JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i1.5882

ALGORITMA NAIVE BAYES UNTUK MENINGKATKAN MODEL KLASIFIKASI PENERIMA PROGRAM INDONESIA PINTAR DI SDN 2 PURWAWINANGUN

Luthvi Nurfauzi Darussalam[1*], Rudi Kurniawan[2], Yudhistira Arie Wijaya[3], Tati Suprapti[4]

[1,2]STMIK IKMI CIREBON; Jl. Perjuangan No. 10B Majasem Kec. Kesambi Kota Cirebon Tlp. 0231-490480-490481
[3]STMIK IKMI CIREBON; Tlp. 0231-490480-490481

Received: 30 Desember 2024
Accepted: 14 Januari 2025
Published: 20 Januari 2025

Keywords:
Naive Bayes, Program Indonesia Pintar (PIP),dan Kalasifikasi

Corespondent Email:
nluthvi@gmail.com

Penelitian ini bertujuan meningkatkan keakuratan klasifikasi penerima Program Indonesia Pintar (PIP) di SDN 2 Purwawinangun, Kabupaten Kuningan. Metode konvensional yang lambat dan kurang akurat digantikan dengan algoritma Naive Bayes untuk menganalisis data siswa berdasarkan kriteria tertentu. Proses penelitian meliputi pengumpulan data sekunder, preprocessing data, dan implementasi algoritma Naive Bayes. Hasilnya, model ini mencapai akurasi 96,47% dalam menentukan kelayakan penerima PIP, dengan mempertimbangkan atribut seperti latar belakang sosial ekonomi dan performa akademik siswa. Temuan menunjukkan bahwa algoritma ini efisien dalam mengolah dataset kompleks dibandingkan metode manual. Namun, performa model sangat bergantung pada kualitas data awal, sehingga data yang tidak lengkap dapat memengaruhi hasil.Penelitian merekomendasikan penerapan metode ini di sekolah lain dan integrasi algoritma tambahan, seperti Decision Tree, untuk validasi hasil. Dengan pendekatan ini, seleksi penerima PIP menjadi lebih tepat sasaran, efisien, dan transparan.

This study aims to improve the accuracy of classifying recipients of the Indonesia Smart Program (PIP) at SDN 2 Purwawinangun, Kuningan Regency. Conventional methods, which are slow and less accurate, are replaced with the Naive Bayes algorithm to analyze student data based on specific criteria. The research process includes collecting secondary data, preprocessing data, and implementing the Naive Bayes algorithm. The results show that the model achieves 94.64% accuracy in determining PIP recipient eligibility, considering attributes such as socioeconomic background and students' academic performance. The findings indicate that this algorithm is efficient in processing complex datasets compared to manual methods. However, the model's performance heavily depends on the quality of the initial data, as incomplete data can affect classification results. The study recommends applying this method to other schools and integrating additional algorithms, such as Decision Trees, for result validation. This approach ensures that the selection of PIP recipients becomes more targeted, efficient, and transparent.

1. PENDAHULUAN

Penelitian ini membahas penerapan algoritma *Naive Bayes* untuk meningkatkan akurasi klasifikasi penerima bantuan sosial pendidikan di tingkat sekolah dasar. Naive Bayes, berdasarkan *Teorema Bayes*, dikenal karena kesederhanaan, efisiensi, dan kemampuannya menangani data besar dan kompleks [1].Hasil penelitian menunjukkan bahwa algoritma ini dapat mencapai akurasi tinggi, terutama dengan optimasi seperti *Laplace*

