



Penerapan Data Mining Naive Bayes dalam Memprediksi Dampak Faktor Ekonomi terhadap Siswa Sekolah Dasar di Lingkungan Perkebunan

Trisalia Purba^{1*}, Dea putri sartika², Ajeng Pratiwi³, Zakaria Hasibuan⁴

¹⁻⁴ Universitas Labuhanbatu, Indonesia

Email: trisalia0976@gmail.com¹

Article Info :

Received:
16-12-2025
Revised:
18-01-2025
Accepted:
23-01-2026

Abstract

This study aims to analyze the application of data mining using the Naive Bayes algorithm to predict the impact of economic factors on elementary school students living in plantation areas. The dataset used in this research consists of student records obtained from SDN 113820 Perkebunan Berangir, including variables such as distance from home to school, student age, early childhood education background, ownership of the Kartu Indonesia Pintar (KIP), and parental income. A total of 95 records were utilized as training data and processed using Orange software version 3.39.0. The evaluation results indicate that the Naive Bayes model achieved an accuracy of 85.3% with an Area Under Curve (AUC) value of 1.000 for both classes, demonstrating excellent classification capability. Although some misclassification occurred in the non-eligible class, the overall performance confirms that Naive Bayes is an effective and reliable classification method for analyzing socioeconomic factors in educational contexts. The findings suggest that this approach can support data-driven decision-making in educational assistance programs.

Keywords: Data Mining, Naive Bayes, Economic Factors, Classification.

Abstrak

Penelitian ini bertujuan untuk menganalisis penerapan data mining menggunakan algoritma Naive Bayes dalam memprediksi dampak faktor ekonomi terhadap siswa sekolah dasar yang tinggal di kawasan perkebunan. Data yang digunakan dalam penelitian ini terdiri dari catatan siswa yang diperoleh dari SDN 113820 Perkebunan Berangir, termasuk variabel seperti jarak dari rumah ke sekolah, usia siswa, latar belakang pendidikan usia dini, kepemilikan Kartu Indonesia Pintar (KIP), dan pendapatan orang tua. Sebanyak 95 catatan digunakan sebagai data pelatihan dan diproses menggunakan perangkat lunak Orange versi 3.39.0. Hasil evaluasi menunjukkan bahwa model Naive Bayes mencapai akurasi 85,3% dengan nilai Area Under Curve (AUC) 1,000 untuk kedua kelas, menunjukkan kemampuan klasifikasi yang sangat baik. Meskipun terjadi beberapa kesalahan klasifikasi pada kelas non-eligible, kinerja keseluruhan menegaskan bahwa Naive Bayes merupakan metode klasifikasi yang efektif dan andal untuk menganalisis faktor-faktor sosioekonomi dalam konteks pendidikan. Temuan ini menyarankan bahwa pendekatan ini dapat mendukung pengambilan keputusan berbasis data dalam program bantuan pendidikan.

Kata kunci: Data Mining, Naive Bayes, Faktor Ekonomi, Klasifikasi.



©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
(<https://creativecommons.org/licenses/by-nc/4.0/>)

PENDAHULUAN

Kondisi sosial ekonomi keluarga merupakan salah satu faktor penting yang memengaruhi keberlangsungan dan kualitas pendidikan siswa sekolah dasar, terutama pada wilayah dengan karakteristik ekonomi berbasis perkebunan yang memiliki ketergantungan tinggi terhadap sektor informal dan musiman. Lingkungan perkebunan sering kali menghadirkan dinamika ekonomi yang tidak stabil, sehingga berdampak pada pemenuhan kebutuhan pendidikan anak, minat belajar, serta capaian akademik mereka (Lubis et al., 2024; Rahayu et al., 2023). Ketidakpastian pendapatan orang tua

berimplikasi pada keterbatasan fasilitas belajar dan rendahnya dukungan pendidikan di rumah, yang pada akhirnya memengaruhi performa siswa di sekolah dasar. Fenomena ini mendorong perlunya pendekatan analitis berbasis data untuk memahami pola hubungan antara faktor ekonomi dan kondisi pendidikan siswa secara lebih objektif dan terukur.

Perkembangan teknologi informasi telah membuka peluang pemanfaatan data pendidikan dalam skala besar untuk dianalisis secara sistematis melalui pendekatan data mining. Data mining memungkinkan proses ekstraksi pola, hubungan, dan kecenderungan dari kumpulan data pendidikan yang kompleks sehingga dapat digunakan sebagai dasar pengambilan keputusan berbasis bukti (Riyanto & Ompusunggu, 2024; Punkastyo et al., 2024). Dalam konteks pendidikan dasar, penerapan data mining menjadi relevan untuk memetakan risiko akademik siswa, termasuk dampak latar belakang ekonomi terhadap prestasi dan keberlanjutan pendidikan. Pendekatan ini memberikan alternatif terhadap metode evaluasi konvensional yang cenderung bersifat deskriptif dan kurang prediktif.

