Implementasi *Data Mining Naïve Bayes* untuk Sistem Peringatan Dini Hasil Belajar Siswa Sekolah Dasar

Adyatma Ramadhani Sugihrahma[1], Atiqah Meutia Hilda[2*], Muhammad Jafar Elly[3], Nani Mintarsih[4]

[1,2]*Fakultas Teknologi Industri dan Informatika, Universitas Muhammadiyah Prof. Dr. HAMKA, Jakarta, Indonesia*
E-mail:2103015186@uhamka.ac.id, atiqahmeutiahilda@uhamka.ac.id

[3]*Pusat Penelitian Ilmu Data dan Informasi, Badan Riset dan Inovasi Nasional (BRIN), Indonesia*
E-mail:mujaly13@gmail.com

[4]*Fakultas Teknologi dan Sistem Informasi, Universitas Gunadarma, Depok, Indonesia*
E-mail:nanim@staff.gunadarma.ac.id
*Corresponding Author

Abstrak— **Penelitian ini memiliki tujuan untuk menggunakan teknik data mining dan algoritma *Naïve Bayes untuk mengklasifikasikan hasil belajar siswa sebagai sistem peringatan dini. Model klasifikasi dikembangkan menggunakan 90 catatan skor siswa dan dikelompokkan menjadi dua kategori, "Baik" dan "Buruk". Hasil eksperimen menunjukkan bahwa model yang diusulkan mampu secara konsisten mengklasifikasikan hasil belajar siswa dan berpotensi berfungsi sebagai alat pendukung keputusan bagi guru dan administrator sekolah.***

Kata Kunci — ***Data Mining*; Klasifikasi; *Naïve Bayes*; Hasil Belajar Siswa.**

Abstract— **This study aims to apply data mining techniques using the *Naïve Bayes algorithm to classify student learning outcomes as an early warning system. The classification model was developed using 90 student score records and grouped into two categories, namely "Good" and "Poor". The experimental results show that the Naïve Bayes model achieved 100% accuracy, precision, and recall. These findings indicate that the proposed model is capable of consistently classifying student learning outcomes and has the potential to serve as a decision support tool for teachers and school administrators.***

Keywords — **Data Mining, Classification, Naïve Bayes, Student Learning Outcomes.**

I. PENDAHULUAN

Pendidikan di Sekolah Dasar (SD) merupakan jenjang pendidikan formal yang memiliki kontribusi fundamental dalam membangun landasan kompetensi kognitif, afektif, dan psikomotorik peserta didik. Pada jenjang ini, proses pembelajaran tidak hanya berorientasi pada penguasaan pengetahuan dasar, tetapi juga pada pembentukan karakter serta kesiapan siswa melanjutkan ke jenjang pendidikan berikutnya.Oleh karena itu, pemantauan dan evaluasi capaian hasil belajar siswa menjadi aspek krusial dalam menjamin efektivitas proses pembelajaran yang diselenggarakan oleh institusi pendidikan dasar.

SDN Tridayasakti 03 Tambun Selatan masih menghadapi kendala dalam pemanfaatan data historis hasil belajar siswa kelas 6 sebagai dasar evaluasi pembelajaran. Hingga saat ini, proses penilaian bersifat manual dan intuitif, sehingga siswa yang berpotensi mengalami kesulitan belajar sering kali tidak teridentifikasi sejak dini. Kondisi ini berdampak pada keterlampatan pemberian intervensi pembelajaran yang seharusnya dapat dilakukan lebih awal dan tepat waktu.

Merujuk pada regulasi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, struktur kurikulum pada *level* pendidikan dasar dirancang untuk mengembangkan kemampuan literasi dan numerasi, serta memperkuat karakter profil pelajar Pancasila melalui pembelajaran yang bermakna serta berkesinambungan. Dalam konteks ini, capaian akademis pada mata pelajaran fundamental seperti Bahasa Indonesia, Matematika, Ilmu Pengetahuan Alam dan Sosial, serta Pendidikan Pancasila menjadi indikator penting dalam menilai kesiapan siswa untuk melanjutkan pendidikan ke jenjang menengah. Capaian tersebut tidak hanya merefleksikan kompetensi intelektual siswa, tetapi juga mencerminkan efektivitas proses belajar dan mengajar yang berlangsung di sekolah.

Suatu institusi pendidikan dasar yang berada dalam pembinaan Dinas Pendidikan Kabupaten Bekasi, SDN Tridayasakti 03 Tambun Selatan menunjukkan komitmen dalam meningkatkan mutu pembelajaran melalui optimalisasi pemanfaatan data capaian belajar siswa. Namun demikian, masih terbatas kajian ilmiah secara eksplisit mengeksplorasi karakteristik hasil belajar siswa di lingkungan sekolah ini menggunakan pendekatan berbasis data. Kesenjangan riset tersebut membuka peluang untuk menerapkan teknik *data mining* sebagai pendekatan ilmiah dalam mendukung keputusan pendidikan yang lebih objektif dan berbasis bukti empiris[1].

Perkembangan *Educational Data Mining* (EDM) telah menunjukkan efektivitasnya dalam memprediksi capaian akademis peserta didik melalui berbagai pendekatan komputasi. Sejumlah penelitian terdahulu melaporkan bahawa rekam jejak akademis siswa merupakan atribut yang sangat relevan untuk memprediksi performa, dengan *Artificial Neural Network* (ANN) dan *Random Forest* banyak digunakan dalam konteks tersebut [2],[3]. Selain itu, teknik pembelajaran mendalam juga terbukti unggul dalam menganalisis data pendidikan yang kompleks, khususnya pada skenario prediksi performa dan sistem peringatan dini [4],[5], [6].

Dalam ranah pendidikan dasar, pemanfaatan teknik *data mining*, khususnya metode klasifikasi*,* masih belum optimal. Salah satu algoritma klasifikasi yang banyak digunakan adalah *Naïve Bayes,* yang dibangun berdasarkan teorema *Bayes* dan memiliki keunggulan dalam efisiensi komputasi serta kemampuan menghasilkan prediksi yang baik meskipun menggunakan dataset yang relatif terbatas [7]. Berbagai penelitian menunjukkan bahwa lagoritma ini memiliki tingkat akurasi tinggi dalam memprediksi capaian belajar siswa [8], [9], [10], [11], [12], [13].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan model klasifikasi hasil belajar siswa berbasis data mining sebagai sistem peringatan dini. Model ini dirancang untuk membantu guru dan pihak sekolah dalam mengidentifikasi siswa yang memerlukan perhatian khusus secara lebih dini dan objektif. Klasifikasi hasil belajar dilakukan ke dalam dua kategori yaitu ”Baik” dan ”Buruk”, sehingga hasilnya dapat dimanfaatkan sebagai dasar penyusunan strategi pembelajaran yang lebih efektif dan adaptif.

Kebaruan penelitian ini terletak pada penerapan algoritma *Naïve Bayes* dalam kerangka *Knowledge Discovery in Database* (KDD), untuk membangun sistem peringatan dini berbasis data nilai siswa sekolah dasar, khususnya pada siswa kelas 6 sebagai fase transisi menuju pendidikan menengah. Hingga saat ini, kajian sistem peringatan dini berbasis data akademik masih didominasi oleh jenjang pendidikan menengah dan perguruan tinggi, sehingga penelitian ini memberikan kontribusi empiris pada konteks pendidikan dasar.