Smoothing dan *Particle Swarm Optimization* [2]dan[3]. Namun, tantangan utama meliputi pemilihan atribut yang relevan, asumsi independensi antar atribut yang sering tidak terpenuhi, pengolahan data tidak seimbang, serta perlakuan data kategorikal dan numerik yang memengaruhi hasil klasifikasi [4].Penelitian oleh [2]dan[5] menunjukkan bahwa Naive Bayes efektif untuk klasifikasi penerima bantuan sosial seperti PKH dan rehabilitasi sekolah, meskipun kompleksitas data tetap menjadi tantangan. Dibandingkan algoritma lain seperti C4.5, *Naive Bayes* unggul dalam kecepatan tetapi dapat dioptimalkan lebih lanjut untuk akurasi [6], Dalam konteks penerima bantuan sosial, jumlah penerima yang memenuhi syarat mungkin jauh lebih sedikit dibandingkan dengan yang tidak memenuhi syarat. Hal ini dapat menyebabkan model lebih cenderung untuk memprediksi kelas mayoritas, sehingga mengurangi akurasi untuk kelas minoritas [7], Penelitian ini diharapkan dapat memberikan wawasan baru bagi para peneliti dan praktisi di bidang informatika, serta membantu pengambil kebijakan dalam merumuskan strategi distribusi bantuan yang lebih efektif. Selain itu, hasil dari penelitian ini diharapkan dapat memberikan manfaat praktis, seperti peningkatan efisiensi dalam proses seleksi penerima bantuan, yang pada gilirannya dapat meningkatkan akses pendidikan bagi anak-anak di tingkat sekolah dasar. Dengan menggunakan algoritma Naive Bayes, diharapkan proses klasifikasi dapat dilakukan dengan lebih cepat dan akurat, sehingga bantuan sosial dapat disalurkan kepada mereka yang benar-benar membutuhkan [8]. Penelitian ini juga berpotensi untuk menjadi acuan bagi penelitian lebih lanjut dalam penerapan teknologi informasi dan komunikasi dalam sektor pendidikan dan sosial [9].Penelitian ini memberikan kontribusi pada pengembangan model klasifikasi yang lebih adaptif dan efisien, mendukung distribusi bantuan sosial yang lebih tepat sasaran, dan membuka peluang untuk eksplorasi lebih lanjut dalam integrasi teknologi informasi di sektor pendidikan [10].

2. TINJAUAN PUSTAKA

1.Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam mengklasifikasikan penduduk di Desa Tanoh Anou, Kecamatan Idi Rayeuk, berdasarkan status kesejahteraan mereka, yaitu miskin, rentan miskin, dan mampu. Dengan menggunakan data yang dikumpulkan dari wawancara langsung dan analisis data, penelitian ini mengidentifikasi faktor-faktor yang mempengaruhi kemiskinan, seperti luas lantai rumah, pendidikan, pekerjaan, dan penghasilan.Algoritma Naive Bayes menunjukkan akurasi sebesar 90% berdasarkan 30 kali percobaan. Hasil ini diperoleh dari perbandingan antara data aktual dan prediksi [11].

2.[12] meneliti penerapan algoritma *Naïve Bayes* untuk menentukan status gizi balita, sebuah permasalahan penting dalam kesehatan masyarakat. Algoritma ini digunakan untuk mengklasifikasikan data balita berdasarkan parameter seperti jenis kelamin, umur, berat badan, tinggi badan, dan gaji orang tua. Penelitian ini menguji berbagai rasio pembagian data (90:10, 80:20, 70:30, dan 60:40), dengan hasil akurasi optimal (100%) pada rasio 90:10, 80:20, dan 70:30. Namun, akurasi menurun menjadi 75% pada rasio 60:40.

Penelitian ini menunjukkan bahwa *Naïve Bayes* efektif dalam klasifikasi status gizi balita, terutama pada rasio pembagian data yang lebih tinggi untuk pelatihan. Kesimpulan ini mendukung efisiensi algoritma dalam membantu proses penentuan status gizi yang lebih cepat dan akurat di posyandu dan puskesmas.

3.[13] menyoroti pengaruh rasio pembagian data pelatihan dan pengujian terhadap kinerja model pre-trained dalam klasifikasi gambar. Penelitian ini menggunakan tiga model pre-trained—MobileNetV2, ResNet50v2, dan VGG19—dengan dataset yang terdiri dari 1000 gambar berimbang. Hasil menunjukkan bahwa rasio pembagian data lebih dari 70% untuk pelatihan (80:20 dan 90:10) menghasilkan kinerja optimal berdasarkan metrik sensitivitas, spesifisitas, dan akurasi.

Dataset sederhana memberikan hasil terbaik, menegaskan pentingnya kualitas data dalam memengaruhi kinerja model. Penelitian ini merekomendasikan eksplorasi dataset dan arsitektur model tambahan untuk memperluas pemahaman tentang pengaruh rasio pembagian data terhadap performa model

Penelitian pertama menunjukkan bahwa algoritma Naive Bayes efektif dalam mengklasifikasikan status kesejahteraan penduduk dengan tingkat akurasi sebesar 90% setelah 30 kali percobaan. Faktor-faktor seperti luas lantai rumah, pendidikan, pekerjaan, dan penghasilan terbukti memengaruhi kemiskinan.