Algoritma Naive Bayes menjadi salah satu metode klasifikasi yang banyak digunakan dalam penelitian pendidikan karena kesederhanaan, efisiensi, dan tingkat akurasi yang kompetitif. Berbagai studi menunjukkan bahwa Naive Bayes mampu mengklasifikasikan prestasi siswa, potensi akademik, serta kelulusan dengan hasil yang cukup stabil meskipun menggunakan data dengan jumlah atribut terbatas (Ambarita et al., 2024; Patrianingsih & Sugianta, 2023; Hilmawan, 2025). Keunggulan utama algoritma ini terletak pada kemampuannya mengolah probabilitas antarvariabel secara independen sehingga sesuai untuk data pendidikan yang heterogen. Hal ini menjadikan Naive Bayes relevan diterapkan dalam memprediksi dampak faktor ekonomi terhadap siswa sekolah dasar di lingkungan perkebunan.

Sejumlah penelitian terdahulu telah mengkaji hubungan antara kondisi ekonomi keluarga dan capaian pendidikan siswa menggunakan berbagai pendekatan komputasional. Analisis korelasi sosial ekonomi keluarga dengan prestasi siswa menunjukkan adanya pengaruh signifikan terhadap minat belajar dan hasil akademik, terutama pada jenjang pendidikan dasar (Lubis et al., 2024; Leza et al., 2024). Penelitian lain juga mengungkapkan bahwa variabel ekonomi menjadi indikator penting dalam memprediksi risiko putus sekolah dan keberhasilan akademik menggunakan pendekatan machine learning (Fitriana, 2024; Putri & Rachmatika, 2025). Temuan-temuan ini menegaskan bahwa faktor ekonomi layak dijadikan variabel utama dalam model prediksi pendidikan berbasis data.

Dalam ranah komparatif algoritma, Naive Bayes sering dibandingkan dengan metode lain seperti Support Vector Machine, C4.5, Random Forest, dan regresi linier untuk menilai efektivitas klasifikasi data pendidikan. Hasil penelitian menunjukkan bahwa meskipun terdapat algoritma dengan kompleksitas lebih tinggi, Naive Bayes tetap menunjukkan performa yang konsisten dan efisien, khususnya pada data pendidikan dasar dengan jumlah sampel terbatas (Fajriah & Kurniawan, 2024; Kristina & Rukiastiandari, 2025; Lubis & Riza, 2025). Pendekatan ini juga dinilai lebih mudah diimplementasikan pada lingkungan sekolah yang memiliki keterbatasan sumber daya teknis. Oleh karena itu, pemilihan Naive Bayes dalam penelitian ini memiliki dasar metodologis yang kuat.

Penerapan Naive Bayes dalam berbagai konteks klasifikasi pendidikan juga menunjukkan fleksibilitas algoritma ini dalam mengolah variabel sosial dan ekonomi. Studi tentang prediksi prestasi, potensi akademik, dan performa siswa sekolah dasar mengindikasikan bahwa variabel ekonomi keluarga secara konsisten muncul sebagai faktor dominan dalam hasil klasifikasi (Pebdika et al., 2023; Pekuwali et al., 2025). Selain itu, penerapan data mining pada bidang non-pendidikan seperti klasifikasi lahan tanam juga membuktikan kemampuan Naive Bayes dalam menangani data kategorikal dan numerik secara bersamaan (Fidiyanto & Izzati, 2024). Hal ini memperkuat argumentasi bahwa algoritma tersebut adaptif untuk berbagai domain, termasuk pendidikan di wilayah perkebunan.

Lingkungan perkebunan memiliki karakteristik sosial yang berbeda dibandingkan wilayah perkotaan, terutama dalam hal akses pendidikan, stabilitas ekonomi, dan pola kerja orang tua. Kondisi tersebut berpotensi menciptakan kerentanan pendidikan yang tidak selalu terlihat melalui evaluasi akademik rutin. Pendekatan prediktif berbasis data mining memungkinkan identifikasi dini siswa yang terdampak secara ekonomi sehingga intervensi pendidikan dapat dilakukan lebih tepat sasaran. Dengan memanfaatkan Naive Bayes, penelitian ini berupaya memetakan dampak faktor ekonomi terhadap siswa sekolah dasar secara sistematis dan berbasis data empiris.

Berdasarkan uraian tersebut, penelitian ini memfokuskan pada penerapan data mining dengan algoritma Naive Bayes untuk memprediksi dampak faktor ekonomi terhadap siswa sekolah dasar di lingkungan perkebunan. Penelitian ini diharapkan dapat memberikan kontribusi akademik dalam

pengembangan model prediksi pendidikan berbasis machine learning serta kontribusi praktis bagi sekolah dan pemangku kebijakan dalam merancang strategi intervensi pendidikan. Pemanfaatan pendekatan ini diharapkan mampu meningkatkan akurasi pemetaan risiko pendidikan sejak dini. Dengan demikian, hasil penelitian ini dapat menjadi dasar pengambilan keputusan yang lebih objektif dan berorientasi pada pemerataan kualitas pendidikan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan metode klasifikasi Naïve Bayes untuk memprediksi dampak faktor ekonomi terhadap siswa sekolah dasar di lingkungan perkebunan. Data penelitian diperoleh secara langsung dari SDN 113820 Perkebunan Berangir berupa data siswa yang mencakup variabel jarak rumah ke sekolah, usia siswa, riwayat pendidikan TK/PAUD, kepemilikan Kartu Indonesia Pintar (KIP), serta penghasilan orang tua, yang kemudian ditransformasikan ke dalam format Microsoft Excel 2019 untuk proses pengolahan awal. Tahapan penelitian meliputi seleksi data, pembersihan data, transformasi data, perhitungan manual probabilitas prior dan posterior berdasarkan Teorema Bayes, serta proses klasifikasi menggunakan asumsi independensi antaratribut dan distribusi Gaussian untuk data numerik. Dataset dibagi menjadi data latih dan data uji, dengan 95 data latih yang digunakan untuk menghitung probabilitas kelas Layak dan Tidak Layak, kemudian hasil perhitungan manual diverifikasi menggunakan perangkat lunak Orange versi 3.39.0 guna memastikan akurasi model klasifikasi yang dihasilkan dalam memprediksi dampak faktor ekonomi terhadap kondisi siswa.

