

IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI KELAYAKAN PENERIMA PROGRAM INDONESIA PINTAR (PIP) PADA SD NEGERI X

M. Yhogha Ismail Ibn Ibrahim¹, Istiana Ruswita²

¹ Ilmu Komputer PSDKU Waykanan, Matematika Dan IPA, Universitas Lampung,

² Ilmu Komputer, Matematika Dan IPA, Universitas Lampung,

¹ Jl. Prof. Dr. Sumantri Brojonegoro No. 1, Gedong Meneng, Kecamatan Rajabasa, Kota Bandar Lampung¹, ² Jl. Prof. Dr. Sumantri Brojonegoro No. 1, Gedong Meneng, Kecamatan Rajabasa, Kota Bandar Lampung

¹ yogaismail@fmipa.unila.ac.id, ² Istiana.ruswita@fmipa.unila.ac.id

Abstrak : Efektivitas distribusi Program Indonesia Pintar (PIP) sangat bergantung pada akurasi proses seleksi di tingkat sekolah. Di SD Negeri X, mekanisme penentuan penerima bantuan masih menghadapi kendala akibat proses verifikasi manual yang cenderung subjektif dan tidak efisien. Penelitian ini mengeksplorasi penggunaan teknik *data mining* dengan algoritma Naïve Bayes untuk mengotomatisasi klasifikasi kelayakan siswa. Atribut yang diuji meliputi indikator ekonomi dan kondisi sosial keluarga siswa yang diolah menggunakan metode *10-Fold Cross Validation*. Temuan penelitian menunjukkan bahwa model Naïve Bayes berhasil mencapai tingkat akurasi sebesar **78,20%**. Walaupun nilai *recall* (52,68%) dan presisi (53,64%) dipengaruhi oleh karakteristik dataset yang ada, penggunaan algoritma ini memberikan landasan pengambilan keputusan yang lebih terukur dan transparan. Implementasi sistem ini diharapkan dapat meminimalisir kesalahan manusia dalam identifikasi penerima bantuan sehingga penyaluran PIP menjadi lebih tepat sasaran.

Kata Kunci: Naïve Bayes, Seleksi PIP, Klasifikasi, *Data Mining*, Akurasi.

Abstract: *The effectiveness of the Indonesia Smart Program (PIP) distribution significantly depends on the accuracy of the selection process at the school level. At SD Negeri X, the mechanism for determining aid recipients still faces challenges due to manual verification processes that tend to be subjective and inefficient. This study explores the application of data mining techniques using the Naïve Bayes algorithm to automate student eligibility classification. The tested attributes include economic indicators and the social conditions of the students' families, which were processed using the 10-Fold Cross-Validation method. The research findings demonstrate that the Naïve Bayes model successfully achieved an accuracy rate of 78.20%. Although the recall (52.68%) and precision (53.64%) values were influenced by the characteristics of the existing dataset, the use of this algorithm provides a more measurable and transparent basis for decision-making. The implementation of this system is expected to minimize human error in identifying aid recipients, thereby ensuring that PIP distribution is more targeted.*

Keywords: *Naïve Bayes, PIP Selection, Classification, Data Mining, Accuracy.*

PENDAHULUAN

Kesenjangan ekonomi tidak seharusnya memutus akses pendidikan yang merupakan elemen krusial kemajuan bangsa. Oleh karena itu, diluncurkanlah Program Indonesia Pintar (PIP) sebagai pengganti Bantuan Siswa Miskin (BSM) yang lebih komprehensif. Program ini bertujuan melindungi keberlangsungan sekolah bagi siswa dari keluarga prasejahtera agar terhindar dari risiko putus sekolah. Dengan menyasar kelompok usia sekolah dari keluarga penerima manfaat PKH dan KKS, PIP menjadi jembatan bagi mereka untuk meraih ijazah pendidikan menengah melalui berbagai jalur pendidikan yang tersedia.

Namun, efektivitas distribusi bantuan PIP sangat bergantung pada akurasi proses seleksi di tingkat satuan pendidikan. Di SD Negeri X yang berlokasi di Kecamatan Terbanggi Besar, Kabupaten Lampung Tengah, proses penentuan penerima bantuan hingga saat ini masih menghadapi kendala teknis. Seleksi dilakukan secara konvensional dengan memeriksa berkas pengajuan siswa satu per satu berdasarkan kriteria tertentu. Mekanisme manual ini tidak hanya memakan waktu yang lama, tetapi juga memiliki risiko subjektivitas yang tinggi dalam pengambilan keputusan, mengingat tidak adanya pola sistematis yang baku dalam pengolahan data siswa dari kelas 1 hingga kelas 6.

Penerapan teknik *data mining* menawarkan solusi untuk mengubah data mentah menjadi informasi strategis guna mendukung sistem pengambilan keputusan. Salah satu algoritma klasifikasi

yang dapat diandalkan adalah Naïve Bayes. Algoritma ini bekerja berdasarkan teori probabilitas yang sangat efisien dalam menangani dataset besar dan memiliki kemampuan klasifikasi yang cepat. Dengan menerapkan Naïve Bayes, proses seleksi penerima bantuan dapat dilakukan secara otomatis berdasarkan pola data historis, sehingga penilaian kelayakan menjadi lebih objektif dan konsisten.