Penelitian kedua mengonfirmasi efektivitas algoritma Naive Bayes dalam klasifikasi status gizi balita. Hasil menunjukkan akurasi optimal (100%) pada rasio pembagian data yang lebih tinggi untuk pelatihan (90:10, 80:20, dan 70:30) serta penurunan akurasi menjadi 75% pada rasio 60:40. Temuan ini menyoroti pentingnya pemilihan rasio data yang

tepat untuk memastikan hasil yang akurat dalam proses klasifikasi.

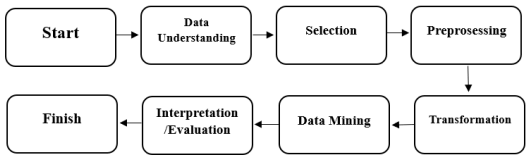
Penelitian ketiga menyoroti dampak rasio pembagian data terhadap kinerja model pre-trained dalam klasifikasi gambar. Rasio pembagian data pelatihan lebih besar dari 70% (80:20 dan 90:10) menghasilkan performa optimal berdasarkan sensitivitas, spesifisitas, dan akurasi. Selain itu, kualitas dataset terbukti menjadi faktor penting dalam meningkatkan hasil model.

Secara keseluruhan, ketiga penelitian ini menegaskan bahwa algoritma Naive Bayes dan model pre-trained memiliki potensi yang tinggi dalam klasifikasi data. Pemilihan rasio pembagian data yang optimal dan kualitas dataset yang baik menjadi faktor kunci dalam meningkatkan akurasi model. Temuan ini dapat menjadi dasar untuk penelitian lanjutan yang berfokus pada pengembangan dan penerapan algoritma pembelajaran mesin di berbagai bidang.

3. METODE PENELITIAN

3.1 Tahapan Metode Penelitian

Penelitian ini menggunakan metode kuantitatif dengan algoritma Naive Bayes untuk meningkatkan model klasifikasi penerima Program Indonesia Pintar (PIP) di Sekolah Dasar Negeri 2 Purwawinangun, Kabupaten Kuningan, metode penelitian kuantitatif adalah penelitian yang memanfaatkan data berupa angka yang dianalisis secara statistik untuk mengukur hubungan antar variabel secara objektif. Proses ini melibatkan pemodelan algoritmik untuk menghasilkan klasifikasi yang lebih akurat dan efisien.Penelitian ini memanfaatkan perangkat lunak RapidMiner sebagai alat bantu analisis data, yang memungkinkan pengolahan data secara sistematis melalui teknik Naive Bayes.Tahapan metode penelitian dapat di lihat pada gambar 3.1 berikut.



Gambar 3.1 Tahapan Metode Penelitian

Tahapan metode penelitian :

1. Pengumpulan Data
   Mengumpulkan data dari Sekolah Dasar Negeri 2 Purwawinangun dan instansi terkait, mencakup demografi siswa, ekonomi keluarga, dan akademik. Dataset tambahan diambil dari sumber sekunder.
2. Pemilihan Data
   Menyeleksi variabel penting seperti pendapatan keluarga dan jumlah tanggungan, yang relevan untuk klasifikasi penerima bantuan Program Indonesia Pintar (PIP).
3. Pemrosesan Data
   Membersihkan data dari nilai hilang, duplikasi, atau inkonsistensi. Data dinormalisasi untuk konsistensi dan siap dianalisis menggunakan RapidMiner.
4. Transformasi Data
   Mengonversi data kategorikal menjadi format numerik melalui diskritisasi dan encoding, agar kompatibel dengan algoritma Naive Bayes.
5. Data Mining
   Membuat model klasifikasi dengan algoritma Naive Bayes, melibatkan validasi silang dan evaluasi performa

3.2 Sumber Data

Penelitian menggunakan data primer dan sekunder untuk hasil yang komprehensif:

1. Data Primer

   Metode: Survei dan observasi langsung terhadap siswa SDN 2 Purwawinangun, Kabupaten Kuningan.

   Tujuan: Mendapatkan informasi faktual terkait kelayakan penerimaan Program Indonesia Pintar (PIP).

2. Data Sekunder

   Sumber: Artikel ilmiah, laporan penelitian, dataset publik dari Kemendikbud, dan jurnal bereputasi.

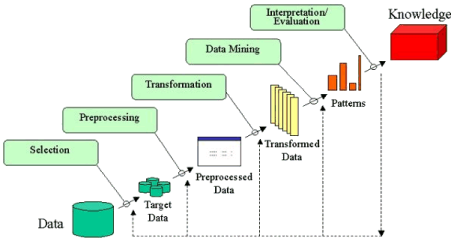
   Evaluasi: Data diuji validitas dan reliabilitasnya dengan memeriksa reputasi sumber dan kesesuaian dengan penelitian.