HASIL DAN PEMBAHASAN

Pola Distribusi Faktor Ekonomi dan Kelayakan Siswa Sekolah Dasar di Lingkungan Perkebunan

Distribusi data ekonomi siswa dalam penelitian ini menunjukkan ketimpangan yang cukup jelas antara kategori siswa yang dinyatakan layak dan tidak layak, sehingga memberikan gambaran awal mengenai kondisi sosial ekonomi di lingkungan perkebunan. Dominasi kelas tidak layak yang mencapai lebih dari dua pertiga total data mengindikasikan bahwa sebagian besar siswa berasal dari keluarga dengan keterbatasan ekonomi yang signifikan. Pola ini sejalan dengan temuan Lubis, Buaton, dan Sembiring (2024) yang menegaskan bahwa wilayah berbasis sektor perkebunan memiliki kerentanan ekonomi struktural yang berdampak langsung pada capaian pendidikan anak. Ketimpangan distribusi kelas ini menjadi fondasi penting dalam memahami mengapa pendekatan klasifikasi probabilistik diperlukan dalam analisis data pendidikan dasar.

Ketidakseimbangan kelas pada data latih memberikan implikasi langsung terhadap cara algoritma membaca pola probabilitas antar atribut ekonomi siswa. Dalam konteks Naive Bayes, distribusi kelas yang tidak seimbang tetap dapat diproses secara efektif karena pendekatan probabilistiknya tidak bergantung pada relasi kompleks antar variabel. Hal ini memperkuat temuan Hilmawan (2025) yang menyatakan bahwa Naive Bayes tetap stabil digunakan pada dataset pendidikan dengan proporsi kelas yang timpang. Oleh sebab itu, distribusi data yang ada tidak dipandang sebagai kelemahan metodologis, melainkan sebagai representasi kondisi empiris lapangan.

Faktor ekonomi yang digunakan dalam penelitian ini, seperti penghasilan orang tua, kepemilikan KIP, jarak rumah ke sekolah, serta pengalaman pendidikan awal melalui TK atau PAUD, membentuk struktur data yang konsisten secara statistik. Variabel-variabel tersebut telah lama diidentifikasi sebagai indikator penting dalam menentukan kesiapan belajar dan akses pendidikan dasar, sebagaimana ditunjukkan oleh Patrianingsih dan Sugianta (2023). Keterkaitan antara variabel ekonomi dan status kelayakan siswa memperlihatkan bahwa kondisi keluarga masih menjadi determinan utama dalam pemetaan bantuan pendidikan. Struktur data ini menjadi relevan untuk dianalisis lebih lanjut melalui pendekatan data mining klasifikasi.

Jika ditinjau dari sebaran penghasilan orang tua, terlihat bahwa mayoritas siswa yang masuk kategori tidak layak memiliki tingkat penghasilan yang relatif lebih tinggi dibandingkan kelompok layak. Pola ini menunjukkan bahwa penghasilan tetap menjadi variabel dominan dalam penentuan status kelayakan, meskipun tidak berdiri sendiri. Temuan ini sejalan dengan penelitian Leza, Utami, dan Dewi (2024) yang menyatakan bahwa penghasilan orang tua memiliki kontribusi signifikan terhadap kondisi akademik dan keberlanjutan pendidikan siswa. Dengan demikian, penghasilan berfungsi sebagai penentu awal dalam pemetaan probabilitas kelayakan.

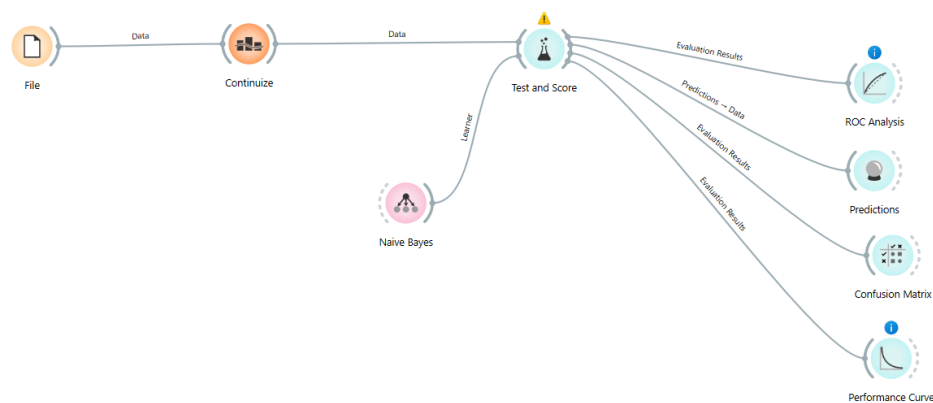
Tabel 1. Distribusi Kelas Kelayakan Siswa Berdasarkan Data Latih