Penelitian ini bertujuan untuk menganalisis performa algoritma Naïve Bayes dalam mengklasifikasikan kelayakan penerima bantuan PIP di SD Negeri X. Guna memastikan kualitas hasil klasifikasi, performa model akan diuji menggunakan instrumen *confusion matrix* untuk menghasilkan nilai evaluasi yang komprehensif, meliputi tingkat *accuracy*, *precision*, *recall*. Melalui pendekatan ini, diharapkan proses distribusi bantuan menjadi lebih tepat sasaran, transparan, dan akurat berdasarkan parameter-parameter yang telah ditentukan.

KAJIAN PUSTAKA DAN LANDASAN TEORI

Ringkasan Penelitian

Berikut adalah ringkasan beberapa penelitian terdahulu yang berkaitan dengan pemberian beasiswa, sebagaimana disajikan pada tabel di bawah ini:

Tabel 1 Ringkasan Penelitian

No	Penulis (Tahun)	Algoritma	Hasil Akurasi
1.	(Fitri Nuril Halizah and Kartikasari 2024)	Naive Bayes	98,89%
2.	Muliawan Nur et al., 2024	Naïve Bayes	92,81%
3.	Hidayat et al., 2023	Naive Bayes	91,67%
4	Wahyuningsih et al., 2022	Naïve Bayes	90,48%
5	Muhammad et al., 2025	Naïve Bayes	77,11%

Beberapa penelitian terdahulu telah membuktikan bahwa algoritma Naïve

Bayes memiliki efektivitas yang sangat tinggi dalam melakukan klasifikasi penerima bantuan pendidikan atau beasiswa. Berdasarkan hasil studi yang dilakukan oleh Fitri Nuril Halizah & Kartikasari, (2024), penggunaan metode ini mampu menghasilkan tingkat akurasi yang sangat signifikan mencapai 98,89%. Hal ini menunjukkan bahwa pendekatan probabilitas dalam Naïve Bayes sangat handal untuk menangani kriteria seleksi yang kompleks.

Sejalan dengan hasil tersebut, Muliawan Nur et al., (2024) dan (Hidayat et al., (2023) juga menerapkan algoritma yang sama dengan perolehan akurasi masing-masing sebesar 92,81% dan 91,67%. Konsistensi nilai akurasi di atas 90% pada berbagai studi ini memperkuat alasan penggunaan Naïve Bayes sebagai solusi untuk meminimalisir kesalahan sasaran dalam distribusi bantuan.

Meskipun demikian, performa algoritma ini juga dipengaruhi oleh karakteristik dataset yang digunakan. Hal ini terlihat pada penelitian Muhammad et al., (2025) yang memperoleh akurasi sebesar 77,11%, serta Wahyuningsih et al., (2022) dengan akurasi 90,48%. Perbedaan hasil ini sering kali dipicu oleh kualitas data awal dan proses *preprocessing* yang dilakukan. Secara keseluruhan, keterkaitan antar penelitian ini menunjukkan bahwa Naïve Bayes tetap menjadi salah satu algoritma klasifikasi yang paling relevan dan objektif untuk mendukung sistem pengambilan keputusan dalam pengusulan kelayakan bantuan seperti Program Indonesia Pintar (PIP).

Program Indonesia Pintar PIP

Sebagai bentuk transformasi dari skema BSM terdahulu, Program Indonesia Pintar (PIP) hadir untuk meringankan beban finansial peserta didik dari keluarga prasejahtera melalui penyaluran bantuan

uang tunai sebagaimana diatur dalam Permendikbud No. 19 Tahun 2016. Fokus utama dari program ini tidak hanya terbatas pada pemenuhan kebutuhan biaya pendidikan semata, namun juga sebagai instrumen pemerintah dalam menekan angka putus sekolah serta memfasilitasi kembali layanan pendidikan bagi anak-anak yang sebelumnya sempat terhenti dari proses pembelajaran. (Edrial, Putrama A, and Sujastiawan 2022)

Program Indonesia Pintar (PIP) sejatinya dirancang sebagai instrumen dukungan finansial bagi peserta didik dari kalangan prasejahtera guna menjamin keberlangsungan pendidikan mereka. Namun dalam praktiknya, distribusi bantuan ini masih menghadapi tantangan akurasi sasaran, di mana ditemukan oknum penerima manfaat yang justru berasal dari latar belakang ekonomi berkecukupan. (Andri and Suparmadi 2022)

Penerima bantuan PIP mencakup siswa yang memegang Kartu Indonesia Pintar (KIP) serta siswa dengan kriteria khusus, seperti:

- Keluarga peserta Program Keluarga Harapan (PKH).
- Keluarga pemegang Kartu Keluarga Sejahtera (KKS).
- Siswa dengan status yatim piatu, korban bencana alam, atau penyandang disabilitas.
- Siswa dari keluarga yang mengalami pemutusan hubungan kerja (PHK).