II. METODE PENELITIAN

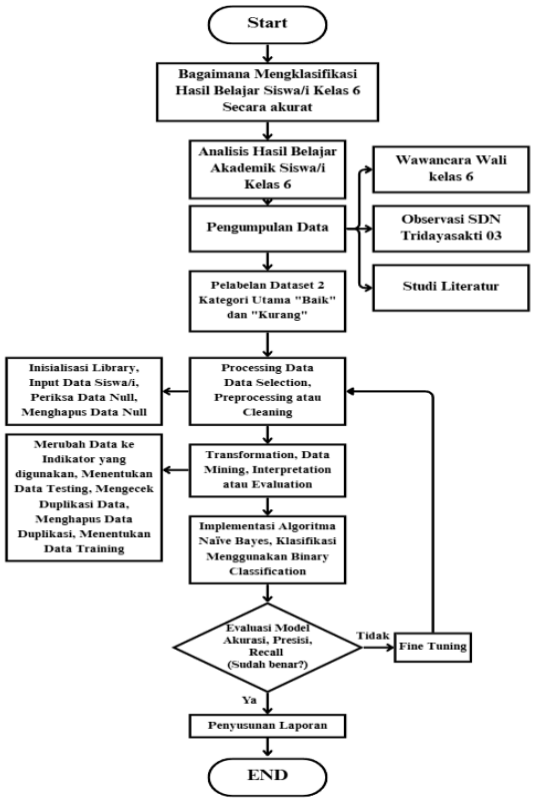
Pendekatan *Knowledge Discovery in Database* (KDD) dalam penelitian ini sebagai kerangka metodologis untuk menggali pola dan informasi bermakna dari data hasil belajar siswa. Pendekatan KDD dipilih karena mampu mengelola data secara sistematis mulai dari tahap seleksi data hingga evaluasi hasil, sehingga menghasilkan model prediksi yang dapat mendukung pengambilan keputusan berbasis data. Algoritma Naïve Bayes diimplementasikan pada tahap data mining sebagai metode klasifikasi hasil belajar siswa.

Gambar 1 menunjukkan diagram konseptual alur penelitian yang menggambarkan keterkaitan antara data nilai siswa, tahapan KDD, penerapan algoritma Naïve Bayes, serta hasil klasifikasi yang digunakan sebagai sistem peringatan dini. Proses penelitian diawali dengan pemelilihan dan pembersihan data, kemudian dilanjutkan dengan tahap transformasi dan pemodelan data. Pendekatan ini tidak hanya berfokus pada pengumpulan data historis, tetapi juga pada analisis komprehensif untuk menemukan pola dan hubungan bermakna yang mampu memberikan pemahaman baru.

Dalam pengelolaan data berskala besar, pendekatan KDD memiliki peran strategis karena memungkinkan proses identifikasi pola dan kecenderungan yang tidak mudah ditemukan melalui analisi data konvensional. Pendekatan ini mengintegrasikan berbagai metode analitis, termasuk teknik data mining, analisis statistik, serta algoritma pembelajaran mesin, guna meningkatkan akurasi dan efektivitas dalam mengekstraksi informasi yang bernilai bagi pengambilan keputusan. Melalui penerapan KDD penelitian ini membangun model prediksi yang mendukung proses pengambilan keputusan secara lebih tepat dan objektif. Dengan demikian, KDD tidak hanya berfungsi sebagai alat

https://doi.org/10.35760/ik.2025.v30i3.58

untuk mengidentifikasi pola dalam data, tetapi juga sebagai landasan inovasi dan perumusan strategi yang lebih adaptif di berbagai domain [14].



Gambar 1. Diagram Konseptual Alur Penelitian

Analisis Masalah

Tahap awal penelitian difokuskan pada penelaahan permasalahan yang berkaitan pada capaian hasil belajar siswa kelas 6 di SDN Tridayasakti 03 Tambun Selatan. Hasil analisis menunjukkan bahwa pemanfaatan data nilai siswa sebagai dasar evaluasi pembelajaran dan mekanisme deteksi dini masih belum dilakukan secara optimal. Akibatnya, siswa yang berpotensi mengalami kendala dalam proses belajar tidak selalu teridentifikasi sejak tahap awal. Oleh karena itu, sangat diperlukan suatu pendekatan analisis data yang mampu membantu pihak sekolah dalam mengenali pola hasil belajar siswa secra lebih objektif dan sistematis.

Pengumpulan Data

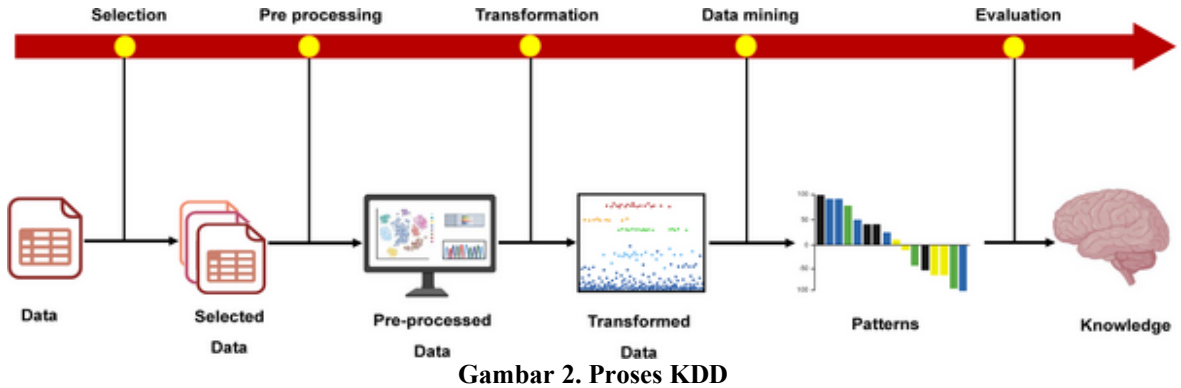
Pengumpulan data dilakukan sebagai proses yang bertujuan untuk memastikan bahwa data yang dikumpulkan relevan serta dapat mendukung analisis dan pemodelan yang akan dilakukan. Teknik pengumpulan data dalam penelitian ini meliputi:

a. Observasi, yaitu secara langsung meninjau proses pembelajaran dan mencermati sistem penilaian yang diterapkan di SDN Tridayasakti 03 Tambun Selatan.
b. Wawancara, dilakukan dengan wali kelas untuk memperoleh data nilai siswa serta informasi terkait proses evaluasi hasil belajar.
c. Studi Literatur, dilakukan dengan tujuan untuk memperoleh landasan teoritis serta gambaran penelitian terdahulu yang mendukung pengembangan dan analisis dalam penelitian ini.

Pelabelan *Dataset*

Dataset merupakan data nilai siswa kelas 6 yang kemudian diberi label untuk keperluan pembelajaran terawasi (*supervised learning*). Pelabelan dilakukan dengan mengelompokkan data ke dalam dua kategori, yaitu ”Baik” dan ”Buruk”, berdasarkan kriteria penilaian yang ditetapkan oleh pihak

sekolah. Proses pelabelan ini bertujuan agar algoritma *Naïve Bayes* dapat melakukan klasifikasi hasil belajar siswa secara optimal dan menghasilkan informasi yang dapat ditindaklanjuti oleh pihak sekolah.



