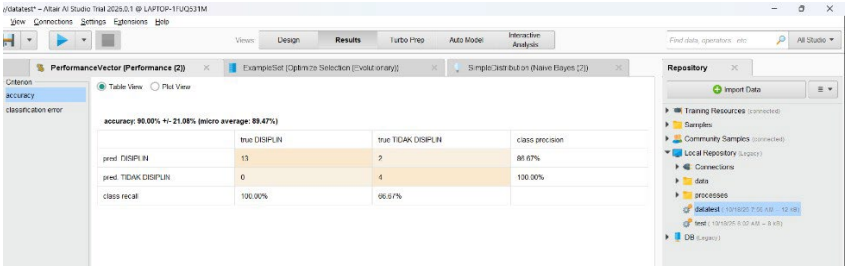
Gambar 6. Pengolahan lanjutan

5. Proses cross validation.
   Ketika kita sudah menghubungkan semua bagian dari read excel-set, role-optimize selection evolutionary, optimize parameters evolutionary sendiri adalah operator yang bekerja didasarkan metode GA (Genetic Algorithm) yaitu algoritma evolusioner dengan menemukan perpaduan atribut yang unggul untuk memperoleh hasil pemodelan yang tinggi[17]. tahap selanjutnya adalah melakukan pengujian model menggunakan Cross Validation, pada operator Cross Validation terdapat dua bagian utama yaitu Subprocess Training dan Subprocess Testing. Pada bagian Training, model pembelajaran dibangun dengan menempatkan algoritma Naive Bayes di dalamnya. Model ini berfungsi untuk mempelajari pola dari data pelatihan yang telah ditentukan oleh proses validasi silang. Sementara itu, pada proses di dalam cross validation digunakan operator Apply Model untuk menerapkan model yang telah terbentuk terhadap data uji, kemudian hasil prediksi tersebut dievaluasi menggunakan operator Performance (Classification) untuk memperoleh nilai akurasi, presisi, recall, dan F1-score.



Gambar 7. *Cross validation*

6. Hasil pengujian
   Confusion matrix berikut diperoleh berdasarkan hasil yang dihasilkan dari tab Hasil RapidMiner



Gambar 8. Hasil pengujian

7. Perbandingan hasil pengujian
   Untuk membuktikan bahwa optimasi genetic algorithm memberikan kontribusi yang efektif penulis menguji pengolahan data menggunakan Naïve bayes dengan optimasi GA dan Naïve bayes tanpa GA. Pada pengujian model tanpa GA menghasilkan accuracy sebesar 84,21% sedangkan jika menggunakan optimasi GA menghasilkan accuracy sebesar 89,47%. Genetic algorithm mampu menyeleksi dan melakukan pencarian parameter terbaik jadi bisa disimpulkan jika hanya menggunakan naïve bayes kurang mampu memberikan klasifikasi data yang optimal. Pada masing masing pengujian ini masih menggunakan number of fold 10 maksudnya data dibagi menjadi 10 bagian agar pembagian 2 label seimbang. Berikut tabel perbandingan dari pengujian naïve bayes tanpa GA dan pengujian naïve bayes dengan optimasi GA.

Table 5. Perbandingan akurasi

Metode	Akurasi (%)	Keterangan
Naïve bayes tanpa Genetic Algorithm	84,21%	Model awal tanpa proses optimasi GA
Naïve bayes dengan optimasi Genetic Algorithm	89,,47%	Model yang dihasilkan setelah optimasi GA, terjadi peningkatan akurasi

8. Analisa pengujian
   Pada model naive bayes mampu mengklasifikasikan disiplin siswa dengan sangat baik (recall 100%), yang berarti bahwa semua siswa yang memiliki disiplin terdeteksi dengan benar. Namun, ada dua siswa yang tidak memiliki disiplin yang diklasifikasikan sebagai disiplin, yang membuat nilai tepatnya menjadi kurang tepat. Secara keseluruhan, model ini sangat tepat, menurut nilai akurasi dengan optimasi GA sebesar 89,47%. Hasil tujuh atribut adalah hasil seleksi terbaik menggunakan optimize selection evolutionary, tujuh atribut yang dipilih untuk mengukur tingkat kedisiplinan siswa dari hasil optimasi GA yaitu absensi, ketepatan hadir, kepatuhan tata tertib, ketertiban seragam, tidak bolos pelajaran, pengumpulan tugas, perilaku di kelas.