Data Mining

Penggalian data atau *data mining* merupakan sebuah metodologi yang mengekstraksi pola-pola tersembunyi dari sekumpulan informasi berskala besar. Melalui proses ini, pengetahuan yang sebelumnya tidak terdeteksi dapat diidentifikasi untuk kemudian dijadikan landasan strategis dalam proses

penentuan kebijakan atau pengambilan keputusan. (Sari et al. 2025)

Induction-based learning merupakan sebuah proses di mana definisi umum dikembangkan melalui analisis terhadap sampel-sampel nyata yang sifatnya khusus. Alih-alih berangkat dari teori besar, metode ini bekerja dengan cara mengumpulkan fakta-fakta spesifik terlebih dahulu guna membentuk sebuah kesimpulan atau konsep yang bersifat general. (Simanjuntak Priscila 2024)

Data mining adalah ekstraksi pola yang menarik dari besarnya jumlah data. (Ramadanti and Muslih 2022)

Penambangan data atau *data mining* dapat dipahami sebagai sebuah rangkaian prosedur untuk menggali wawasan bermakna dan pola-pola intelegen dari himpunan informasi berskala besar. Dengan mengandalkan sinergi antara metode statistik, kecerdasan buatan, dan algoritma pembelajaran mesin, teknologi ini berupaya menyingkap korelasi yang tidak tampak guna memperkuat basis pengambilan keputusan di berbagai sektor. Operasionalnya mencakup persiapan data secara menyeluruh melalui tahap prapemrosesan hingga tahap interpretasi temuan, yang mana sangat aplikatif dalam dunia bisnis maupun kesehatan untuk mendeteksi tren perilaku masa depan. (Anwar Setia Mubarakah, Bambang Irawan 2025)

Klasifikasi

Klasifikasi merupakan tahapan krusial dalam data mining yang berfungsi melabeli objek baru berdasarkan analisis variabel data yang ada melalui tiga fase utama yakni konstruksi model menggunakan data pelatihan sebagai contoh aplikasi model serta evaluasi guna memprediksi kelas dari item yang belum diketahui sebelumnya. (Pebdika et al. 2023)

Sebagai metode *supervised learning* yang berupaya menemukan korelasi antara atribut masukan dan target, klasifikasi berfungsi memetakan objek data ke dalam label kelas yang tersedia melalui proses pelatihan fungsi target berdasarkan data latih yang ada guna memprediksi kategori sampel baru secara akurat. Meskipun bertujuan meningkatkan keandalan hasil melalui berbagai model seperti *k-Nearest Neighbor*, *Naive Bayes*, *Support Vector Machine*, hingga *Random Forest*, kinerja sistem ini tetap harus diuji menggunakan matriks konfusi karena hasil klasifikasi tidak selalu mencapai akurasi sempurna 100 persen. (Andri and Suparmadi 2022)

Pendekatan klasifikasi dalam bidang *data mining* menjadi opsi strategis untuk menyelesaikan kendala yang ada dengan cara mengekstraksi pola dari informasi terdahulu yang menggunakan variabel target sebagai parameter utama dalam melakukan kategorisasi objek secara akurat. (Albert, Lobo, and Talakua 2024)

Knowledge Discovery in Database

Penelitian ini mengimplementasikan kerangka kerja *Knowledge Discovery in Database* (KDD) yang merupakan standar prosedur dalam penambangan data untuk mengekstraksi wawasan serta pola-pola strategis yang terkandung di dalam dataset. Pendekatan KDD ini diaplikasikan melalui serangkaian tahapan sistematis yang dimulai dari seleksi data, prapemrosesan, dan transformasi informasi, yang kemudian dilanjutkan dengan fase inti penambangan data (*data mining*) serta diakhiri dengan tahap interpretasi atau evaluasi hasil. (Sitanggang, Umaidah, and Adam 2024)

Algoritma Naive Bayes

Demi menjamin keadilan dan keterbukaan dalam penyaluran bantuan pendidikan,

pemerintah perlu merumuskan kriteria seleksi yang baku sekaligus memanfaatkan sistem klasifikasi berbasis Naive Bayes untuk memetakan peluang siswa secara objektif. Langkah ini tidak hanya bertujuan untuk memastikan ketepatan sasaran bagi mereka yang membutuhkan, tetapi juga sebagai upaya standarisasi agar proses evaluasi data siswa terhindar dari praktik diskriminatif. (Yudha Prawira, Pratama, and Yanti 2024)

Metode Naïve Bayes Classifier diterapkan dalam penelitian ini sebagai algoritma klasifikasi utama. Algoritma ini merupakan teknik dalam *data mining* yang mengandalkan prinsip probabilitas dan pendekatan statistik untuk memproses data. (Samsu and Amnia 2023)

Pendekatan dari Teorema Naïve Bayes adalah sebagai berikut :

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Keterangan :

$P(C_i|X)$: Merepresentasikan probabilitas dokumen X untuk diklasifikasikan ke dalam kategori C_i

$P(X|C_i)$: Menunjukkan peluang munculnya sekumpulan kata dari dokumen X dalam lingkup kategori C_i

$P(C_i)$: Merupakan probabilitas awal (*prior*) dari suatu kategori dibandingkan dengan seluruh kategori yang tersedia.