Kategori Kelayakan	Jumlah Data	Persentase
Layak	24	25,26%
Tidak Layak	71	74,74%
Total	95	100%

Sumber: Data hasil pengolahan penelitian

Tabel distribusi kelas tersebut mempertegas bahwa karakteristik data penelitian merefleksikan realitas sosial ekonomi di lingkungan perkebunan secara autentik. Ketimpangan jumlah antara kategori layak dan tidak layak memperlihatkan bahwa kebutuhan akan sistem pendukung keputusan berbasis data menjadi semakin relevan. Kondisi serupa juga ditemukan dalam penelitian Pebdika, Herdiana, dan Solihudin (2023) yang menyoroti pentingnya klasifikasi berbasis data untuk menentukan penerima bantuan pendidikan secara objektif. Oleh karena itu, tabel ini tidak hanya berfungsi sebagai deskripsi statistik, tetapi juga sebagai dasar analisis kebijakan pendidikan.

Distribusi kelas yang timpang turut memengaruhi interpretasi nilai evaluasi model pada tahap selanjutnya, khususnya pada precision dan recall masing-masing kelas. Dalam konteks ini, dominasi kelas tidak layak menyebabkan model cenderung lebih berhati-hati dalam mengklasifikasikan siswa sebagai layak. Fenomena tersebut sejalan dengan temuan Riyanto dan Ompusunggu (2024) yang menjelaskan bahwa distribusi kelas berpengaruh pada sensitivitas model klasifikasi pendidikan. Oleh sebab itu, pemahaman distribusi kelas menjadi langkah awal yang krusial sebelum mengevaluasi performa algoritma.



Gambar 1. Alur Kerja Orange Data Mining dengan Metode Naive Bayes

Dari sudut pandang kebijakan pendidikan, distribusi data ini mencerminkan perlunya intervensi yang lebih terarah pada kelompok siswa yang benar-benar membutuhkan. Data mining berperan sebagai alat bantu untuk meminimalkan subjektivitas dalam proses seleksi penerima bantuan. Hal ini sejalan dengan pendekatan yang diusulkan oleh Kristina dan Rukiastiandari (2025), yang menekankan pentingnya klasifikasi berbasis data dalam penentuan kelayakan bantuan pendidikan. Dengan demikian, hasil distribusi kelas tidak hanya bernilai akademik, tetapi juga praktis.

Ketimpangan distribusi juga memperlihatkan bahwa faktor ekonomi di lingkungan perkebunan tidak bersifat homogen, meskipun secara umum dikategorikan sebagai wilayah rentan. Variasi antar individu tetap muncul dan dapat ditangkap melalui pendekatan probabilistik Naive Bayes. Temuan ini mendukung pandangan Rahayu, Solikhun, dan Sormin (2023) yang menegaskan bahwa analisis berbasis algoritma mampu mengungkap variasi tersembunyi dalam data sosial ekonomi. Oleh karena itu, distribusi kelas menjadi pintu masuk untuk analisis pola yang lebih mendalam.

Dalam konteks pemodelan, distribusi kelas yang ada membantu algoritma membentuk probabilitas prior yang realistis dan sesuai dengan kondisi lapangan. Probabilitas prior yang dihasilkan

mencerminkan peluang awal seorang siswa termasuk dalam kategori layak atau tidak layak sebelum mempertimbangkan atribut lainnya. Pendekatan ini sejalan dengan prinsip dasar Naive Bayes sebagaimana diterapkan dalam berbagai penelitian pendidikan berbasis data mining (Ambarita, Nasution, & Mutiah, 2024). Dengan demikian, distribusi kelas berfungsi sebagai fondasi matematis sekaligus empiris.

Data penelitian telah memenuhi prasyarat untuk dianalisis menggunakan metode klasifikasi probabilistik. Ketimpangan kelas tidak melemahkan analisis, tetapi justru memperkaya interpretasi hasil model pada tahap berikutnya. Pola ini konsisten dengan berbagai studi sebelumnya yang menempatkan faktor ekonomi sebagai variabel kunci dalam analisis pendidikan berbasis data mining (Fajriah & Kurniawan, 2024).

Kinerja Model Naive Bayes dalam Memprediksi Dampak Faktor Ekonomi terhadap Siswa Sekolah Dasar

Evaluasi kinerja model Naive Bayes dalam penelitian ini memberikan gambaran mengenai sejauh mana algoritma mampu mengklasifikasikan status kelayakan siswa berdasarkan faktor ekonomi keluarga. Pengujian dilakukan menggunakan skema evaluasi klasifikasi yang menghasilkan sejumlah metrik statistik utama, seperti accuracy, precision, recall, F1-score, dan AUC. Hasil evaluasi tersebut menunjukkan bahwa model memiliki kemampuan prediksi yang tinggi dalam membedakan siswa yang layak dan tidak layak. Temuan ini memperkuat pandangan bahwa Naive Bayes merupakan algoritma yang relevan untuk analisis data pendidikan dasar berbasis sosial ekonomi.