Gambar 2. Proses KDD

Tahapan *Knowledge Discovery In Database (KDD)*

Tahapan ini dipergunakan secara bergantian dengan tujuan menjelaskan proses pemahaman mendalam informasi yang belum teridentifikasi dalam basis data berkapasitas besar[14]. Proses pengolahan dataset dalam penelitian ini mengikuti tahapan KDD pada Gambar 2

Proses Knowledge Discovery in Database (KDD) dalam penelitian ini dapat dijabarkan melalui beberapa tahapan utama yaitu:

a. *Data selection*
   Tahap ini bertujuan memilih atribut nilai siswa kelas 6 yang relevan untuk dianalisis. Data yang digunakan berasal dari periode Juli 2024 hingga Desember 2024 serta Januari 2025 hingga Juni 2025. Pemilihan Siswa kelas 6 dilakukan karena jenjang ini merupakan fase akhir pendidikan dasar yang capaian akademiknya menjadi indikator kesiapan menuju pendidikan menengah.
b. *Preprocessing* (*Cleaning)*
   Proses *Preprocessing* dilakukan melalui pemeriksaan konsistensi dan kelengkapan data. Tahapan ini mencakup identifikasi dan duplikat, pengecekan inkonsistensi, serta perbaikan kesalahan penulisan. Hasil pemeriksaan menunjukkan bahwa seluruh data nilai siswa telah memenuhi kriteria kelengkapan, sehingga seluruh 90 data digunakan pada tahap analisis selanjutnya.
c. *Transformation*
   Pada tahap ini dilakukan penyesuaian format data agar sesuai dengan kebutuhan proses penambahan data. Dalam penelitian ini, tidak dilakukan transformasi tambahan karena struktur data yang diperoleh telah sesuai untuk proses klasifikasi menggunakan algoritma *Naïve Bayes*.
d. *Data Mining*
   Tahap data mining dilakukan dengan menerapkan algoritma *Naïve Bayes* untuk mengklasifikasikan hasil belajar siswa berdasarkan data nilai yang telah diproses.
e. *Interpretation* atau *evaluation*
   Hasil klasifikasi kemudian dianalisis dan dievaluasi untuk menilai kinerja model serta kesesuaian hasil dengan tujuan penelitian.

Implementasi Algoritma *Naïve Bayes*

Implementasi algoritma *Naïve Bayes* dilakukan menggunakan perangkat lunak RapidMiner. Dataset dalam penelitian ini dibagi menjadi 80% data latih dan 20% data uji. dibagi menjadi 80% data latih dan 20% data uji. Pembagian ini bertujuan untuk menguji kemampuan model dalam mengklasifikasikan data yang belum pernah digunakan pada proses pelatihan. Algoritma *Naïve Bayes* digunakan untuk melakukan klasifikasi biner, yaitu mengelompokkan siswa ke dalam kategori ”Baik” dan ”Buruk”.

Evaluasi Model

Evaluasi kinerja model dilakukan dengan memanfaatkan metrik akurasi, presisi, dan *recall*. Ketiga metrik tersebut digunakan untuk memberi suatu penilaian secara menyeluruh terhadap performa

https://doi.org/10.35760/ik.2025.v30i3.58

model klasifikasi yang dikembangkan. Akurasi digunakan untuk merepresentasikan tingkat ketepatan prediksi model secara umum, sementara presisi dan recall berfungsi untuk menilai kemampuan model dalam mengklasifikasikan masing-masing kelas secara lebih detail dan spesifik.

Fine Tuning Model

Tahap fine-tuning dilakukan apabila hasil evaluasi menunjukkan performa model yang belum optimal. *Fine-tuning* dapat mencakup penyesuaian pada tahap preprocessing, seperti seleksi fitur dan penanganan noise, maupun penyesuaian parameter algoritma *Naïve Bayes.* Tujuan dari tahap ini adalah untuk meningkatkan kinerja model serta mengurangi risiko overfitting [16], sehingga model yang dihasilkan memiliki tingkat akurasi yang lebih tinggi serta tingkat keandalan yang lebih baik dalam proses klasifikasi.

III. HASIL DAN PEMBAHASAN

Data

Data yang digunakan pada penelitian ini merupakan data primer hasil observasi di SDN Tridayasakti 03 Kecamatan Tambun Selatan. Data berupa nilai siswa kelas 6 yang dikumpulkan selama dua semester (semester 1 dan semester 2) dengan total 90 data siswa. Atribut yang digunakan meliputi nilai harian, nilai PTS, nilai PAS, dan nilai rapor, karena atribut tersebut merepresentasikan performa akademik siswa secara komprehensif selama satu tahun ajaran.

Proses preprocessing dilakukan melalui pemeriksaan konsistensi dan kelengkapan data. Hasil pemeriksaan menunjukkan bahawa seluruh data memenuhi kriteria kelengkapan, sehingga seluruh 90 data digunakan pada tahap analisis selanjutnya.

Data Selection

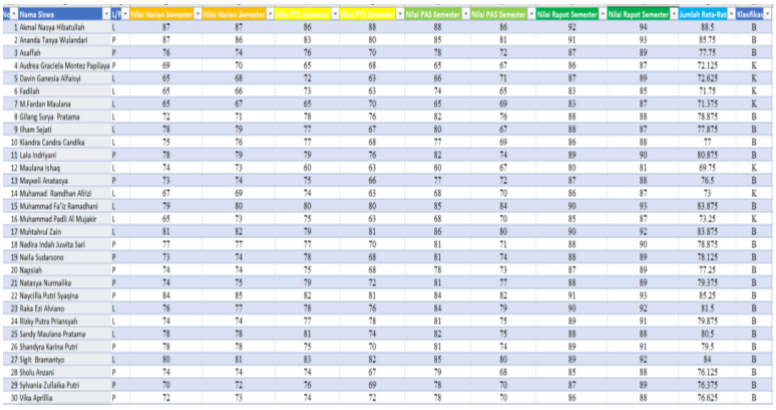
Tahap data selection bertujuan memastikan bahawa atribut yang digunakan relevan dengan tujuan klasifikasi hasil belajar siswa. Pada penelitian ini, pemilihan atribut difokuskan pada nilai yang menggambarkan capaian belajar siswa sepanjang tahun ajaran, yaitu PTS, PAS, dan rapor. Fokus penelitian pada siswa kelas 6 dipilih karena jenjang ini merupakan fase akhir pendidikan dasar yang menjadi indikator kesiapan siswa menuju pendidikan menengah[16]. Ilustrasi kondisi data awal sebelum diproses ditunjukkan pada Gambar 3.

Preprocessing

Pada tahapan *preprocessing* dilakukan dalam rangka meningkatkan kualitas data sebelum dimasukan kedalam pemodelan. Proses yang dilakukan meliputi pengecekan duplikasi, koreksi kesalahan penulisan, dan pemeriksaan konsistensi format data. Tahap ini bertujuan untuk memastikan data yang digunakan akurat dan siap diproses oleh algoritma klasifikasi [14]. Hasil data setelah preprocessing ditunjukkan pada Gambar 4.

