

Klasifikasi Penerima Bantuan Program Indonesia Pintar (PIP) Pada Siswa SMK Menggunakan Algoritma KNN, NBC dan C4.5

Tandra Adiyatma Putra*, Inggih Permana, Zarnelly, Megawati

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11950311584@students.uin-suska.ac.id, ²inggihpermana@uin-suska.ac.id, ³zarnelly@uin-suska.ac.id, ⁴megawati@uin-suska.ac.id

Email Penulis Korespondensi: 11950311584@students.uin-suska.ac.id

Submitted: 04/12/2024; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Program Indonesia Pintar (PIP) adalah inisiatif pemerintah yang bertujuan untuk memberikan bantuan pendidikan kepada siswa dari keluarga kurang mampu. Penelitian ini dilakukan di SMKN 4 Pekanbaru untuk meningkatkan ketepatan penyaluran bantuan PIP menggunakan metode data mining. Tiga algoritma klasifikasi digunakan untuk mengidentifikasi siswa yang berhak menerima bantuan, yaitu K-Nearest Neighbor (KNN), Naive Bayes Classifier (NBC), dan C4.5. Data yang digunakan dalam penelitian ini meliputi atribut seperti pekerjaan orang tua, penghasilan, dan alat transportasi yang digunakan. Proses pengolahan data mencakup pembersihan, normalisasi, dan pembagian data untuk uji dan latih. Hasil penelitian menunjukkan bahwa algoritma KNN memiliki performa terbaik dengan akurasi sebesar 84.20%, presisi sebesar 89.83%, dan recall sebesar 99.18%. Algoritma C4.5 menonjol dalam hal kesederhanaan model, sementara NBC menunjukkan hasil yang kurang optimal dibandingkan KNN.

Kata Kunci: C4.5; Klasifikasi; KNN; NBC; Penerima PIP

Abstract—The Indonesia Smart Program (PIP) is a government initiative aimed at providing educational assistance to students from underprivileged families. This research was conducted at SMKN 4 Pekanbaru to enhance the accuracy of distributing PIP aid using data mining methods. Three classification algorithms were used to identify students eligible for assistance: K-Nearest Neighbor (KNN), Naive Bayes Classifier (NBC), and C4.5. The data used in this study included attributes such as parental occupation, income, and the type of transportation used. The data processing involved cleaning, normalization, and splitting into test and training sets. The results showed that the KNN algorithm performed best with an accuracy of 84.20%, precision of 89.83%, and recall of 99.18%. The C4.5 algorithm excelled in model simplicity, while NBC showed less optimal results compared to KNN.

Keywords: C4.5; Classification; KNN; NBC; PIP Recipient

1. PENDAHULUAN

Sekolah Menengah Kejuruan Negeri (SMKN) 4 Pekanbaru adalah salah satu institusi pendidikan formal yang menyelenggarakan pendidikan kejuruan di tingkat menengah, yang menjadi lanjutan dari jenjang SMP atau sederajat. SMKN 4 Pekanbaru didirikan pada tahun 1994 dengan nama Sekolah Menengah Industri Kerajinan (SMIK), berfokus pada pendidikan industri. Beralamat di Jalan Purwodadi, Tampan, Pekanbaru, sekolah ini menawarkan program kejuruan di bidang teknologi industri. Saat ini, SMKN 4 Pekanbaru memiliki 118 guru dan 1.586 siswa, dengan 7 program kompetensi keahlian [1].

Sekolah ini juga menawarkan sejumlah program bantuan pendidikan yang ditujukan kepada siswa dari kalangan kurang mampu atau siswa miskin. Beberapa program bantuan yang disediakan antara lain Beasiswa Peningkatan Prestasi Akademik (PPA), Beasiswa SPP, Beasiswa Bidikmisi, dan Program Indonesia Pintar (PIP)[2]. Program-program ini bertujuan untuk mendukung siswa dalam memenuhi kebutuhan pendidikan sehari-hari, terutama bagi mereka yang memiliki keterbatasan ekonomi. Dalam implementasinya, bantuan biasanya diberikan dalam bentuk uang tunai, yang disalurkan melalui Kemendikbud kepada sekolah untuk siswa yang membutuhkan [3].