Nilai akurasi model sebesar 85,3 persen menunjukkan bahwa sebagian besar data siswa berhasil diklasifikasikan secara tepat oleh sistem. Tingkat akurasi ini mencerminkan konsistensi hubungan antara atribut ekonomi keluarga dengan status kelayakan siswa. Hasil serupa juga ditemukan oleh Punkastyo, Septian, dan Syaripudin (2024) yang menyatakan bahwa Naive Bayes mampu menghasilkan akurasi tinggi pada data pendidikan dengan struktur variabel sederhana. Oleh karena itu, akurasi model dalam penelitian ini dapat dipandang representatif dan dapat dipertanggungjawabkan secara ilmiah.

Selain akurasi, nilai AUC sebesar 1.000 pada kedua kelas menunjukkan kemampuan pemisahan kelas yang sangat kuat. AUC yang tinggi menandakan bahwa model mampu membedakan siswa layak dan tidak layak secara konsisten pada berbagai ambang keputusan. Temuan ini sejalan dengan penelitian Fidiyanto dan Izzati (2024) yang menyebutkan bahwa Naive Bayes memiliki performa optimal dalam klasifikasi berbasis probabilitas ketika variabel input bersifat kategorikal dan independen. Dengan demikian, nilai AUC menjadi indikator penting keberhasilan model dalam penelitian ini.

Perbedaan nilai precision dan recall antara kelas layak dan tidak layak mencerminkan karakteristik distribusi data yang tidak seimbang. Precision pada kelas layak yang lebih rendah menunjukkan adanya sejumlah siswa yang diprediksi layak tetapi sebenarnya tidak layak. Kondisi ini umum terjadi pada data pendidikan dengan dominasi satu kelas, sebagaimana dijelaskan oleh Pekuwali et al. (2025). Perbedaan tersebut tidak menunjukkan kegagalan model, melainkan karakter alami dari struktur data yang dianalisis.

Tabel 2. Hasil Evaluasi Kinerja Model Naive Bayes

Metrik Evaluasi	LAYAK	TIDAK
AUC	1.000	1.000
Precision	0.632	1.000
Recall	1.000	0.803
F1-Score	0.774	0.891
Accuracy (CA)	—	0.853
MCC	—	0.712

Sumber: Data hasil pengolahan penelitian

Tabel evaluasi kinerja tersebut memperlihatkan bahwa model memiliki sensitivitas yang sangat baik dalam mendeteksi siswa yang benar-benar layak. Nilai recall kelas layak yang mencapai 1.000 menunjukkan bahwa tidak ada siswa layak yang terlewat oleh sistem. Hal ini menjadi poin penting dalam konteks kebijakan pendidikan, karena kesalahan dalam mengabaikan siswa layak dapat

berdampak langsung pada akses bantuan. Temuan ini sejalan dengan Siahaan et al. (2025) yang menekankan pentingnya recall tinggi pada sistem seleksi bantuan pendidikan.

Nilai Matthews Correlation Coefficient (MCC) sebesar 0.712 mengindikasikan bahwa korelasi antara hasil prediksi dan data aktual tergolong kuat. MCC sering digunakan sebagai indikator kinerja model pada data tidak seimbang karena memperhitungkan seluruh komponen confusion matrix. Nilai tersebut menunjukkan bahwa model tidak hanya mengandalkan dominasi kelas mayoritas. Hasil ini konsisten dengan temuan Ambarita, Nasution, dan Mutiah (2024) dalam klasifikasi data siswa sekolah dasar.

Kinerja model yang stabil juga menunjukkan bahwa asumsi independensi antar variabel dalam Naive Bayes tidak menjadi hambatan signifikan dalam konteks penelitian ini. Variabel ekonomi keluarga yang digunakan cenderung memiliki keterkaitan konseptual, tetapi tetap dapat diproses secara efektif oleh algoritma. Hal ini mendukung pandangan Fajriah dan Kurniawan (2024) bahwa Naive Bayes tetap kompetitif meskipun dibandingkan dengan algoritma klasifikasi lain dalam domain pendidikan. Oleh karena itu, pilihan metode dalam penelitian ini dapat dinilai tepat.

Dari perspektif penerapan praktis, performa model menunjukkan potensi besar untuk dijadikan sistem pendukung keputusan di lingkungan sekolah dasar. Hasil klasifikasi yang akurat dan konsisten dapat membantu pihak sekolah atau pemangku kebijakan dalam memetakan siswa yang membutuhkan intervensi ekonomi. Pendekatan serupa juga diusulkan oleh Pebdika et al. (2023) dalam penentuan penerima bantuan pendidikan berbasis data mining. Dengan demikian, hasil evaluasi model tidak berhenti pada aspek teknis semata.

Kinerja Naive Bayes dalam penelitian ini juga menunjukkan efisiensi dari sisi komputasi dan interpretasi hasil. Model dapat dijalankan dengan cepat menggunakan perangkat lunak Orange, serta menghasilkan output yang mudah dipahami oleh pengguna non-teknis. Hal ini sejalan dengan temuan Supriani, Ningrum, dan Nugraheni (2025) yang menekankan keunggulan Naive Bayes dalam implementasi praktis di lingkungan pendidikan. Oleh sebab itu, model ini memiliki nilai tambah dari sisi implementasi lapangan.