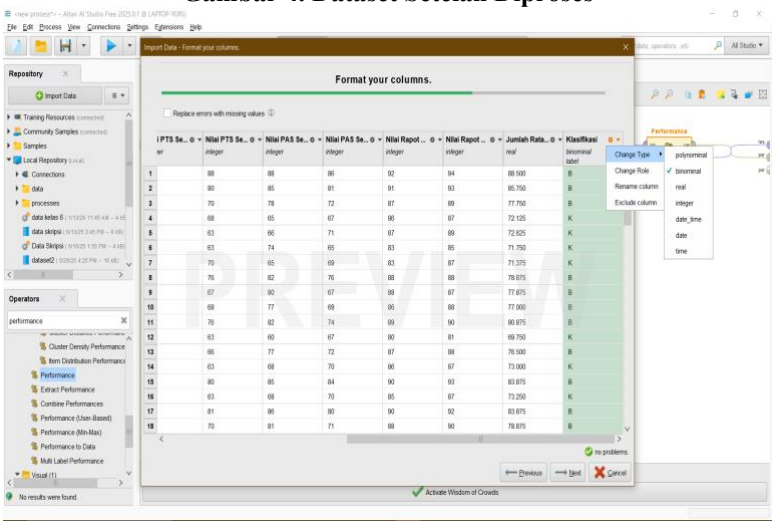
No	Nama Siswa	L/P	Nilai Harian Semester	Nilai Harian Semester	Nilai PTS Semester	Nilai PTS Semester	Nilai PAS Semester	Nilai PAS Semester	Nilai Raport Semester	Nilai Raport Semester	Jumlah Rata-Rat
1	Akmal Nasya Hibatullah	L	87	87	86	88	88	86	92	94	88.5
2	Ananda Tasya Wulandari	P	87	86	83	80	85	81	91	93	85.75
3	Asaffah	P	76	74	76	70	78	72	87	89	77.75
4	Audrea Graciela Montez Papilaya	P	69	70	65	68	65	67	86	87	72.125
5	Davin Ganesia Alfaisyi	L	65	68	72	63	66	71	87	89	72.625
6	Fadilah	L	65	66	73	63	74	65	83	85	71.75
7	M.Fardan Maulana	L	65	67	65	70	65	69	83	87	71.375
8	Gilang Surya Pratama	L	72	71	78	76	82	76	88	88	78.875
9	Ilham Sejati	L	78	79	77	67	80	67	88	87	77.875
10	Kiandra Candra Candika	L	75	76	77	68	77	69	86	88	77
11	Lala Indriyani	P	78	79	79	76	82	74	89	90	80.875
12	Maulana Ishaq	L	74	73	60	63	60	67	80	81	69.75
13	Mayveli Anatasya	P	73	74	75	66	77	72	87	88	76.5
14	Muhamad Ramdhan Afrizi	L	67	69	74	63	68	70	86	87	73
15	Muhammad Fa'iz Ramadhani	L	79	80	80	80	85	84	90	93	83.875
16	Muhammad Padli Al Mujakir	L	65	73	75	63	68	70	85	87	73.25
17	Muhtahrul Zain	L	81	82	79	81	86	80	90	92	83.875
18	Nadira Indah Juwita Sari	P	77	77	77	70	81	71	88	90	78.875
19	Naila Sudarsono	P	73	74	78	68	81	74	88	89	78.125
20	Napsiah	P	74	74	75	68	78	73	87	89	77.25
21	Natasya Nurmalika	P	74	75	79	72	81	77	88	89	79.375
22	Naycilla Putri Syaqina	P	84	85	82	81	84	82	91	93	85.25
23	Raka Ezi Alviano	L	76	77	78	76	84	79	90	92	81.5
24	Rizky Putra Prianyah	L	74	74	77	78	81	75	89	91	79.875
25	Sandy Maulana Pratama	L	78	78	81	74	82	75	88	88	80.5
26	Shandyra Karina Putri	P	78	78	75	70	81	74	89	91	79.5
27	Sigit Bramantyo	L	80	81	83	82	85	80	89	92	84
28	Sholu Anzani	P	74	74	74	67	79	68	85	88	76.125
29	Sylvania Zullaika Putri	P	70	72	76	69	78	70	87	89	76.375
30	Vika Aprillia	P	72	73	74	72	78	70	86	88	76.625

Gambar 3. Data Sebelum Diproses



No	Nama Siswa	L/P	Nilai Harian Semester	Nilai Harian Semester	Nilai PTS Semester	Nilai PTS Semester	Nilai PAS Semester	Nilai PAS Semester	Nilai Raport Semester	Nilai Raport Semester	Jumlah Rata-Rata	Klasifikasi
1	Akmal Nasya Hibatullah	L	87	87	86	88	88	86	92	94	88.5	B
2	Ananda Tasya Wulandari	P	87	86	83	80	85	81	91	93	85.75	B
3	Asaffah	P	76	74	76	70	78	72	87	89	77.75	B
4	Audrea Graciela Montez Papilaya	P	69	70	65	68	65	67	86	87	72.125	K
5	Davin Ganesia Alfaisyl	L	65	68	72	63	66	71	87	89	72.625	K
6	Fadilah	L	65	66	73	63	74	65	83	85	71.75	K
7	M.Fardan Maulana	L	65	67	65	70	65	69	83	87	71.375	K
8	Gilang Surya Pratama	L	72	71	78	76	82	76	88	88	78.875	B
9	Ilham Sejati	L	78	79	77	67	80	67	88	87	77.875	B
10	Kiandra Candra Candika	L	75	76	77	68	77	69	86	88	77	B
11	Lala Indriyani	P	78	79	79	76	82	74	89	90	80.875	B
12	Maulana Ishaq	L	74	73	60	63	60	67	80	81	69.75	K
13	Maxwell Anataya	P	73	74	75	66	77	72	87	88	76.5	B
14	Muhamad Ramdhan Afrizi	L	67	69	74	63	68	70	86	87	73	K
15	Muhammad Fa'iz Ramadhani	L	79	80	80	80	85	84	90	93	83.875	B
16	Muhammad Padli Al Mujakir	L	65	73	75	63	68	70	85	87	73.25	K
17	Muhtahrul Zain	L	81	82	79	81	86	80	90	92	83.875	B
18	Nadira Indah Juwita Sari	P	77	77	77	70	81	71	88	90	78.875	B
19	Naila Sudarsono	P	73	74	78	68	81	74	88	89	78.125	B
20	Napsiah	P	74	74	75	68	78	73	87	89	77.25	B
21	Natasya Nurmalika	P	74	75	79	72	81	77	88	89	79.375	B
22	Naycilla Putri Syaqina	P	84	85	82	81	84	82	91	93	85.25	B
23	Raka Ezi Alviano	L	76	77	78	76	84	79	90	92	81.5	B
24	Rizky Putra Priansyah	L	74	74	77	78	81	75	89	91	79.875	B
25	Sandy Maulana Pratama	L	78	78	81	74	82	75	88	88	80.5	B
26	Shandyra Karina Putri	P	78	78	75	70	81	74	89	91	79.5	B
27	Sigit Bramantyo	L	80	81	83	82	85	80	89	92	84	B
28	Sholu Anzani	P	74	74	74	67	79	68	85	88	76.125	B
29	Sylvania Zullaika Putri	P	70	72	76	69	78	70	87	89	76.375	B
30	Vika Aprilia	P	72	73	74	72	78	70	86	88	76.625	B

Gambar 4. Dataset Setelah Diproses



Gambar 5. Tahap Transformasi

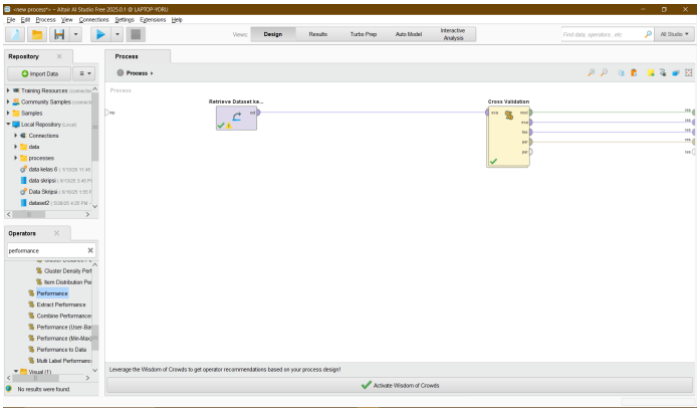
Transformasi

Pada tahap transformasi, dilakukan penyesuaian tipe atribut agar sesuai dengan kebutuhan pemodelan pada RapidMiner. Perubahan yang dilakukan meliputi mengubah atribut klasifikasi dari polynomial menjadi binominal, serta mengganti role atribut target menjadi label, sehingga dataset siap dipergunakan untuk proses pengklasifikasian dengan menggunakan *Naïve Bayes*[17]. Proses transformasi dataset ditunjukkan pada Gambar 5.