3.3 Teknik Pengumpulan Data

1. Observasi: Mengambil data calon penerima PIP di SDN 2 Purwawinangun.
2. Wawancara: Melibatkan wakil kepala sekolah dan pengurus sekolah untuk memperoleh informasi mendalam tentang calon penerima PIP.

3.4 Teknik Analisis Data

- Pendekatan: Kuantitatif dengan metode eksperimen komputasional.
- Tahapan: Melibatkan pengumpulan, pengolahan, analisis data, dan penerapan algoritma Naive Bayes.
- Tujuan: Menguji efektivitas algoritma Naive Bayes dalam meningkatkan akurasi klasifikasi kelayakan penerimaan PIP.
- Metode: Menggunakan tahapan Knowledge Discovery in Databases (KDD) sebagai kerangka kerja analisis.



Gambar 3.2 Tahapan KDD

Tahapan Perancangan KDD

1. Data Selection – Proses pemilihan data dari sekumpulan data operasional sebelum ekstraksi data. Data yang telah dipilih akan ditampilkan dalam satu halaman sebagai data operasional yang siap digunakan.

2. Preprocessing – Meliputi pembersihan data, penghapusan duplikasi, pengecekan inkonsistensi, dan koreksi kesalahan untuk memastikan data siap diproses lebih lanjut.
3. Transformation – Proses transformasi data melalui coding agar sesuai dengan kebutuhan analisis data mining. Tahap ini bersifat kreatif dan bergantung pada jenis data yang digunakan.
4. Data Mining – Menggunakan teknik atau algoritma tertentu untuk menemukan pola atau informasi yang menarik sesuai tujuan penelitian.
5. Interpretation/Evaluation – Mengevaluasi hasil data mining untuk memastikan informasi yang diperoleh mudah dimengerti dan relevan dengan kebutuhan analisis atau kebijakan yang akan diterapkan.

4. HASIL DAN PEMBAHASAN

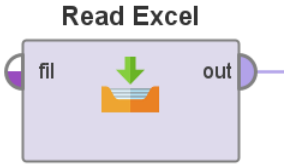
4.1 Hasil Penelitian

Penelitian ini membahas proses klasifikasi Program Indonesia Pintar (PIP) menggunakan metode Naïve Bayes untuk memperoleh nilai akurasi tertinggi dengan bantuan machine learning RapidMiner (Ai Studio 2024.1.0).

4.1.1 Data Understanding

Tahap awal dalam Knowledge Discovery in Database (KDD) adalah memahami data. Proses ini meliputi pengumpulan, deskripsi, eksplorasi, dan validasi data untuk memastikan data relevan dan siap digunakan.

Operator *Read Excel* digunakan untuk membaca file *Excel* dan memilih atribut yang relevan dari dataset PIP. Selanjutnya, data diimpor dari file Data Siswa PIP SDN 2 Purwawinangun.xlsx. Operator *Read Excel* ditampilkan pada Gambar 4.1.



Gambar 4.1 Oprator *Read Excel*

Parameter yang di gunakan dalam oprator Read Excel bisa di lihat pada tabel 4.1

Tabel 4.1 Parameter *Read Excel*

Dari pembacan oprator Read Excel di dapat informasi sebagai berikut :

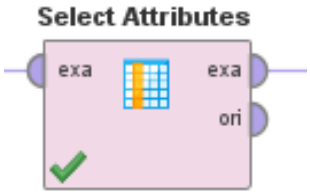
Tabel 4.1 Statistik Dataset

No.	Parameter	Isi
1	Read Excel	C:\STMIK IKMI CIREBON\SKRIPSI\SEKRIPSI\Data Siswa PIP SDN 2 Purwawinangun.xlsx
2	Sheet Selection	Sheet number
3	Sheet number	1
4	Imported cell range	A1
5	Encoding	System
6	Use header row	✓
7	Header row	1

No.	Uraian	Keterangan
1	Record	238
2	Regular Atribut	38

4.1.2 Selection (Pemilihan Data)

Tahap berikutnya yaitu memasukkan operator *Select Attributes*, yang berfungsi untuk memilih atau menentukan kolom (atribut) mana saja yang akan dipakai dalam analisis. Operator *Select Attributes* dapat dilihat pada Gambar 4.2.