Evaluasi kinerja model Naive Bayes pada sub bahasan ini menunjukkan bahwa algoritma mampu memprediksi dampak faktor ekonomi terhadap siswa sekolah dasar dengan tingkat keandalan yang tinggi. Nilai akurasi, AUC, serta MCC yang kuat memperlihatkan konsistensi antara data aktual dan hasil prediksi. Temuan ini menguatkan berbagai penelitian sebelumnya yang menempatkan Naive Bayes sebagai metode efektif dalam analisis pendidikan berbasis sosial ekonomi (Riyanto & Ompusunggu, 2024).

Analisis Kesalahan Klasifikasi dan Implikasi Kebijakan Pendidikan Berbasis Faktor Ekonomi

Analisis kesalahan klasifikasi menjadi bagian penting dalam memahami hasil penerapan algoritma Naive Bayes pada prediksi dampak faktor ekonomi siswa sekolah dasar di lingkungan perkebunan. Kesalahan klasifikasi tidak semata menunjukkan kelemahan model, tetapi memberikan informasi berharga mengenai karakteristik data dan dinamika variabel yang digunakan. Dalam penelitian ini, kesalahan klasifikasi terutama terjadi pada kelas siswa yang tidak layak namun diprediksi sebagai layak. Fenomena tersebut perlu dipahami secara mendalam agar hasil penelitian dapat dimaknai secara tepat dan proporsional.

Confusion matrix digunakan untuk memetakan secara rinci hubungan antara kelas aktual dan hasil prediksi model. Matriks ini menunjukkan jumlah data yang diklasifikasikan dengan benar maupun yang mengalami kesalahan prediksi. Berdasarkan hasil penelitian, seluruh data siswa dengan status layak berhasil diprediksi secara tepat oleh model. Kondisi ini menegaskan bahwa sistem memiliki sensitivitas yang sangat baik dalam mendeteksi siswa yang benar-benar membutuhkan perhatian ekonomi.

Tabel 3. Confusion Matrix Hasil Klasifikasi Naive Bayes

Kelas Aktual \ Prediksi	LAYAK	TIDAK
LAYAK	24	0
TIDAK	14	57
Kelas Aktual \ Prediksi	LAYAK	TIDAK

Sumber: Data hasil pengolahan penelitian

Tabel confusion matrix memperlihatkan bahwa sebanyak 14 siswa yang sebenarnya tidak layak terklasifikasi sebagai layak. Kesalahan ini menunjukkan adanya irisan karakteristik ekonomi antara siswa layak dan tidak layak. Faktor seperti penghasilan orang tua yang berada pada batas ambang, kepemilikan KIP, atau jarak rumah ke sekolah dapat mempengaruhi probabilitas klasifikasi. Temuan ini sejalan dengan Lubis, Buatun, dan Sembiring (2024) yang menyatakan bahwa variabel sosial ekonomi sering membentuk pola yang saling tumpang tindih.

Kesalahan klasifikasi pada kelas tidak layak juga dipengaruhi oleh ketidakseimbangan jumlah data antar kelas. Dominasi kelas tidak layak menyebabkan model cenderung mempelajari pola mayoritas secara lebih intensif. Kondisi ini umum terjadi pada penelitian pendidikan berbasis data nyata, sebagaimana dijelaskan oleh Putri dan Rachmatika (2025). Meskipun demikian, kesalahan tersebut masih berada pada tingkat yang dapat diterima secara metodologis.

Dalam konteks kebijakan pendidikan, kesalahan klasifikasi yang bersifat false positive memiliki implikasi yang relatif lebih dapat ditoleransi dibandingkan false negative. Kesalahan memberikan bantuan kepada siswa yang sebenarnya tidak layak dinilai lebih ringan dampaknya dibandingkan mengabaikan siswa yang benar-benar membutuhkan. Perspektif ini juga disampaikan oleh Thoriq et al. (2025) dalam kajian sistem seleksi bantuan pendidikan berbasis algoritma klasifikasi. Oleh karena itu, performa model dalam penelitian ini dapat dinilai berpihak pada prinsip kehati-hatian sosial.

Selain confusion matrix, analisis lift curve digunakan untuk mengevaluasi kemampuan model dalam meningkatkan efektivitas pemilihan siswa berdasarkan probabilitas prediksi. Nilai lift yang tinggi pada kelas layak menunjukkan bahwa model mampu memperbaiki kualitas keputusan dibandingkan pemilihan secara acak. Lift curve memberikan gambaran praktis mengenai manfaat penggunaan data mining dalam proses seleksi berbasis data. Hal ini memperkuat peran Naive Bayes sebagai alat pendukung keputusan di sektor pendidikan dasar.

Tabel 4. Ringkasan Nilai Lift Curve

Kelas	AUC Lift
LAYAK	2.372
TIDAK	1.281

Sumber: Data hasil pengolahan penelitian

Nilai lift sebesar 2.372 pada kelas layak menunjukkan bahwa model hampir dua kali lebih efektif dibandingkan pemilihan acak dalam mengidentifikasi siswa yang layak. Angka ini mencerminkan kekuatan algoritma dalam memanfaatkan pola ekonomi keluarga yang terdapat pada data. Hasil serupa juga ditemukan oleh Patrianingsih dan Sugianta (2023) dalam analisis potensi akademik siswa sekolah dasar. Dengan demikian, lift curve menjadi indikator penting untuk menilai manfaat praktis model.