Data Mining

Tahap *Data Mining* dilakukan menggunakan perangkat lunak *RapidMiner* untuk membangun model klasifikasi hasil belajar siswa menggunakan algoritma *Naïve Bayes*. Alur pemodelan meliputi:

a. *Retrieve* data nilai siswa-siswi
   Retrieve merupakan proses seleksi model pembelajaran dari beberapa model yang telah dilatih sebelumnya dengan mempertimbangkan kriteria tertentu. Berbagai faktor memengaruhi seleksi model ini, seperti jenis permasalahan yang hendak dipecahkan, ukuran serta kualitas data yang ada, ketersediaan sumber daya komputasi, dan tingkat akurasi prediksi yang direncanakan.
b. *Cross Validation*
   Cross Validation merupakan metode validasi yang dilaksanakan secara berulang, dimana *dataset* dipartisi menjadi beberapa *subset* yang terdiri dari dua data yaitu, data latih dan data uji. Pada setiap iterasi, satu *subset* diambil untuk data uji, sementara *subset* lainnya dipergunakan untuk melatih model. Dengan demikian, proses validasi berlangsung secara komprehensif dan bergantian.



Gambar 6. Proses Validasi

Naïve Bayes

Naïve Bayes digunakan untuk mengklasifikasikan hasil belajar siswa berdasarkan prinsip probabilitas yang dikembangkan oleh *Thomas Bayes*. Metode ini bekerja dengan asumsi bahwa penggunaan atribut pada data memiliki sifat independen, meskipun pada kenyataannya hal tersebut tidak selalu sepenuhnya terpenuhi[18]. Hasil penelitian menunjukkan bahwa *Naïve Bayes* tetap mampu memberikan kinerja optimal dengan akurasi, presisi, dan *recall* mencapai 100%. Dengan demikian, penerapan algoritma ini tidak hanya berfungsi untuk klasifikasi data, tetapi juga dapat menjadi dasar pengembangan sistem peringatan dini guna membantu sekolah dalam mendeteksi siswa yang berpotensi mengalami kesulitan.

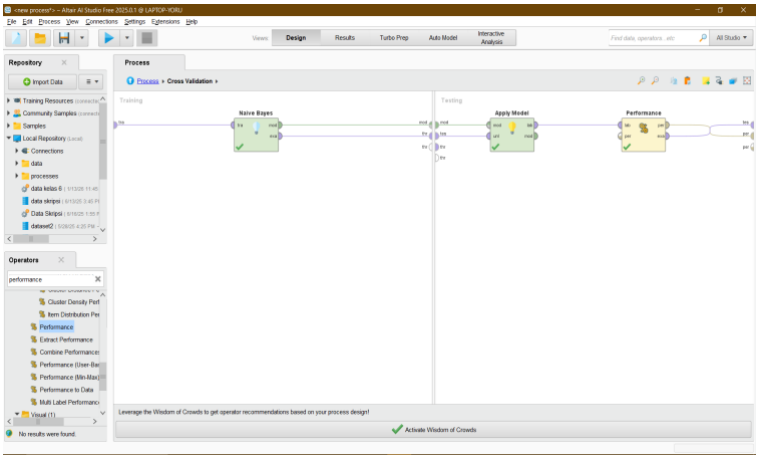
Apply Model

Apply Model dilakukan untuk memberikan kemudahan dalam mengimplementasikan model yang telah dilatih pada *dataset* pengujian atau data baru, sehingga hasil analisis dapat diperoleh dengan cepat dan efisien. Ini juga memungkinkan evaluasi performa model secara *real-time*, sehingga pengguna dapat mengevaluasi seberapa baik model tersebut mampu menggeneralisasi data yang belum dikenal[17]. *Apply Model* memberikan fleksibilitas dalam pengolahan data karena pengguna dapat memilih untuk mengimplementasikan model pada *subset* tertentu dari data atau melakukan prediksi dalam skala besar.

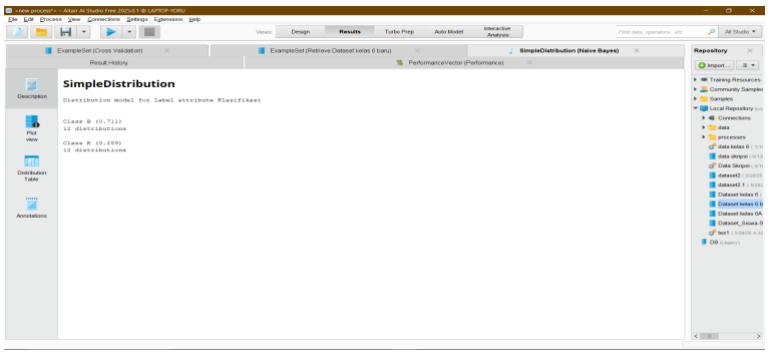
Performa

Performa yang dilakukan berfungsi untuk mengukur seberapa efektif atau berkualitas kinerja dari model prediksi yang telah dikembangkan, baik itu model klasifikasi, regresi, maupun *clustering*. Dalam riset ini, digunakan model klasifikasi yang dievaluasi berdasarkan metrik seperti presisi, *recall* dan akurasinya untuk menilai seberapa baik model dalam melakukan prediksi. Akurasi memberikan gambaran umum mengenai seberapa sering model menghasilkan prediksi yang tepat. Namun, metrik tersebut tidak selamanya mencerminkan performa model secara menyeluruh, terutama ketika terdapat ketidakseimbangan kelas. Oleh karena itu, metrik tambahan seperti presisi dan *recall* menjadi sangat penting. Presisi mengukur proporsi prediksi positif yang akurat, sedangkan *recall* bertujuan mengukur seberapa banyak dari total kasus positif yang berhasil diidentifikasi oleh model. Dengan memanfaatkan kombinasi metrik tersebut, peneliti bisa mendapatkan pemahaman yang lebih komprehensif tentang kinerja model[15].

Evaluasi kinerja model tidak hanya sebatas pengukuran metrik, tetapi juga melibatkan analisis lebih lanjut terhadap hasil yang diperoleh. Sebagai contoh, dengan menggunakan *confusion matrix*, peneliti dapat melihat dengan jelas jumlah perkiraan yang tepat dan keliru untuk masing-masing kelas, yang dapat membantu mengidentifikasi area di mana model perlu disempurnakan. Analisis ini juga dapat memberikan wawasan tentang kemungkinan bias yang ada dalam model serta membantu dalam pengambilan keputusan terkait perbaikan model atau pemilihan algoritma yang lebih sesuai. Dengan demikian, proses evaluasi kinerja melalui operator *Performa* menjadi tahapan yang sangat krusial dalam pengembangan model *machine learning*. Proses ini tidak hanya memastikan bahwa model berfungsi dengan optimal, tetapi juga memberikan fondasi yang lebih baik untuk pengambilan keputusan yang tepat dalam aplikasi yang lebih praktis.