$P(X)$: Peluang spesifik dari dokumen terkait. Dalam implementasinya, nilai ini sering kali diabaikan karena bersifat konstan, sehingga tidak memengaruhi perbandingan antar kategori.,

Confusion Matrik

Dalam mengevaluasi efektivitas model klasifikasi, *Confusion Matrix* sering dimanfaatkan sebagai instrumen

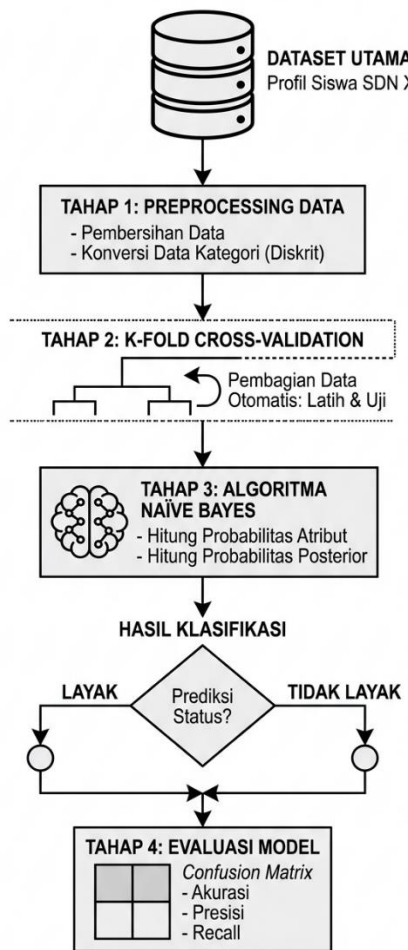
pengukuran performa. Melalui perbandingan antara label yang diprediksi dengan data sebenarnya, matriks ini mampu membedah distribusi kesalahan serta menyajikan gambaran komprehensif mengenai ketepatan hasil klasifikasi. (Putri, Jasmir, and Purnama 2025)

Evaluasi performa model klasifikasi dilakukan melalui *confusion matrix*, yang memetakan perbandingan antara nilai prediksi dengan nilai aktual. Matriks ini terdiri dari empat indikator utama: *true positive*, *true negative*, *false positive*, dan *false negative*. Guna mengukur efektivitas metode yang diterapkan secara komprehensif, data dari matriks tersebut kemudian diolah menjadi nilai metrik evaluasi yang meliputi akurasi, presisi, *recall*, serta *f1-score*. (Nur and Harianto 2022)

METODE

Pendekatan penelitian ini mengadopsi metodologi *Knowledge Discovery in Database* (KDD) guna mengekstraksi pengetahuan dari basis data. Rangkaian sistematis yang dilakukan dalam kerangka KDD pada penelitian ini meliputi:

1. **Seleksi Data**: Mengisolasi subset data yang relevan dari sumber informasi utama.
2. **Pra-pemrosesan**: Membersihkan data dari anomali atau *noise* untuk menjaga kualitas informasi.
3. **Transformasi Data**: Mengonversi format data ke dalam bentuk yang siap diolah oleh algoritma.
4. **Penambangan Data (Data Mining)**: Mengaplikasikan teknik cerdas untuk menemukan pola tersembunyi.
5. **Evaluasi**: Menginterpretasikan hasil temuan untuk memastikan validitas dan kegunaannya.



Gambar 1 Metode Penelitian

Beberapa proses tersebut dapat dijelaskan sebagai berikut :

Pengumpulan Dataset

Dalam penelitian ini, sumber data diperoleh dari PIP di SD Negeri X, dengan jumlah data sebanyak 476 Data dengan Atribut Data Sebagai Berikut :

Tabel 2 Pengumpulan Dataset

No	Nama Atribut	Kategori
1	Penghasilan Orang Tua (Ayah/Ibu)	1: Rp 0; 2: < 1,5jt; 3: 1,5jt - 3,9jt; 4: 4jt
2	Pekerjaan Ayah/Ibu	1: Tidak Bekerja 2: Buruh 3: Karyawan BUMN 4: Karyawan Honorer 5: Karyawan Swasta 6: Nelayan 7: Pedagang Besar 8: Pedagang Kecil

		9: Petani 10: PNS/TNI/Polri 11: Sopir 12: Wiraswasta 13: Wirausaha 14: Sudah Meninggal
3	Jumlah Anak	1: >2 2: <2
4	Alat Transportasi	1: Jalan kaki 2: Sepeda 3: Sepeda motor 4: Abudemen 5: Mobil Pribadi
5	Layak PIP (Label)	Ya; Tidak

Penelitian ini menggunakan variabel kunci yang menentukan prioritas penerima bantuan PIP di SD Negeri X, sebagaimana dirinci dalam Tabel 2.1. Untuk mengoptimalkan kinerja algoritma Naïve Bayes, seluruh data mentah ditransformasikan ke dalam kategori numerik. Variabel Penghasilan Orang Tua diklasifikasikan menjadi empat level angka, mulai dari kode 1 untuk penghasilan Rp 0, hingga kode 4 untuk penghasilan di atas 4 juta rupiah. Penggunaan kode numerik ini memungkinkan sistem melakukan perhitungan probabilitas pada tiap rentang ekonomi secara lebih akurat. Pada atribut Pekerjaan Orang Tua, dilakukan pelabelan angka dari skala 1 hingga 14 yang mencakup berbagai sektor profesi, mulai dari tidak bekerja dan buruh, hingga PNS/TNI/Polri serta status sudah meninggal. Atribut pendukung lainnya, yaitu Jumlah Anak, dikategorikan menjadi kode 1 (>2 anak) dan kode 2 (<2 anak). Demikian pula dengan variabel Alat Transportasi yang dikonversi menjadi skala numerik 1 sampai 5, yang merepresentasikan tingkat kepemilikan aset transportasi siswa, mulai dari jalan kaki hingga penggunaan mobil pribadi. Dalam proses pengolahan menggunakan RapidMiner, setiap atribut numerik ini dihitung bobot peluangnya secara independen terhadap label keputusan, yaitu 'Ya' atau 'Tidak' (Layak PIP).