Gambar 4.2 Oprator Select Attributes

Parameter yang di gunakan pada operator select atribut dapat di lihat pada tabel 4.2

Tabel 4.2 Parameter Operator *Select Attributes*

No.	Parameter	Isi
1	Type	Include attributes
2	Attribute filter type	a subset

Pada parameter *select attributes* terdapat *select subset* seperti dalalam tabel 4.3

Tabel 4.3 *Select Subset*

No.	*Select Attributes*	*Attributes*
1	No	Nama
2	Tanggal Lahir	NIPD
3	NISN	Tempat Lahir
4	Alamat	JK
5	RT	Agama
6	RW	Dusun
7	Kelurahan	Kode POS
8	Kecamatan	Jenis Tinggal
9	Alat Trasportasi	HP
10	HP	Nama Ayah
11	Penerima KPS	Tahun Lahir Ayah
12	Jenjang Pendidikan Ayah	Nama Ibu
13	Pekerjaan Ayah	Tahun Lahir Ibu
14	Penghasilan Ayah	Penerima KIP
15	NIK Ayah	
16	Jenjang Pendidikan Ibu	
17	Pekerjaan Ibu	
18	Penghasilan Ibu	
19	NIK Ibu	
20	Alasan Layak PIP	
21	Anak ke-berapa	
22	No. KK	
23	Jumlah Sodara Kandung	
24	Jarak Rumah Ke Sekolah	

4.1.3 Preprocessing

Tahapan preprocessing dilakukan untuk menangani data yang hilang atau tidak konsisten. Dalam penelitian ini, operator *Replace Missing Values* digunakan untuk mengisi nilai yang kosong atau hilang dengan pendekatan tertentu, seperti rata-rata atau nilai modus. Hal ini dilakukan agar model dapat memproses data secara optimal tanpa terganggu oleh data yang tidak lengkap. Operator *Replace Missing Values* seperti pada Gambar 4.3.

Gambar 4.3 Operator *Replace Missing Values*

Berikut ini parameter yang di gunakan pada oprerator mising values bisa dilihat pada tabel 4.4

Tabel 4.4 Parameter Operator *Replace Missing Values*

No.	Parameter	Isi
1	Attributs Filter Type	all
2	Default	Average

Operator *Replace Missing Values* menggunakan pengaturan standar *(default)*. Namun pada data program Indonesia pintar (PIP) tidak menunjukan adanya perbedaan terkait data yang kosong ataupun *missing*. Karena data tersebut sudah memiliki kualitas yang baik sehingga data sudah siap untuk dianalisis dan tidak memerlukan operator *Replace Missing Values* untuk Pemrosesan data. Seperti yang ditampilkan pada Tabel 4.4.

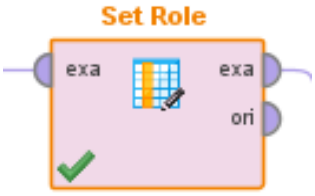
Tabel 4.5 Hasil dari Operator *Replace Missing Values*

No	Atribut	*Type* Atribut	

			Missing Values
1	No	Integer	0
2	Tanggal Lahir	Polynominal	0
3	NISN	Real	0
4	Alamat	Polynominal	0
5	RT	Integer	0
6	RW	Integer	0
7	Kelurahan	Polynominal	0
8	Kecamatan	Polynominal	0
9	Alat Trasportasi	Polynominal	0
10	Jenjang Pendidikan Ayah	Polynominal	0
11	Pekerjaan Ayah	Polynominal	0
12	Penghasilan Ayah	Polynominal	0
13	Jenjang Pendidikan Ibu	Polynominal	0
14	Pekerjaan Ibu	Polynominal	0
15	Penghasilan Ibu	Polynominal	0
16	Penerima KIP	Polynominal	0
17	Layak PIP	Polynominal	0
18	Alasan Layak PIP	Polynominal	0
19	Anak ke-berapa	Integer	0
20	Jumlah Sodara Kandung	Integer	0
21	Jarak Rumah Ke Sekolah	Integer	0
Example sheet		283	
Regular Attributs		21	
Special Attributs		0	

4.1.4 Transformation (Traspormasi Data)

Pada tahapan ini,untuk mengubah data menjadi model analisis data dan memodelkan data agar sesuai dengan analisis data mining yang diharapkan, tujuan transformasi adalah mengubah data yang dipilih ke dalam bentuk prosedur penambangan. Mengubah kode No menjadi ID dan kode Layak PIP menjadi Label. Atribut dalam label berperan sebagai label dan atribut berlabel sebagai operator pembelajaran, label ini juga disebut sebagai variabel atau kelas. Berikut data yang diolah untuk data *mining.*