Gambar 7. Proses Performa



Gambar 8. *Simple Distribution* Pada *Rapidminer*

Evaluation

Setelah model *Naïve Bayes* dibuat, salah satu hal yang penting dalam analisis *data mining* adalah mengetahui seberapa besar probabilitas suatu atribut terhadap kelasnya, yang dikenal sebagai *prior probability*. Dalam konteks penelitian ini, pemahaman tentang *prior probability* sangat krusial karena dapat memberikan wawasan tentang distribusi kelas dalam *dataset* yang digunakan. Dengan mengetahui probabilitas masing-masing kelas, peneliti dapat lebih memahami karakteristik data dan bagaimana model akan berperilaku saat melakukan prediksi[18]. Pada penelitian ini, distribusi kelas diperoleh menggunakan operator *Simple Distribution* di *RapidMiner*, yang didasarkan pada hasil pemodelan dengan algoritma *Naïve Bayes*. Hasil yang diperoleh menunjukkan bahwa kelas "B" (Baik) memiliki probabilitas sebesar 0.711, sedangkan kelas "K" (Buruk) sebesar 0.289. Nilai distribusi tersebut dihasilkan dari keseluruhan data sebanyak 90, yang memberikan gambaran yang jelas tentang proporsi masing-masing kelas dalam dataset.

Analisis terhadap *prior probability* ini juga dapat membantu dalam pengambilan keputusan yang lebih baik dalam konteks aplikasi model. Misalnya, dengan probabilitas kelas "B" yang jauh lebih tinggi dibandingkan dengan kelas "K", peneliti dapat menyimpulkan bahwa data cenderung lebih banyak mengandung contoh positif. Hal ini dapat mempengaruhi strategi dalam pengumpulan data lebih lanjut, di mana peneliti mungkin perlu mencari lebih banyak contoh dari kelas "K" untuk mencapai keseimbangan yang lebih baik. Selain itu, pemahaman tentang *prior probability* juga dapat digunakan untuk menyesuaikan *threshold* dalam pengambilan keputusan model, sehingga dapat meningkatkan kinerja model dalam situasi di mana biaya kesalahan prediksi berbeda untuk setiap kelas. Dengan demikian, analisis *prior probability* tidak hanya memberikan informasi statistik, tetapi juga berfungsi sebagai alat strategis dalam pengembangan dan penerapan model *machine learning* yang lebih efektif. Berikut gambar 8 dari *simple distribution* pada *RapidMiner*. Merujuk pada Gambar 8, perhitungan distribusi nilai kelas secara manual dapat dijelaskan melalui rumus *Naïve Bayes* berikut ini;

$P(H \mid X)=((P(X \mid H)xP(H))/(P(H))$

Dimana

P (Klasifikasi | Baik) = 64/90 = 0, 711
P (Klasifikasi | Buruk) = 26/90 = 0, 289

Setelah mengetahui distribusi probabilitas dari masing-masing kelas, langkah selanjutnya adalah mengevaluasi performa model *Naïve Bayes* dengan menghitung nilai *accuracy*, *precision*, dan *recall* berdasarkan *confusion matrix* yang dihasilkan. Tabel 1 adalah Struktur Tabulasi *Confusion Matrix.*

Berikut adalah penjelasan terkait hasil perhitungan *accuracy*, *precision*, dan *recall* yang dilakukan secara manual maupun menggunakan bantuan *tools RapidMiner*:

1. *Accuracy* menggambarkan seberapa dekat hasil prediksi terhadap jumlah data sebenarnya. Dari hasil klasifikasi yang dilakukan, diperoleh nilai *accuracy* sebesar 100%, yang menandakan bahwa seluruh data telah terklasifikasi dengan tepat oleh model. Perhitungan *accuracy* ditunjukkan pada rumus berikut:

   Accuracy = (TP+TN)/(TP+TN+FP+FN) x 100%
   = (64+26)/(64+26 +0+0) x 100% =
   = 90/90 = 1 x 100% = 100 %

Hasil perhitungan *accuracy* di *RapidMiner* dapat ditampilkan pada Gambar 9. Hasil Gambar 9 memperlihatkan bahwa dari 90 data siswa diklasifikasikan yang masuk dalam kategorik ’B’ sebanyak 64 siswa.

2. *Precision*

   Precision merupakan rasio antara jumlah data yang sesungguhnya termasuk dalam kategori 'B' dengan keseluruhan data yang diprediksi sebagai 'baik', serta jumlah data yang sesungguhnya termasuk dalam kategori 'K' dibandingkan dengan keseluruhan data yang diprediksi sebagai 'Buruk'. Berdasarkan hasil pengujian, diperoleh nilai *precision* sebesar 100%. Perhitungan nilai *precision* dapat dilihat pada rumus berikut:

   Precision = TP/(TP+FP) x 100% =
   = 64/(64+0) x 100% =
   = 1 x 100% = 100 %

Hasil *precision* pada *RapidMiner* dapat dilihat pada Gambar 10.

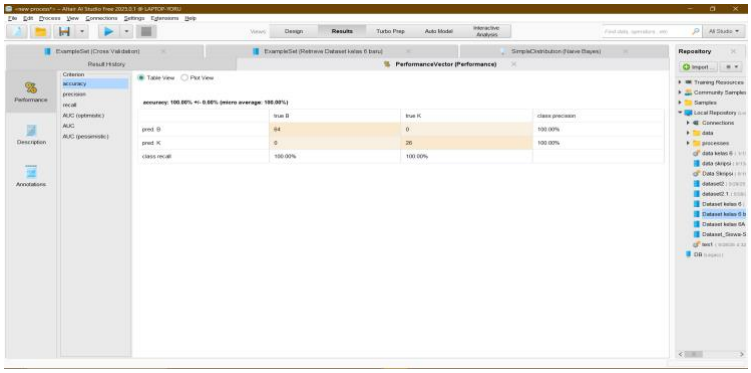
3. *Recall*

   Recall merupakan rasio antara jumlah data yang sesungguhnya baik dengan keseluruhan data yang seharusnya baik, sementara *true* Buruk adalah rasio antara jumlah data yang sesungguhnya Buruk dengan keseluruhan data yang seharusnya Buruk. Nilai *recall* dalam hal ini adalah 100%. Perhitungan nilai *recall* disajikan dalam rumus berikut:

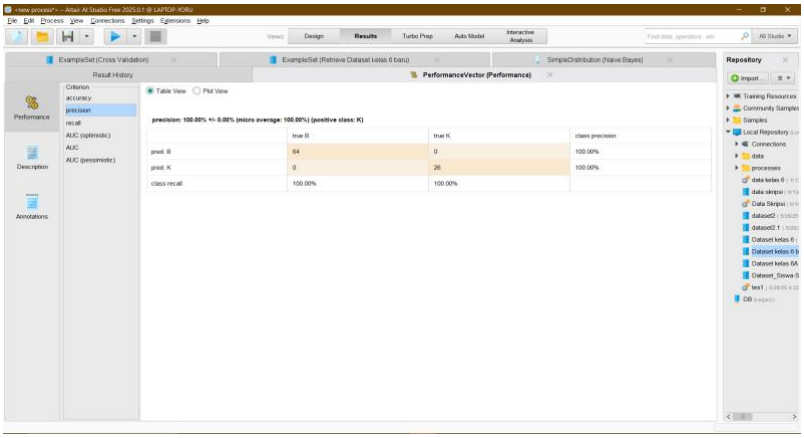
   Recall = TP/(TP+FN) x 100% =
   = 64/(64+0) x 100% =
   = 1 x 100% = 100%