Tahap *preprocessing* dilakukan terlebih dahulu untuk memastikan tidak ada data yang kosong (*missing values*) yang dapat mengganggu hasil klasifikasi. Melalui transformasi ke dalam format numerik yang terstruktur, proses seleksi penerima bantuan di SD Negeri X menjadi lebih sistematis, terukur, dan terhindar dari penilaian subjektif petugas.

Persiapan Data

Setelah data terkumpul, dilakukan proses persiapan data (*preprocessing*). Pada tahap ini, data dibersihkan dari nilai yang kosong (*missing values*) maupun data yang tidak konsisten. Selain itu, dilakukan transformasi data dengan mengubah nilai nominal menjadi kategori numerik melalui proses pelabelan angka (misalnya, mengonversi rentang gaji menjadi representasi angka tertentu) agar seluruh data siap diolah secara optimal oleh algoritma klasifikasi.

K-Fold Validation

Untuk menjamin validitas dan reliabilitas hasil pengujian, penelitian ini menerapkan metode *K-Fold Cross Validation*. Dataset akan dibagi ke dalam sejumlah *k* bagian (fold) secara acak. Dalam setiap iterasi, satu bagian akan digunakan sebagai data uji (*testing data*), sementara bagian sisanya digunakan sebagai data latih (*training data*). Penggunaan metode ini bertujuan untuk memastikan model Naïve Bayes yang dihasilkan tetap stabil, terhindar dari bias pemilihan data, dan tidak mengalami *overfitting*.

Klasifikasi Algoritma Naïve Byes

Data yang telah divalidasi kemudian diproses menggunakan algoritma Naïve Bayes Classifier. Algoritma ini dipilih karena kemampuannya dalam memprediksi peluang di masa depan berdasarkan pengalaman pada data

historis. Proses klasifikasi dilakukan dengan menghitung probabilitas setiap kriteria (seperti penghasilan orang tua, status siswa, dan kepemilikan kartu bantuan) terhadap kelas Layak atau Tidak Layak. Dengan menggunakan pendekatan statistik ini, sistem dapat memberikan keputusan yang objektif berdasarkan bobot probabilitas tertinggi dari setiap atribut yang tersedia di SD Negeri X.

Evaluasi

Tahap akhir dalam penelitian ini adalah Evaluasi untuk mengukur sejauh mana performa algoritma Naïve Bayes dalam melakukan klasifikasi. Evaluasi dilakukan dengan alat ukur Confusion Matrix. Dari matriks ini, akan diperoleh nilai parameter berupa Akurasi (tingkat kedekatan prediksi dengan data asli), Presisi (tingkat ketepatan prediksi pada kelas tertentu), dan Recall (tingkat keberhasilan model dalam mengenali kembali data). Hasil evaluasi ini akan menjadi pembuktian apakah algoritma Naïve Bayes efektif digunakan sebagai standar sistem pendukung keputusan bagi pihak sekolah.

HASIL DAN PEMBAHASAN

Pada bab ini dipaparkan hasil dari penerapan algoritma Naïve Bayes dalam mengklasifikasikan kelayakan penerima bantuan PIP di SD Negeri X. Proses pengujian dilakukan menggunakan perangkat lunak RapidMiner.

Deskripsi Dataset

Dataset yang digunakan berjumlah 476 data siswa dengan 6 atribut kriteria dan 1 atribut label (Kelayakan). Sebelum masuk ke algoritma, data telah dipastikan bersih melalui tahap *preprocessing*.

Persiapan Data

Tahap persiapan data merupakan langkah dilakukan setelah pengumpulan data (Tanti Cahya Herdiyani 2022) untuk memastikan

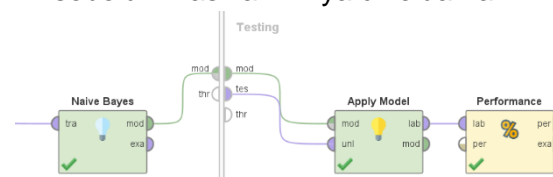
Proses pemodelan dilakukan dengan menyusun rangkaian operator secara sistematis pada perangkat lunak RapidMiner sebagaimana disajikan pada Gambar 5 Alur kerja dalam rangkaian ini dirancang untuk memastikan data diproses secara objektif sebelum dilakukan pengujian. Berikut adalah fungsi dari masing-masing operator tersebut:

Adapun tahapan implementasi pada RapidMiner adalah sebagai berikut:

1. *Read Excel*: Berfungsi sebagai gerbang utama untuk memasukkan dataset siswa SD Negeri X yang telah disimpan dalam format .x/sx ke dalam lingkungan kerja RapidMiner.
2. *Set Role*: Operator ini digunakan untuk menetapkan atribut "Kelayakan PIP" sebagai label. Penetapan ini krusial karena algoritma Naïve Bayes memerlukan target sasaran (kunci jawaban) untuk diprediksi.
3. *Select Attributes*: Tahap ini merupakan bagian dari optimasi model. Berdasarkan analisis probabilitas sebelumnya, operator ini digunakan untuk melakukan seleksi fitur dengan hanya menyertakan atribut-atribut yang memiliki pengaruh signifikan dan mengeliminasi atribut yang bersifat *noise* (seperti Jarak Rumah, Disabilitas, dan Status Siswa).
4. *Cross Validation*: Merupakan operator inti untuk pengujian. Di dalam operator ini, data di menggunakan metode 10-Fold Cross Validation, di mana algoritma Naïve Bayes dilatih dan diuji secara bergantian pada sepuluh lipatan data yang berbeda guna menghasilkan nilai akurasi, presisi, dan recall yang stabil dan valid.

Sub-Proses Training dan Testing pada Cross Validation

Di dalam operator *Cross Validation*, proses klasifikasi dibagi menjadi dua tahapan utama yang saling terhubung, sebagaimana ditunjukkan pada Gambar 6 Tahapan ini memastikan bahwa model yang dibangun telah diuji secara objektif sebelum hasil akhirnya dikeluarkan.



Gambar 6 Sub-Proses Training dan Testing pada Cross Validation

1. *Sisi Training (Kiri):*

Pada panel ini, data latih dialirkan menuju operator Naïve Bayes. Di sinilah algoritma mempelajari pola probabilitas dari dataset siswa SD Negeri X berdasarkan kriteria yang telah ditentukan. Hasil dari proses ini adalah sebuah Model Klasifikasi (ditandai dengan garis hijau *mod*) yang berisi aturan-aturan statistik untuk menentukan kelayakan PIP.

2. *Sisi Testing (Kanan):*

Model yang dihasilkan dari tahap *training* kemudian diteruskan ke operator Apply Model. Di sini, model tersebut digunakan untuk menebak atau memprediksi data uji (*testing data*) yang belum pernah dilihat sebelumnya.

3. *Evaluasi Performa:*

Hasil prediksi dari *Apply Model* kemudian dibandingkan dengan data asli melalui operator Performance. Dari perbandingan inilah sistem secara otomatis menghitung nilai Akurasi, Presisi, dan Recall. Penggunaan alur ini menjamin bahwa angka Recall

69% yang diperoleh merupakan hasil pengujian yang valid dan dapat dipertanggungjawabkan secara ilmiah.

Analisis Distribusi Probabilitas Naïve Bayes

Berdasarkan hasil pengolahan data menggunakan algoritma Naïve Bayes, diperoleh nilai parameter statistik berupa rata-rata (*mean*) dan simpangan baku (*standard deviation*) untuk setiap atribut kriteria. Nilai-nilai ini mencerminkan karakteristik sebaran data pada masing-masing kelas kelayakan, sebagaimana disajikan dalam Tabel 5.

Tabel 3. Distribusi Parameter Statistik Naïve Bayes

Penghasilan Ayah	<i>standard deviation</i>	0.64	0.44
Penghasilan Ibu	<i>standard deviation</i>	0.51	1.04
Pekerjaan Ayah	<i>standard deviation</i>	4.44	2.78
Pekerjaan Ibu	<i>standard deviation</i>	2.79	4.41
Jumlah Anak	<i>standard deviation</i>	0.46	0.50
Alat Transportasi	<i>standard deviation</i>	0.80	0.56
Penghasilan Ayah	<i>mean</i>	2.58	3.06
Penghasilan Ibu	<i>mean</i>	1.19	1.66
Pekerjaan Ayah	<i>mean</i>	8.07	10.64
Pekerjaan Ibu	<i>mean</i>	2.05	3.80
Jumlah Anak	<i>mean</i>	1.29	1.55
Alat Transportasi	<i>mean</i>	2.71	2.87

Secara teoritis, nilai simpangan baku (*standard deviation*) yang rendah

menunjukkan bahwa data pada atribut tersebut memiliki tingkat variasi yang kecil atau cenderung homogen. Sebagai contoh, pada atribut Penghasilan Ayah, diperoleh nilai simpangan baku sebesar 0,64 untuk kelas 'Layak' dan 0,44 untuk kelas 'Tidak Layak'. Hal ini mengindikasikan bahwa data penghasilan pada dataset siswa SD Negeri X memiliki pola yang cukup konsisten, sehingga algoritma Naïve Bayes dapat lebih mudah menentukan batas probabilitas kelayakan berdasarkan kriteria tersebut.

Sebaliknya, pada atribut Pekerjaan Ayah dan Ibu, nilai simpangan baku terlihat lebih tinggi (berkisar antara 2,78 hingga 4,44). Hal ini menunjukkan adanya variasi jenis pekerjaan yang cukup beragam di kalangan orang tua siswa, yang memberikan tantangan tersendiri bagi model dalam menghitung bobot probabilitasnya. Namun, melalui kombinasi seluruh atribut yang ada, sistem tetap mampu menghasilkan prediksi dengan tingkat akurasi yang optimal. Analisis terhadap parameter distribusi ini membuktikan bahwa faktor ekonomi keluarga merupakan variabel kunci yang mendominasi keputusan akhir sistem dalam menentukan penerima bantuan PIP."