Gambar 4.4 Oprator Set-Role

Set-Role merupakan operator yang mengklasifikasikan atribut menjadi atribut spesifik atau atribut standar . Operator *Set Role* memisahkan baris yang mengambil atribut koordinat serta prediksi posisi yang diberikan ke kelas label. Dibawah ini merupakan hasil transformation data dengan mengganti No sebagai ID serta Layak PIP sebagai label. Hasil transformasi data dapat di lihat pada tabel 4.6 hasil transformasi dan gambardata berikut ini. 9 (tambahkan Atribur sebelum dan setelah antara atribut special dan atribut biasa)

Tabel 4.6 Parameter Operator Set-Role

No.	Parameter	Isi
1.	Attribut name :	Target role :
	No	Id
	Layak PIP	Label

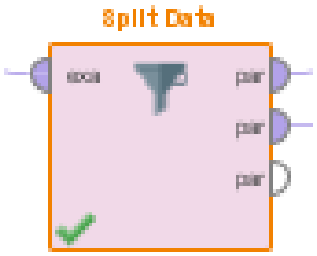
Hasil transformasi data dengan menggunakan operator Set-Role dapat di lihat pada tabel 4.7

Tabel 4.7 Hasil Transformasi Data

Example sheet	283
Regular Attributs	19
Special Attributs	2

4.1.5 Data Mining

Sebelum tahap Klasifikasi dilakukan, maka operator Operator *Split data* diperlukan untuk membagi data training dan testing . Operator *Split Data* dapat dilihat pada Gambar 4.5.

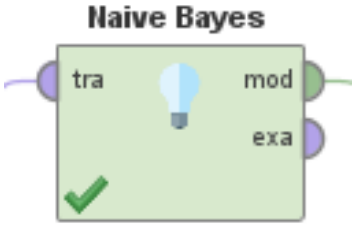


Gambar 4.5 Operator Split Data
Parameter yang di gunakan dalam Operator Split Data dapat di lihat pada tabel 4.10.

Tabel 4.8 Parameter Split Data

No.	Parameter	Isi
1	Sampling type	Stratifiled
2	local rendem seed	1

Selanjutnya penggunaan Operator *Naive Bayes* berfungsi untuk melatih model klasifikasi berbasis probabilitas yang cepat dan efisien, terutama untuk data kategorikal, Operator Naive Bayes ditampilkan pada Gambar 4.6.



Gambar 4.6 Operator *Naïve Bayes*

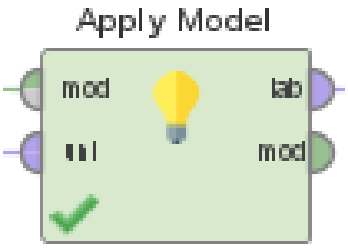
Parameter yang di gunakan dapat di lihat pada Tabel 4.9 berikut.

Tabel 4.9 Parameter *Naïve Bayes*

No.	Parameter	Isi
1	Laplace correction	✓

Hasil dari penggunaan Operator *Naive Bayes* bisa di lihat sebagai berikut.
Simpel Distribution.
Class Ya (0,641)
19 distributions
Class Tidak (0,359)
19 distributions

Setelah Operator *Naive Bayes* diterapkan, maka tahap berikutnya memasukkan Operator *Apply Model* diterapkan untuk memprediksi hasil pada data uji. Operator *Apply Data* dapat dilihat pada Gambar 4.8.



Gambar 4.8 Operator *Apply Data*

Parameter yang di gunakan dalam oprator *Naive Bayes* mengguanakan *default*

Selanjutnya menggunakan Operator *Performance* digunakan untuk mengukur kualitas prediksi model melalui metrik evaluasi seperti akurasi, presisi, dan recall. Operator Performance dapat di lihat pada gambar 4.8.

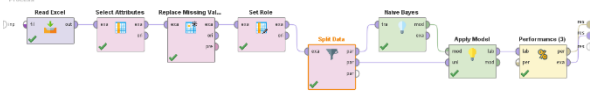
Gambar 4.9 Operator *Performance*

Parameter yang di gunakan dapat di lihat pada Tabel 4.10 berikut ini.