Table 1. Struktur Tabulasi *Confusion Matrix*

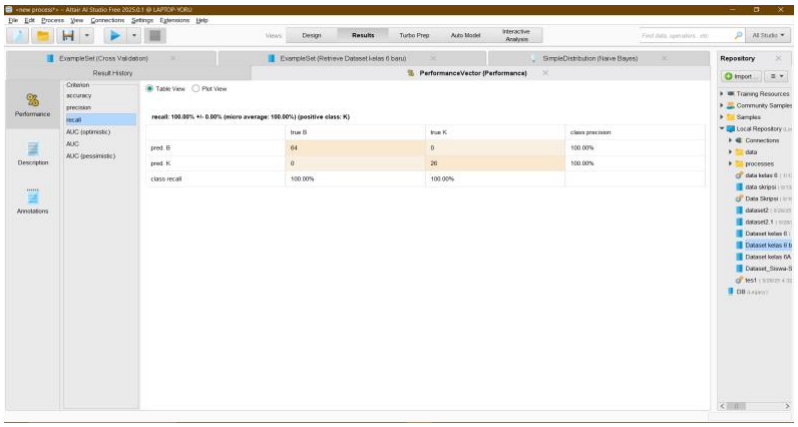
Confusion Matrix		**Kelas Hasil Prediksi (*Predicted)***	
		Baik (B) (+)	**Buruk (K) (-)**
Kelas Asli (*Actual*)	**Baik (B) (+)**	64 (TP)	0 (FN)
	Buruk (K) (-)	0 (FP)	26 (TN)



Gambar 9. Hasil Perhitungan *Accuracy* di *RapidMiner*



Gambar 10. Hasil *Precision* pada *Rapidminer*



Gambar 11. Hasil *Recall* pada Aplikasi *RapidMiner*

Hasil *recall* pada aplikasi *RapidMiner* ditunjukkan pada Gambar 11. Hasil riset ini menunjukkan konsistensi dengan berbagai studi empiris terkini dalam ranah *Educational Data Mining* yang mengimplementasikan algoritma klasifikasi untuk prediksi performa akademik. Temuan akurasi 100% yang dicapai oleh model *Naïve Bayes* dalam riset ini sejalan dengan hasil yang mengimplementasikan algoritma serupa pada siswa sekolah menengah di Arab Saudi dan memperoleh akurasi prediksi mencapai 99,34% [16]. Keberhasilan ini mengindikasikan bahwa algoritma *Naïve Bayes* memiliki kapabilitas *robust* dalam menangani *dataset* pendidikan dengan karakteristik yang beragam, baik dari aspek konteks geografis maupun jenjang pendidikan [19]. Dalam konteks *Knowledge Discovery in Databases* (KDD), pendekatan yang diterapkan dalam riset ini merefleksikan prinsip yang diuraikan, di mana proses *data mining* secara otomatis mempertimbangkan sejumlah besar prediktor untuk mengungkap pola tersembunyi yang valid dalam data dan meningkatkan model *goodness of fit*[18]. Implementasi tahapan KDD yang sistematis—mulai dari *data selection*, *preprocessing*, *transformation*,

hingga *evaluation*—memastikan bahwa model yang dihasilkan tidak hanya akurat namun juga *reliable* untuk pengambilan keputusan berbasis bukti empiris dalam konteks pendidikan.

Meskipun riset ini mencapai performa optimal dengan *dataset* yang relatif seimbang (distribusi kelas 71,1% untuk "Baik" dan 28,9% untuk "Buruk"), penting untuk mempertimbangkan implikasi ketidakseimbangan kelas yang mungkin terjadi pada *dataset* dengan karakteristik berbeda[17]. Permasalahan ketidakseimbangan kelas dalam *dataset* pendidikan dapat menghambat akurasi model prediktif karena sebagian besar model dirancang dengan asumsi kelas yang seimbang, dan merekomendasikan penggunaan teknik *random oversampling* untuk data yang *moderately imbalanced* atau *hybrid resampling* untuk data yang *extremely imbalanced* [10]. Temuan mengidentifikasi bahwa teknik *Deep Learning* dalam EDM telah menunjukkan superioritas signifikan dalam mengatasi kompleksitas dan keberagaman data akademik, khususnya dalam *performance prediction* yang menjadi salah satu dari empat skenario pendidikan tipikal [15]. Lebih lanjut, perspektif *interdisipliner* mengilustrasikan bahwa sinergi antara *database mining* dan *machine learning* dalam *knowledge discovery* telah merevolusi riset di berbagai *domain*, di mana temuan dari ekstraksi data dapat menjadi *pivotal* untuk pengambilan keputusan, stratifikasi risiko, dan upaya peningkatan kualitas layanan [20]. Dalam konteks yang lebih luas, ditekankan bahwa implementasi *data mining* dan *business intelligence* tidak hanya meningkatkan efisiensi operasional namun juga memfasilitasi pengambilan keputusan strategis berbasis data, meskipun tantangan seperti kompleksitas teknologi dan kesenjangan keterampilan perlu diatasi melalui program edukasi yang tertarget dan pengembangan solusi yang *user-friendly* [21]. Implikasi dari temuan-temuan tersebut menegaskan bahwa model klasifikasi yang dikembangkan dalam riset ini memiliki potensi besar untuk diintegrasikan ke dalam sistem informasi sekolah sebagai instrumen prediktif yang mendukung intervensi dini terhadap siswa yang berisiko mengalami kesulitan belajar[22].

Meskipun model *Naïve Bayes* dalam penelitian ini mencapai nilai akurasi, presisi, dan recall sebesar 100%, hasil tersebut perlu ditafsirkan secara hati-hati. Hal ini disebabkan oleh ukuran dataset yang relatif terbatas serta karakteristik data yang cenderung homogen, sehingga berpotensi menimbulkan overfitting. Oleh karena itu, untuk meningkatkan kemampuan generalisasi model, penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan beragam, serta membandingkan performa *Naïve Bayes* dengan algoritma klasifikasi lainnya.

IV. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa algoritma *Naïve Bayes* terbukti sangat efektif dalam mengklasifikasikan capaian belajar siswa kelas 6 di SDN Tridayasakti 03 Tambun Selatan. Model ini berhasil mengelompokkan 90 data nilai siswa ke dalam kategori "Baik" dan "Buruk" dengan akurasi, presisi, dan *recall* yang mencapai 100%. Hal ini jelas menunjukkan bahwa *Naïve Bayes* merupakan instrumen yang *reliable* untuk mendeteksi pola belajar siswa, sehingga sangat bermanfaat untuk evaluasi dan perencanaan strategi pembelajaran di sekolah. Keberhasilan ini menggarisbawahi pentingnya teknologi dalam pendidikan untuk meningkatkan kualitas proses belajar mengajar.