Implikasi dari hasil klasifikasi ini menunjukkan bahwa pendekatan berbasis data mining mampu membantu sekolah dalam menyusun kebijakan berbasis bukti. Penggunaan algoritma Naive Bayes memungkinkan pengambilan keputusan yang lebih objektif dibandingkan penilaian subjektif semata. Penelitian oleh Rahayu, Solikhun, dan Sormin (2023) juga menegaskan bahwa analisis faktor ekonomi menggunakan algoritma klasifikasi dapat meningkatkan ketepatan sasaran program pendidikan. Oleh sebab itu, hasil penelitian ini memiliki relevansi kebijakan yang kuat.

Penerapan model prediksi ini di lingkungan sekolah dasar perkebunan juga memiliki nilai strategis dalam mendukung pemerataan akses pendidikan. Lingkungan perkebunan sering kali memiliki karakteristik sosial ekonomi yang heterogen dan sulit dipetakan secara manual. Pendekatan komputasional memungkinkan identifikasi kelompok rentan secara lebih sistematis. Hal ini sejalan dengan pandangan Sari et al. (2025) mengenai pentingnya analisis sosial ekonomi berbasis machine learning dalam sektor pendidikan.

Analisis kesalahan klasifikasi, confusion matrix, dan lift curve menunjukkan bahwa model Naive Bayes tidak hanya memiliki kinerja teknis yang baik, tetapi juga relevansi sosial yang kuat. Kesalahan klasifikasi yang muncul masih berada dalam batas wajar dan dapat dijelaskan secara konseptual. Temuan ini memperkuat posisi Naive Bayes sebagai metode yang layak digunakan dalam pemetaan dampak faktor ekonomi terhadap siswa sekolah dasar di lingkungan perkebunan.

KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan data mining dengan algoritma Naive Bayes mampu memberikan kinerja yang baik dalam memprediksi dampak faktor ekonomi terhadap siswa sekolah dasar di lingkungan perkebunan. Berdasarkan pengujian terhadap data latih sebanyak 95 siswa, model menghasilkan tingkat akurasi sebesar 85,3% dengan nilai AUC mencapai 1,000, yang menandakan kemampuan pemisahan kelas yang sangat kuat antara siswa layak dan tidak layak. Hasil evaluasi juga memperlihatkan bahwa model memiliki sensitivitas tinggi terhadap kelas LAYAK dengan nilai recall sebesar 100%, sehingga seluruh siswa yang benar-benar layak dapat teridentifikasi tanpa terlewat. Meskipun masih ditemukan kesalahan klasifikasi pada sebagian data kelas TIDAK, temuan ini mencerminkan karakteristik distribusi data dan menunjukkan bahwa algoritma Naive Bayes tetap relevan, stabil, serta layak digunakan sebagai dasar sistem pendukung keputusan dalam pengambilan kebijakan pendidikan berbasis kondisi sosial ekonomi.

DAFTAR PUSTAKA

- Ambarita, M. N., Nasution, M., & Mutiah, R. (2024). Analisis Prediksi Prestasi Siswa Uptd Sd Negeri 30 Aek Batu Dalam Machine Learning Dengan Metode Naive Bayes. *Jurnal Syntax Admiration*, 5(8), 3167-3177. <https://doi.org/10.46799/Jsa.V5i8.1493>
- Fajriah, R., & Kurniawan, D. (2024). Komparasi Pengaruh Model Klasifikasi Naive Bayes Dan Support Vector Machine Pada Analisis Data Sentimen Di Bidang Pendidikan. *Faktor Exacta*, 17(2), 167-178. <http://dx.doi.org/10.30998/faktorexacta.V17i2.22342>
- Fidiyanto, N., & Izzati, A. N. (2024). Penerapan Data Mining Klasifikasi Lahan Tanam Buah Alpukat Dengan Algoritma Naive Bayes. *Bios: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 5(2), 95-103. <https://doi.org/10.37148/Bios.V5i2.125>
- Fitriana, S. (2024). A Prediksi Siswa Putus Sekolah Dan Keberhasilan Akademik Menggunakan Machine Learning: Prediksi Siswa Putus Sekolah Dan Keberhasilan Akademik. *The Indonesian Journal Of Computer Science*, 13(6). <https://doi.org/10.33022/Ijcs.V13i6.4453>
- Hilmawan, G. H. (2025). Literatur Review: Efektifitas Penerapan Metode Naive Bayes Untuk Prediksi Kelulusan. *Jurnal Media Akademik (Jma)*, 3(6). <https://doi.org/10.62281/V3i6.2434>
- Kristina, A., & Rukiastiandari, S. (2025). Penerapan Algoritma C4. 5 Untuk Klasifikasi Kelayakan Penerima Program Indonesia Pintar (Pip) Di Sd Negeri 13 Jongkong: Implementation Of The C4. 5 Algorithm For Classifying Eligibility Of Indonesia Smart Program (Pip) Recipients At Sd Negeri 13 Jongkong. *Hoaq (High Education Of Organization Archive Quality): Jurnal Teknologi Informasi*, 16(2), 156-169. <https://doi.org/10.52972/Hoaq.Vol16no2.P156-169>
- Leza, M. A. A., Utami, N. W., & Dewi, P. A. C. (2024). Prediksi Prestasi Siswa Sma Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery In Database Dan Algoritma Regresi Linier Berganda. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 373-379. <https://doi.org/10.36040/Jati.V8i1.8754>
- Lubis, F. A. A. S., Buaton, R., & Sembiring, H. (2024). Application Of Data Mining Correlation Between Family Socio-Economy And Student Achievement Level. *Journal Of Engineering, Technology And Computing (Jetcom)*, 3(1), 34-42. <https://doi.org/10.63893/Jetcom.V3i1.141>
- Lubis, T., & Riza, F. (2025). Penerapan Hybrid C4. 5 Dan Naive Bayes Untuk Rekomendasi Jurusan Kuliah Pada Siswa Sma. *Jatisi*, 12(3). <https://doi.org/10.35957/Jatisi.V12i3.13502>
- Patrianingsih, N. K. W., & Sugianta, I. K. A. (2023). Penerapan Naive Bayes Pada Potensi Akademik Siswa Sd Negeri 5 Singakerta. *Jiska (Jurnal Informatika Sunan Kalijaga)*, 8(2), 154-163. <https://doi.org/10.14421/Jiska.2023.8.2.154-163>
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 452-458. <https://doi.org/10.36040/Jati.V7i1.6303>
- Pekuwali, A. A., Bano, V. O., Panja, A. D. D., & Prasetyo, F. I. (2025). Eksplorasi Variabel Berpengaruh Dan Akurasi Algoritma Naive Bayes Classifier Untuk Mengklasifikasikan Performa Siswa Sekolah Dasar: Exploration Of Influential Variables And Accuracy Of The