Tabel 4.10 Parameter *Performance*

No.	Parameter	Isi
1	*Performance*	
2	*Main Criterion*	*Accuracy*
3	*Accuracy*	✓

Gambar proses data mining dapat di lihat pada gambar 4.10



Gambar 4.10 Proses Data Mining
Hasil dari Oprator *Performance* dapat di lihat sebagai berikut.
Tabel 4.14 Hasil dengan Rasio 60 : 40

Accuracy : 92,92 %

	***True* Ya**	***True* Tidak**	***Class Precision***
Pred. Ya	67	3	95,71%
Pred. Tidak	5	38	88,37%

Class Recall	93,06%	92,68%	

Tabel 4.15 Hasil dengan Rasio 70 : 30

Accuracy : 96,47 %

	***True* Ya**	***True* Tidak**	***Class Precision***
Pred. Ya	51	0	100,00%
Pred. Tidak	3	31	91,18%
Class Recall	94,44%	100,00%	

Tabel 4.14 Hasil dengan Rasio 80 : 20

Accuracy : 96,43 %

	***True* Ya**	***True* Tidak**	***Class Precision***
Pred. Ya	34	0	100,00%
Pred. Tidak	2	20	90,91%
Class Recall	94,44%	100,00%	

Tabel 4.14 Hasil dengan Rasio 90 : 10

Accuracy : 96,43 %

	***True* Ya**	***True* Tidak**	***Class Precision***
Pred. Ya	17	0	100,00%
Pred. Tidak	1	10	90,91%
Class Recall	94,44%	100,00%	

Akurasi adalah metrik yang paling umum digunakan untuk mengevaluasi kinerja model. Akurasi mengukur persentase prediksi yang benar (baik kelas positif maupun negatif) terhadap seluruh data testing. Rumus untuk menghitung akurasi adalah:

1. *Accuracy*

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\%$$
$$= \frac{51+31}{51+31+0+3} \text{ x } 100\%$$
$$= 0{,}9647 = 96{,}47\%$$

4.2 Pembahasan

Penelitian ini membahas penerapan algoritma Naïve Bayes untuk mengklasifikasikan kelayakan peserta didik calon penerima Program Indonesia Pintar (PIP).

4.2.1 Pengaruh Penggunaan Berbagai Rasio Pembagian Data

Algoritma Naïve Bayes diuji dengan empat skenario pembagian data:

1. 60:40 Akurasi: 92,92%
2. 70:30 Akurasi: 96,47%
3. 80:20 *(Terbaik)* Akurasi: 96,43%
4. 90:10 Akurasi: 96,43%

Hasil menunjukkan bahwa rasio 70:30 memberikan keseimbangan terbaik antara akurasi, precision, dan recall.

Perbandingan dengan Penelitian Sebelumnya:

1. [12]: Naïve Bayes efektif untuk klasifikasi status gizi balita dengan akurasi optimal pada rasio 90:10, 80:20, dan 70:30.
2. [13]: Rasio pembagian data lebih dari 70% untuk pelatihan meningkatkan kinerja model pre-trained dalam klasifikasi gambar.

4.2.2 Akurasi Terbaik dalam Klasifikasi

Model Naïve Bayes menunjukkan performa tertinggi pada rasio 70:30 dengan akurasi 96,47%, precision 100,00%, dan recall 94,44%.

5. KESIMPULAN

Penelitian ini membahas penerapan algoritma Naive Bayes untuk mengklasifikasikan kelayakan peserta didik calon penerima Program Indonesia Pintar (PIP).

1. Pengaruh Rasio Pembagian Data

Penggunaan rasio pembagian data 60:40, 70:30, 80:20, dan 90:10 memengaruhi akurasi model. Semakin besar proporsi data training, semakin baik performa model:

1. 60:40 – Akurasi 92,92%, Precision 95,71%, Recall 93,06%.
2. 70:30 – Akurasi 96,47%, Precision 100,00%, Recall 94,44%.
3. 80:20 – Akurasi 96,43%, Precision 100,00%, Recall 94,44%.
4. 90:10 – Akurasi 96,43%, Precision 100%, Recall 94,44%.

Rasio 70:30 direkomendasikan karena memberikan keseimbangan terbaik antara akurasi, precision, dan recall.

2. Nilai Akurasi Terbaik

Model Naive Bayes mencapai akurasi terbaik sebesar 96,47% pada rasio 70:30. Precision 100,00% dan recall 94,44% menunjukkan bahwa model ini akurat dan andal untuk seleksi penerima PIP dengan tingkat kesalahan rendah.