Meskipun hasil yang diperoleh sangat positif, terdapat beberapa rekomendasi yang dapat dipertimbangkan untuk pengembangan ke depannya. Pertama, disarankan untuk melakukan validasi model dengan menggunakan *dataset* yang lebih besar dan beragam untuk meningkatkan kemampuan generalisasi dan keandalan model. Selain itu, menambahkan atribut relevan lainnya seperti motivasi belajar dan dukungan orang tua dapat memberikan wawasan lebih komprehensif mengenai faktor-faktor yang memengaruhi capaian belajar siswa. Penting juga untuk membandingkan performa algoritma ini dengan algoritma klasifikasi lainnya guna menemukan metode yang paling optimal. Terakhir, mengintegrasikan model ini ke dalam sistem informasi sekolah dapat memfasilitasi guru dalam mengakses hasil klasifikasi dan merancang intervensi yang tepat bagi siswa yang memerlukan perhatian lebih.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada pihak sekolah SDN Tridayasakti 03 Tambun Selatan, khususnya kepada Kepala Sekolah dan Seluruh staf pengajar, atas dukungan serta kerja sama yang

diberikan selama proses pengumpulan data. Bantuan dan izin yang diberikan sangat berperan sehingga peneltian ini dapat terselesaikan.

DAFTAR PUSTAKA

[1] D. Abisono Punkastyo, F. Septian, And A. Syaripudin, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa," 2024.

[2] D. Bañeres, M. E. Rodríguez-González, A. E. Guerrero-Roldán, And P. Cortadas, "An Early Warning System To Identify And Intervene Online Dropout Learners," *International Journal Of Educational Technology In Higher Education*, Vol. 20, No. 1, Dec. 2023, Doi: 10.1186/S41239-022-00371-5.

[3] S. Batool, J. Rashid, Wasif Nisar Muhammad, J. Kim, H.-Y. Kwon, And A. Hussain, "Educational Data Mining To Predict Students' Academic Performa: A Survey Study," *Advances In Cement Research*, Vol. 36, No. 5, Pp. 230–239, Aug. 2023, Doi: 10.1680/Jadcr.23.00031.

[4] S. V. Tsiu, M. Ngobeni, L. Mathabela, And B. Thango, "Applications And Competitive Advantages Of Data Mining And Business Intelligence In Smes Performa: A Systematic Review," *Businesses*, Vol. 5, No. 2, P. 22, May 2025, Doi: 10.3390/Businesses5020022.

[5] Y. Lin, H. Chen, W. Xia, F. Lin, Z. Wang, And Y. Liu, "A Comprehensive Survey On Deep Learning Techniques In Educational Data Mining," 2025, *Springer Nature*. Doi: 10.1007/S41019-025-00303-Z.

[6] A. López-García, O. Blasco-Blasco, M. Liern-García, And S. E. Parada-Rico, "Early Detection Of Students' Failure Using Machine Learning Techniques," *Operations Research Perspectives*, Vol. 11, Dec. 2023, Doi: 10.1016/J.Orp.2023.100292.

[7] X. Shu And Y. Ye, "Knowledge Discovery: Methods From Data Mining And Machine Learning," *Soc Sci Res*, Vol. 110, Feb. 2023, Doi: 10.1016/J.Ssresearch.2022.102817.

[8] A. S. Alghamdi And A. Rahman, "Data Mining Approach To Predict Success Of Secondary School Students: A Saudi Arabian Case Study," *Educ Sci (Basel)*, Vol. 13, No. 3, Mar. 2023, Doi: 10.3390/Educsci13030293.

[9] W. Mu *Et Al.*, "Making Food Systems More Resilient To Food Safety Risks By Including Artificial Intelligence, Big Data, And Internet Of Things Into Food Safety Early Warning And Emerging Risk Identification Tools," Jan. 01, 2024, *John Wiley And Sons Inc*. Doi: 10.1111/1541-4337.13296.

[10] T. Wongvorachan, S. He, And O. Bulut, "A Comparison Of Undersampling, Oversampling, And Smote Methods For Dealing With Imbalanced Classification In Educational Data Mining," *Information (Switzerland)*, Vol. 14, No. 1, Jan. 2023, Doi: 10.3390/Info14010054.

[11] P. Jieyang *Et Al.*, "A Systematic Review Of Data-Driven Approaches To Fault Diagnosis And Early Warning," Dec. 01, 2023, *Springer*. Doi: 10.1007/S10845-022-02020-0.

[12] A. Akram, N. Risal, N. I. Yusuf, A. B. Kaswar, And F. Adiba, "Penerapan Data Mining Dalam Mengklasifikasikan Tingkat Kasus Covid-19 Di Sulawesi Selatan Menggunakan Algoritma Na *Naïve* Bayes Indonesian Journal Of Fundamental Sciences," *Indonesian Journal Of Fundamental Sciences*, Vol. 7, No. 1, 2021.

[13] Ina Magdalena, Gilang Ramadhan, Hasanah Dwi Wahyuni, And Nabilah Dwi Safitri, "Pentingnya Proses Evaluasi Dalam Pembelajaran Di Sekolah Dasar," *Ta'rim: Jurnal Pendidikan Dan Anak Usia Dini*, Vol. 4, No. 3, Pp. 167–176, Jul. 2023, Doi: 10.59059/Tarim.V4i3.220.

[14] Y. B. Utomo, I. Kurniasari, And I. Yanuartanti, "Penerapan Knowledge Discovery In Database Untuk Analisis Tingkat Kecelakaan Lalu Lintas," *Jurnal Teknik Informatika Kaputama (Jtik)*, Vol. 7, No. 1, 2023.

[15] Y. Niar, K. Komariah, A. Surip, R. Saputra, And I. Ali, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Rotan," *Kopertip : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, Vol. 4, No. 1, Pp. 28–34, Jun. 2022, Doi: 10.32485/Kopertip.V4i1.112.

[16] M. Zainuri, "Komparasi Metode Klasifikasi Data Mining Algoritma C5.0 Dan Naïve Bayes Untuk Memprediksi Jurusan Siswa (Studi Kasus: Sman 1 Gondanglegi)," Pp. 1–21, 2022.

[17] F. Santoso, Sunardi, And H. Z. Lukman, "Implementasi Data Mining Dengan Metode Naïve Bayes Untuk Memprediksi Penerimaan Siswa Baru Di Mts Nu Islamiyah Asembagus," *G-Tech:*

Jurnal Teknologi Terapan, Vol. 7, No. 4, Pp. 1355–1366, Oct. 2023, Doi: 10.33379/Gtech.V7i4.3086.

[18] Nurmala, E. S. Susanto, And I. M. Widiarta, "Implementasi Data Mining Untuk Memprediksi Kecocokan Gaya Belajar Siswa Sekolah Dasar Menggunakan Algoritma C4.5," Vol. 4, Pp. 1–10, Jul. 2024.

[19] A. Fajriansyah, D. Yusup, And K. Prihandini, "Prediksi Kelulusan Nilai Kalkulus Menggunakan Naïve Bayes," 2025.

[20] T. Adrian And N. Suarna, "Implementasi Data Mining Untuk Mengklasifikasi Hasil Kelulusan Madrasah Menggunakan Algoritma Naïve Bayes," *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, 2024.

[21] W. N. Indah, "Implementasi Machine Learning Menggunakan Rapidminer Untuk Mengidentifikasi Faktor-Faktor Yang Mempengaruhi Kemampuan Penalaran Matematis," Pp. 1–214, 2025.

[22] R. Trisudarmo, "Penerapan Metode Prototype Dalam Sistem E-Government Pada Pelayanan Administrasi Kependudukan," *Jurnal Informatika Dan Teknologi Pendidikan*, Vol. 2, No. 2, Pp. 64–71, 2022, Doi: 10.25008/Jitp.V2i2.3.