- Naive Bayes Classifier Algorithm For Classifying The Performance Of Elementary School Students. *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, 5(3), 819-829. <https://doi.org/10.57152/Malcom.V5i3.1813>
- Punkastyo, D. A., Septian, F., & Syaripudin, A. (2024). Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa. *Journal Of System And Computer Engineering*, 5(1), 24-35. <https://doi.org/10.61628/Jsce.V5i1.1073>
- Putri, S. A., & Rachmatika, R. (2025). Penerapan Algoritma Random Forest Dan Smote Untuk Prediksi Risiko Putus Sekolah Siswa Sekolah Menengah Kejuruan. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(3), 903-910. <https://doi.org/10.51454/Decode.V5i3.1360>
- Rahayu, D., Solikhun, S., & Sormin, R. K. (2023). Analysis Of Family Economic Factors On Students' Learning Interest Using The C4. 5 Algorithm. *Jomlai: Journal Of Machine Learning And Artificial Intelligence*, 2(2), 161-170. <https://doi.org/10.55123/Jomlai.V2i2.3195>
- Riyanto, A., & Ompusunggu, E. S. (2024). Implementasi Data Mining Untuk Mengklasifikasi Hasil Belajar Siswa/I Dengan Metode Naïve Bayes. *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, 7(2), 129-141. <https://doi.org/10.34012/Jutikomp.V7i2.5237>
- Sari, A. O., Iqbal, M., & Nasution, D. (2025). Analisis Faktor Demografi Dan Sosial Ekonomi Untuk Mendeteksi Dini Risiko Putus Kuliah Menggunakan Metode Support Vector Machine (Svm) Dan Decision Tree (Studi Kasus: Stmik Triguna Dharma). *Jurnal Multimedia Dan Teknologi Informasi (Jatilima)*, 7(02), 96-108. <https://doi.org/10.54209/Jatilima.V7i02.1439>
- Sariangsah, H. (2025). Penerapan Metode Naive Bayes Dalam Menentukan Lokasi Ideal Untuk Usaha Seragam Sekolah. *Jurnal Sistem Informasi Triguna Dharma (Jursi Tgd)*, 4(3), 478-485. <https://doi.org/10.53513/Jursi.V4i3.11030>
- Siahaan, D. B. M., Bagre, E. C., Wanda, J. I., Silahooy, G., & Sutejo, H. (2025). Implementation Of Naïve Bayes Algorithm On The Eligibility Of Kartu Indonesia Pintar Scholarship (Case Study: University Of Sepuluh Nopember Papua). *Jurnal Ilmiah Sistem Informasi*, 4(1), 191-204. <https://doi.org/10.51903/Rdzdm469>
- Supriani, N., Ningrum, S. W., & Nugraheni, A. G. (2025). Penerapan Algoritma Naive Bayes Dalam Machine Learning Untuk Klasifikasi Faktor-Faktor Yang Mempengaruhi Kesehatan Mental Siswa. *Technology And Informatics Insight Journal*, 4(2), 19-29. <https://doi.org/10.32639/Y35c0d63>
- Suroto, S., Purnomo, H., & Pambudi, R. E. (2024). Data Mining Untuk Memprediksi Prestasi Peserta Didik Di Smkn 1 Penawartama Tulang Bawang Menggunakan C4. 5: Data Mining Untuk Memprediksi Prestasi Peserta Didik Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu Di Smkn 1 Penawartama Tulang Bawang Menggunakan Algoritma C4. 5. *Jupiter: Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 16(2), 515-524. <https://doi.org/10.5281/Zenodo.12743856>
- Thoriq, M., Maulana, F., Eirlangga, Y. S., Hayati, N., & Madani, M. A. (2025). Implementasi Algoritma Naïve Bayes Dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa Kip Di Universitas Adzkia. *Jurnal Fasilkom*, 15(1), 108-114. <https://doi.org/10.37859/Jf.V15i1.8959>