Kesimpulan Utama:

1. Algoritma Naive Bayes efektif dan konsisten dengan penelitian sebelumnya [12] dan [13]
2. Model ini mampu mengurangi bias subjektif dan mempercepat proses pengambilan keputusan secara otomatis dan akurat di SDN 2 Purwawinangun.

UCAPAN TERIMA KASIH

uji syukur kami panjatkan kepada Tuhan Yang Maha Esa atas kelancaran penelitian ini. Kami mengucapkan terima kasih kepada:

1. Pihak SDN 2 Purwawinangun atas izin dan dukungan dalam pengumpulan data.
2. Kemendikbud atas data sekunder yang relevan.
3. Dosen Pembimbing atas bimbingan dan arahannya.
4. Keluarga, sahabat, dan rekan atas dukungan dan motivasinya.

Semoga penelitian ini bermanfaat bagi dunia pendidikan dan penelitian.

Hormat kami,

DAFTAR PUSTAKA

[1] D. Pradana and E. Sugiharti, "Implementation data mining with naive bayes classifier method and laplace smoothing to predict students learning results," *Recursive J. Informatics*, vol. 1, no. 1, pp. 1–8, 2023, doi: 10.15294/rji.v1i1.63964.

[2] D. Utami and P. Devi, "Klasifikasi kelayakan penerima bantuan program keluarga harapan (pkh) menggunakan metode weighted naïve bayes dengan laplace smoothing," *Jipi (Jurnal Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1373–1384, 2022, doi: 10.29100/jipi.v7i4.3592.

[3] Y. Religia, G. Pranoto, and I. Suwancita, "Analysis of the use of particle swarm optimization on naïve bayes for classification of credit bank applications," *Jisa(jurnal Inform. Dan Sains)*, vol. 4, no. 2, pp. 133–137, 2021, doi: 10.31326/jisa.v4i2.946.

[4] P. Luthfy, "Perencanaan engineering design process pada pembelajaran outdoor di paud," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 7, no. 6, pp. 7397–7408, 2023, doi: 10.31004/obsesi.v7i6.5561.

[5] R. Pahlevi, "Penerapan metode naive bayes untuk menentukan klasifikasi kelayakan penerimaan bantuan rehabilitasi dan pembangunan sekolah pada dinas pendidikan dan kebudayaan kabupaten banyuasin," *J. Teknol. Inform. Dan Komput.*, vol. 9, no. 2, pp. 1176–1188, 2023, doi: 10.37012/jtik.v9i2.1790.

[6] U. Pujianto and P. Y. Ristanti, "Perbandingan kinerja metode C4.5 dan Naive Bayes dalam klasifikasi artikel jurnal PGSD berdasarkan mata pelajaran," *Tekno*, vol. 29, no. 1, p. 50, 2019, doi: 10.17977/um034v29i1p50-67.

[7] V. Lestari, R. Arianto, B. Anindito, Y. Taramita, E. Amalia, and O. Triswidrananta, "Aplikasi pembelajaran rekonstruksi algoritma pseudocode dengan pendekatan element fill-in-blank problems di pemrograman java," *J. Tek. Ilmu Dan Apl.*, vol. 3, no. 2, pp. 153–161, 2022, doi: 10.33795/jtia.v3i1.108.

[8] W. Chandrawati, "Pengembangan e-learning berbasis media interaktif smart apps creator terhadap motivasi belajar siswa smp empat lima 2 kedungpring pada pelajaran informatika," *Jiip - J. Ilm. Ilmu Pendidik.*, vol. 6, no. 11, pp. 9146–9154, 2023, doi: 10.54371/jiip.v6i11.2941.

[9] S. Maarif, N. Muna, and M. Darmawan, "Pengembangan entrepreneurship di kalangan pelajar," *JMK (Jurnal Manaj. Dan Kewirausahaan)*, vol. 8, no. 1, p. 53, 2023, doi: 10.32503/jmk.v8i1.3170.

[10] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: a review," *IEEE Access*, vol. 8, 2020, doi: 10.1109/access.2020.2988510.

[11] N. Imanda, M. Teknologi, I. Universitas, and K. Lhokseumawe, "PENERAPAN ALGORIMA NAIVE BAYES PADA," vol. 12, no. 3, 2024.

[12] M. A. Sembiring *et al.*, "Penerapan Naïve Bayes Untuk Mengetahui Status Gizi Balita," *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 565–570, 2023, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[13] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, pp. 331–339, 2024, doi: 10.14569/IJACSA.2024.0150235.