Program Indonesia Pintar (PIP) merupakan kerjasama antara tiga kementerian, yaitu Kementerian Pendidikan dan Kebudayaan (Kemendikbud), Kementerian Sosial (Kemensos), dan Kementerian Agama (Kemenag) [4]. PIP disalurkan melalui Kartu Indonesia Pintar (KIP) sebagai bentuk bantuan tunai kepada siswa dari keluarga miskin, rentan miskin, pemilik Kartu Penerima Manfaat (KPM), penyandang disabilitas, serta korban bencana alam atau musibah. Program ini bertujuan memastikan anak-anak dari keluarga kurang mampu mendapatkan akses pendidikan hingga menyelesaikan pendidikan menengah, baik di jalur formal maupun non-formal [5].

Bantuan yang diterima siswa dalam program PIP bervariasi berdasarkan tingkat pendidikan. Siswa SD menerima bantuan sebesar Rp 450.000, siswa SMP menerima Rp 750.000, dan siswa SMA/SMK menerima Rp 1.000.000 per tahun [6]. Bantuan ini diharapkan mampu mendukung siswa agar dapat terus melanjutkan pendidikan mereka meskipun berasal dari keluarga yang kurang mampu. Syarat utama penerima PIP antara lain siswa yang terdampak bencana, siswa dari keluarga miskin atau rentan miskin, serta siswa yang putus sekolah namun ingin melanjutkan pendidikan [7].

Tujuan dari Program Indonesia Pintar adalah untuk meningkatkan aksesibilitas, kesetaraan, dan mutu pendidikan bagi siswa yang kurang mampu secara ekonomi. Dalam proses pelaksanaannya, penyaluran bantuan PIP melibatkan beberapa tahap, mulai dari identifikasi dan verifikasi data siswa yang berhak, pembuatan rekening bank, penyaluran dana bantuan, hingga pemantauan dan evaluasi [8]. Namun, dalam pelaksanaannya, terdapat beberapa masalah yang muncul, seperti identifikasi yang tidak tepat sasaran, ketidakmerataan akses, kurangnya informasi dan komunikasi terkait program, serta ketidaktepatan dalam penyaluran dana bantuan [9].

Salah satu masalah signifikan dalam penyaluran PIP adalah identifikasi siswa yang berhak mendapatkan bantuan, yang sering kali tidak tepat sasaran. Selain itu, akses penerima yang tidak merata, serta kurangnya informasi dan komunikasi terkait program ini, memperparah masalah penyaluran dana yang tidak tepat waktu [10]. Kurangnya pemantauan dan evaluasi juga menyebabkan banyak siswa yang seharusnya berhak mendapatkan bantuan tidak menerima dana tersebut. Oleh karena itu, diperlukan evaluasi dan peninjauan ulang agar bantuan benar-benar sampai kepada siswa yang membutuhkan [11].

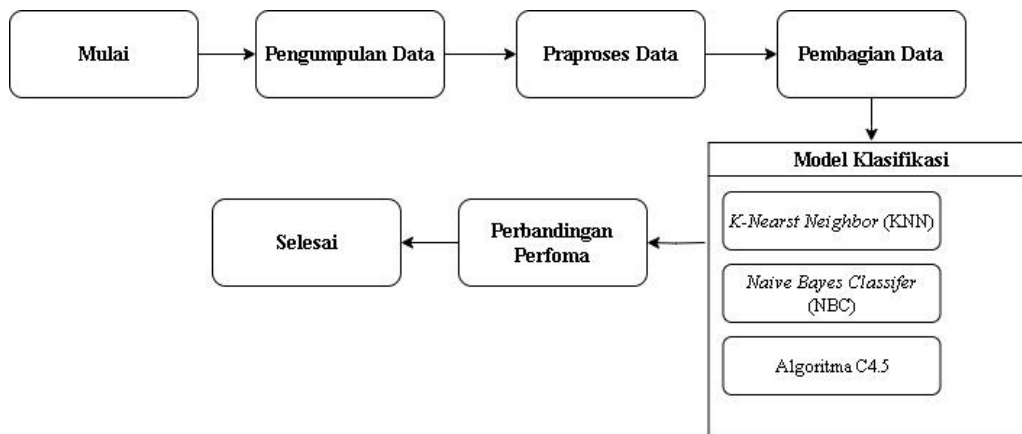
Dalam mengatasi permasalahan yang ada, maka penelitian ini menggunakan model klasifikasi yang tepat untuk menentukan apakah siswa layak atau tidak mendapatkan PIP. Klasifikasi digunakan untuk memastikan bahwa bantuan diterima dengan benar dan tepat sasaran. Penelitian ini menggunakan Algoritma Naïve Bayes, KNN, C4.5 [12]. Algoritma ini dipilih karena dapat mengklasifikasi penerima bantuan PIP, algoritma tersebut dapat digunakan dalam memprediksi kelayakan penerima bantuan PIP [13]. Algoritma ini juga dapat digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Performa dari klasifikasi ini diambil dan dibandingkan mana yang menjadi klasifikasi terbaik. Sehingga, model klasifikasi dapat meningkatkan efektivitas pengambilan keputusan siswa yang menerima PIP [14].