Evaluasi Hasil

Tahap evaluasi merupakan langkah krusial untuk mengukur sejauh mana algoritma Naïve Bayes mampu melakukan klasifikasi terhadap kelayakan penerima bantuan PIP secara akurat. Pengujian dilakukan dengan metode *10-Fold Cross Validation* guna memastikan hasil yang diperoleh bersifat stabil dan terhindar dari bias data. Hasil evaluasi performa model disajikan dalam bentuk *Confusion Matrix* pada Tabel 6 berikut:

Tabel 4 Evaluasi Hasil

	True YA	True Tidak	Class Precision
Pred. YA	59	51	53,64%
Pred. Tidak	53	315	85,60%
Class Recall	52,68%	86,07%	78,20% (Accuracy)

Berdasarkan Tabel 4. diperoleh hasil evaluasi performa sebagai berikut:

1. **Akurasi (Accuracy):**
 Model menghasilkan tingkat akurasi sebesar 78,20%. Hal ini menunjukkan bahwa secara keseluruhan, algoritma Naïve Bayes memiliki kemampuan yang baik dalam memprediksi status kelayakan siswa dengan benar pada sebagian besar dataset.
2. **Presisi (Precision):**
 Untuk kelas "YA" (Layak), diperoleh nilai presisi sebesar 53,64%. Nilai ini menunjukkan tingkat ketepatan sistem dalam mengidentifikasi siswa yang benar-benar layak menerima bantuan dari total prediksi layak yang dihasilkan.
3. **Recall:**
 Nilai *recall* untuk kelas "YA" berada pada angka 52,68%. Hal ini mengindikasikan bahwa sistem berhasil menjaring lebih dari separuh jumlah siswa yang secara nyata berhak menerima PIP berdasarkan kriteria yang telah ditetapkan.

Meskipun terdapat tantangan dalam menyeimbangkan nilai presisi dan *recall* pada kelas minoritas (YA), pencapaian akurasi sebesar 78,20% membuktikan bahwa sistem pendukung keputusan ini jauh lebih efektif dan objektif dibandingkan proses seleksi manual. Sistem berhasil mengklasifikasikan 315 siswa pada kelas 'TIDAK' secara akurat, yang secara signifikan dapat membantu pihak sekolah dalam mereduksi beban kerja verifikasi berkas dan meminimalisir kesalahan manusia (*human error*) dalam penentuan penerima bantuan PIP."

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai penerapan algoritma Naïve Bayes untuk klasifikasi kelayakan penerima bantuan Program Indonesia Pintar (PIP) di SD Negeri X, dapat ditarik beberapa kesimpulan sebagai berikut:

Efektivitas Model: Algoritma Naïve Bayes terbukti efektif dalam melakukan klasifikasi kelayakan bantuan PIP dengan menghasilkan tingkat Akurasi sebesar 78,20%. Hasil ini menunjukkan bahwa sistem memiliki performa yang handal dalam mengolah data ekonomi dan sosial siswa untuk mendukung pengambilan keputusan.

Kinerja Klasifikasi: Melalui pengujian menggunakan *Confusion Matrix*, sistem berhasil memprediksi secara akurat sebanyak 315 data pada kelas "Tidak Layak" dan 59 data pada kelas "Layak". Meskipun nilai *recall* (52,68%) dan presisi (53,64%) pada kelas "Layak" menunjukkan tantangan pada dataset yang tidak seimbang, secara keseluruhan sistem mampu mengurangi objektivitas seleksi manual.

Variabel Dominan: Faktor ekonomi, khususnya atribut Penghasilan dan Pekerjaan Orang Tua, merupakan variabel yang paling berpengaruh dalam perhitungan probabilitas pada model Naïve Bayes di SD Negeri X dibandingkan atribut pendukung lainnya.

Saran

Untuk pengembangan penelitian selanjutnya, disarankan beberapa hal sebagai berikut:

Melakukan penyeimbangan dataset (*data balancing*) menggunakan Teknik *oversampling* atau *under sampling* untuk meningkatkan nilai *recall* dan *precision* pada kelas minoritas (siswa layak).

Mengombinasikan algoritma Naïve Bayes dengan metode seleksi fitur berbasis optimasi seperti *Particle Swarm Optimization* (PSO) atau *Genetic Algorithm* (GA) untuk memperoleh akurasi yang lebih maksimal.

Menambah jumlah dataset dan atribut kriteria yang lebih spesifik guna mempertajam kemampuan model dalam mengenali pola data siswa yang lebih beragam.