Adapun penelitian yang dilakukan oleh Purwanto dkk (2023) [15], dalam penelitian tersebut didapatkan hasil klasifikasi oleh algoritma K-NN dan C4.5, penelitian tersebut menunjukkan algoritma K-Nearest Neighbors memiliki performansi yang lebih baik yaitu presisi 98,08%. Penelitian lainnya oleh Sumiah dan Mirantika (2020) [16], penelitian tersebut menggunakan algoritma K-NN dan Naive Bayes dalam menentukan penerimaan beasiswa, hasil penelitian menunjukkan algoritma K-NN lebih baik dengan akurasi 100%. Penelitian oleh Noviyanto dan Mukti (2021) [17], penelitian ini menggunakan algoritma Naive Bayes dalam menentukan kelayakan penerima calon beasiswa, hasil yang didapatkan dengan nilai akurasi sebesar 90% sehingga algoritma ini dapat dijadikan sebagai pendukung keputusan untuk menentukan calon penerima beasiswa dengan baik.

Penelitian ini memiliki perbedaan dengan yang sebelumnya, pada penelitian ini melakukan klasifikasi terhadap penerima bantuan PIP dengan menggunakan 3 algoritma yaitu, K-NN, Naive Bayes dan C4.5. Penelitian ini diharapkan mampu memberikan kontribusi dalam mengoptimalkan penyaluran bantuan PIP agar lebih tepat sasaran, melalui penggunaan metode data mining dengan membandingkan performa tiga algoritma klasifikasi.

2. METODOLOGI PENELITIAN

Adapun Tahapan dalam melaksanakan penelitian ini, terdiri dari pengumpulan data, praproses data, pembagian data, model klasifikasi, dan tahap perbandingan performa. Alur penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Metodologi Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan data penerima bantuan PIP pada SMKN 4 Pekanbaru. Sumber data pada penelitian ini adalah data usulan dari sekolah yaitu, Sekolah mengusulkan siswa yang dianggap layak menerima bantuan berdasarkan kondisi ekonomi, prestasi, dan partisipasi dalam program sekolah.

2.2 Praproses Data

2.2.1 Pembersihan Data

Dalam proses ini, langkah yang diambil adalah menghapus baris data yang memiliki nilai NULL. Proses pembersihan data dilakukan menggunakan operator *Filter Example* pada aplikasi RapidMiner Studio versi 2024.0.

2.2.2 Normalisasi Data

Penelitian ini menerapkan normalisasi data dengan metode Min-Max Normalization, menggunakan rentang nilai minimum 0 dan maksimum 1. Proses normalisasi dilakukan melalui operator *Normalize* pada aplikasi RapidMiner Studio versi 2024.0.

2.3 Pembagian Data

Pada penelitian ini, proses pembagian data latih dan data uji dilakukan dengan menerapkan teknik K-Fold Cross Validation, di mana nilai K ditetapkan sebanyak 10. Pengujian menggunakan K-Fold Cross Validation ini diimplementasikan melalui operator model Cross Validation yang terdapat dalam aplikasi RapidMiner Studio versi 2024.0, sehingga metode ini memungkinkan evaluasi yang lebih komprehensif dan akurat terhadap model yang dikembangkan.

2.4 Pembuatan Model Klasifikasi

2.4.1 K-Nearest Neighbor (K-NN)

Pada penelitian ini, parameter percobaan K-NN disajikan dalam Tabel 1. Untuk analisis data, digunakan perangkat lunak RapidMiner Studio versi 2024.0 dengan memanfaatkan operator K-NN sebagai algoritma utamanya. Dalam penerapan model K-NN, digunakan jenis parameter *measure types* yang mengacu pada *mixed measure*, dengan metode pengukuran yang diterapkan berupa *mixed euclidean distance*. Pendekatan ini dipilih karena dianggap mampu menggabungkan berbagai tipe data secara optimal dalam proses penghitungan jarak antar titik data.

Tabel 1. Parameter k-NN

No	Parameter	Nilai
1	Perhitungan jarak terdekat	Euclidean distance
2	Nilai k	K=3, K=5, K=7, K=9, K=11, K=13, K=15, K=17, K=19, K=21

2.4.2 Naive Bayes Classifier (NBC)

Model klasifikasi Naïve Bayes dibuat menggunakan aplikasi RapidMiner Studio versi 2024.0, dengan memanfaatkan operator Naïve Bayes untuk membangun model. Pada proses pembentukan model ini, digunakan parameter laplace correction sebagai solusi untuk mengatasi kemungkinan munculnya nilai probabilitas nol, yang seringkali menjadi masalah dalam perhitungan probabilitas pada algoritma tersebut [18].

2.4.3 Algoritma C4.5

Proses pembuatan model klasifikasi C4.5 dilaksanakan dengan memanfaatkan operator Decision Tree pada aplikasi RapidMiner Studio versi 2024.0. Dalam penerapannya, digunakan sejumlah parameter yang tercantum dalam Tabel 2 sebagai acuan. Setelah model selesai dibangun, dilakukan tahap penerapan model (*apply model*) serta evaluasi kinerja (*performance*) untuk memperoleh hasil evaluasi model dan visualisasi struktur pohon keputusan yang terbentuk.

Tabel 2. Parameter C4.5

No	Parameter	Nilai
1	Criterion	Information Gain
2	Maximal Depth	5
3	Apply Pruning	Confidence= 0,5

2.5 Perbandingan Perfoma

Dalam penelitian ini, pengukuran kinerja dilakukan dengan menggunakan metode Confusion Matrix. Dari Confusion Matrix ini dapat menghitung beberapa indikator kinerja seperti nilai akurasi, presisi, dan recall. Proses penghitungan dilakukan melalui penggunaan operator model Performance yang tersedia di dalam perangkat lunak RapidMiner Studio versi 10.1. Setelah hasil pengukuran kinerja diperoleh, dilakukan analisis perbandingan terhadap tiga algoritma yaitu k-Nearest Neighbors (k-NN), Naive Bayes Classifier (NBC), dan C4.5, untuk menilai model mana yang memiliki akurasi terbaik. Rumus yang digunakan untuk menghitung akurasi dapat dilihat pada persamaan 1, sementara rumus recall tercantum pada persamaan 2, dan rumus presisi terdapat pada persamaan 3. Penghitungan ini bertujuan untuk mendapatkan gambaran lebih jelas tentang performa masing-masing algoritma. [19], [20].

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN} \right) \times 100 \tag{1}$$

$$Recall = \left(\frac{TP}{TP+FN} \right) \times 100 \tag{2}$$

$$Precision = \left(\frac{TP}{TP+FP} \right) \times 100 \tag{3}$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Dari hasil pengumpulan data usulan sekolah yang menerima bantuan PIP yang berupa format microsoft excel. Pada data tersebut terdapat 1520 baris calon penerima bantuan PIP pada SMKN 4 Pekanbaru. Data tersebut terdiri dari 11 atribut sebagai kriteria penerima bantuan PIP. Pada data tersebut terdiri dari 408 siswa yang berhak menerima bantuan PIP dan 1112 siswa yang tidak berhak menerima bantuan PIP. Untuk lebih detailnya atribut-atribut data dapat dilihat Tabel 3. Sampel data Penerima PIP dapat dilihat pada Tabel 4.