REFERENSI

- [1]. Albert, Murry, Agustin Lobo, and Alfrian Carmen Talakua. 2024. "Klasifikasi Data Penerima Beasiswa Menggunakan Algoritma C4 . 5." *Jurnal Penelitian Inovatif (JUPIN)* 4(2): 575–82.
- [2]. Andri, Nata, and Suparmadi. 2022. "ANALISIS SISTEM PENDUKUNG KEPUTUSAN DENGAN MODEL KLASIFIKASI BERBASIS MACHINE LEARNING DALAM PENENTUAN PENERIMA PROGRAM." *Journal of Science and Social Research* 3(3): 697–702.
- [3]. Anwar Setia Mubarakah, Bambang Irawan, Willy Prihartono. 2025. "PENERAPAN ALGORITMA NAÏVE BAYES UNTUK." *Jurnal Informatika dan Teknik Elektro Terapan* 13(3): 2295–2305.
- [4]. Edrial, Rangga Putrama A, and Ade Sujastiawan. 2022. "EVALUASI KEBIJAKAN PROGRAM INDONESIA PINTAR (PIP) DI SMA NEGERI 1 UTAN TAHUN 2019-2020." : 109–16.
- [5]. Fitri Nuril Halizah, Cendikia, and Meivi Kartikasari. 2024. "IMPLEMENTASI ALGORITMA NAÏVE BAYES DALAM PENENTUAN PEMBERIAN BEASISWA KIP KULIAH (STUDI KASUS STIKI MALANG) IMPLEMENTATION OF NAÏVE BAYES ALGORITHM IN DETERMINING KIP – KULIAH SCHOLARSHIP (CASE STUDY." *Jurnal Simantec* 12(2): 1–6.
- [6]. Hidayat, Muslim et al. 2023. "STUDI KOMPARASI ALGORITMA NAÏVE BAYES DAN K-NN UNTUK KLASIFIKASI PENERIMAAN BEASISWA DI MI AL – ISLAMİYAH." *Jurnal Ilmiah Teknik dan Ilmu Komputer* 2(4): 172–80.
- [7]. Muhammad, Thoriq, Maulana Fajar, Nova Eirlangga Yofhanda Septi Hayati, and Muhammad Ashim Madani. 2025. "Implementasi Algoritma Naïve Bayes Dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP Di Universitas Adzkia." *JURNAL FASILKOM* 15(1): 108–14.
- [8]. Muliawan Nur, Amri, Weni Nursali, Nurhidayati, and Fathurrahman Imam. 2024. "Penerapan Metode Naïve Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar." *Jurnal Informatika dan Teknologi* 7(1): 93–102.
- [9]. Nur, Fajriyan, and Wahyudi Harianto. 2022. "KOMPARASI TINGKAT AKURASI INFORMATION GAIN DAN GAIN RATIO PADA METODE K-NEAREST NEIGHBOR." 6(1): 386–91.
- [10]. Pebdika, Angga et al. 2023. "KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP." *Jurnal Mahasiswa Teknik Informatika* 7(1): 452–58.
- [11]. Putri, Azhiah, Jasmir, and Benni Purnama. 2025. "Classification of Scholarship Eligibility Using Naïve

- Bayes with Attribute Optimization Based on K-Means Clustering Klasifikasi Kelayakan Penerima Beasiswa Menggunakan Naive Bayes Dengan Optimasi Atribut Berbasis K-Means Clustering.” *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 5(October): 1450–62.
- [12]. Rahmadani, Rafi et al. 2025. “ANALISIS SENTIMEN ULASAN ‘ OJOL THE GAME ’ DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN MODEL EKSTRAKSI FITUR TF-IDF.” *Jurnal Informatika dan Teknik Elektro Terapan* 12(3).
- [13]. Ramadanti, Elsa, and Muhamad Muslih. 2022. “PENERAPAN DATA MINING ALGORITMA K-MEANS CLUSTERING PADA POPULASI AYAM PETELUR DI INDONESIA.” *Jurnal Teknologi dan Sistem Informasi Univrab* 7(1): 1–7.
- [14]. Samsu, L M, and Wajizatul Amnia. 2023. “Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi).” 6(1).
- [15]. Sari, Betha Nurina et al. 2025. “PENERAPAN DATA MINING DALAM KLASIFIKASI DATA TRANSAKSI PRODUK KOPERASI DI SMK PGRI 2 KARAWANG.” *Jurnal Mahasiswa Teknik Informatika* 9(1): 263–69.
- [16]. Simanjuntak Priscila, Lidya. 2024. “Penerapan Data Mining Dalam Proses Hukum Pidana Bagi Pelaku.” *ADA Journal of Information System Research* 1(2): 79–92.
- [17]. Sitanggang, Altolyto, Yuyun Umaidah, and Riza Ibnu Adam. 2024. “ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN SIANG GRATIS PADA MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYES.” *JITET (Jurnal Informatika dan Teknik Elektro Terapan)* 12(3).
- [18]. Tanti Cahya Herdiyani, Achmad Udin Zailani. 2022. “Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia.” 3(2): 154–65.
- [19]. Wahyuningsih, Budiman, and Izzatul Umami. 2022. “IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK MENENTUKAN CALON PENERIMA BEASISWA DI SMK YPM 14 SUMOBITO JOMBANG.” 4(1): 446–54.
- [20]. Yudha Prawira, Dimas, Yovi Pratama, and Elvi Yanti. 2024. “Analisis Dan Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerimaan Beasiswa PIP (Studi Kasus : SMPN 7 Kota Jambi).” *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)* 4(September): 1051–59.