Tabel 3. Atribut-Atribut Penerima PIP

Kode	Atribut	Keterangan	Nilai
A1	Alat Transportasi	Alat transportasi yang digunakan untuk ke sekolah.	1.Andong/bendi/sado/dokar/delman/becak; 2. Angkutan umum/bus/pete-pete; 3. Jalan kaki; 4.Kendaraan pribadi; 5. Mobil pribadi; 6. Mobil/bus antar jemput; 7. Ojek; 8. Sepeda; 9. Sepeda motor; 10. Lainnya
A2	Penerima KPS	Siswa penerima Kartu Perlindungan Sosial (KPS)	1. Ya; 2. Tidak
A3	Pekerjaan Ayah	Jenis pekerjaan yang sedang dilakukan oleh orang tua laki-laki	1. Buruh; 2. Karyawan BUMN; 3. Karyawan Swasta; 4. Pedagang Besar; 5. Pedagang Kecil; 6.Pensiunan; 7. Petani; 8. Peternak; 9. PNS/TNI/Polri; 10. Wiraswasta; 11. Wirausaha; 12. Lainnya; 13. Sudah Meninggal; 14. Tidak bekerja; 15. Tidak dapat diterapkan
A4	Penghasilan Ayah	Kondisi keuangan Ayah	1. Kurang dari Rp. 500,000; 2. Rp. 500,000 - Rp. 999,999; 3. Rp. 1,000,000 - Rp. 1,999,999; 4. Rp. 2,000,000 - Rp. 4,999,999; 5. Rp. 5,000,000 - Rp. 20,000,000; 6. Lebih dari Rp. 20,000,000; 7. Tidak Berpenghasilan
A5	Pekerjaan Ibu	Jenis pekerjaan yang sedang dilakukan oleh orang tua perempuan.	1. Buruh; 2. Karyawan BUMN; 3. Karyawan Swasta; 4. Pedagang Besar; 5. Pedagang Kecil; 6.Pensiunan; 7. Petani; 8. Peternak; 9. PNS/TNI/Polri; 10. Wiraswasta; 11. Wirausaha; 12. Lainnya; 13. Sudah Meninggal; 14. Tidak bekerja; 15. Tidak dapat diterapkan
A6	Penghasilan Ibu	Kondisi keuangan Ibu	1. Kurang dari Rp. 500,000; 2. Rp. 500,000 - Rp. 999,999; 3. Rp. 1,000,000 - Rp. 1,999,999; 4. Rp. 2,000,000 - Rp. 4,999,999; 5. Rp. 5,000,000 - Rp. 20,000,000; 6. Lebih dari Rp. 20,000,000; 7. Tidak Berpenghasilan
A7	Penerima PIP	Penerima usulan dari sekolah untuk Program Indonesia Pintar (PIP)	1. Ya; 2. Tidak
A8	Kebutuhan Khusus	Penerima PIP merupakan penyandang disabilitas	1. Ya; 2. Tidak
A9	Anak Ke-Berapa	Anak ke berapa dalam keluarga	Menyesuaikan masing-masing anak didalam keluarga setiap rumah.
A10	Jumlah Saudara Kandung	Berapa jumlah saudara penerima PIP	Menyesuaikan masing-masing anggota keluarga pada setiap rumah.
A11	Jarak Rumah ke Sekolah	Sejauh mana antara rumah ke sekolah	Menyesuaikan jarak masing-masing siswa dari rumah ke sekolah
A12	Status (Label)	Atribut label yang bernilai kelas untuk calon penerima PIP. Apabila bernilai Yes maka kriteria penerima dan apabila bernilai No maka kriteria tidak menerima	Ya Tidak

Tabel 4. Sampel Data Penerima PIP

Alat Transportasi	Penerima KPS	...	Anak ke-berapa	Jml. Saudara Kandung	Jarak Rumah ke Sekolah	Status
Jalan kaki	Tidak	...	1	0	0	Ya
Sepeda motor	Tidak	...	1	2	4	Tidak
Sepeda motor	Ya	...	3	3	2	Ya
Sepeda motor	Tidak	...	1	2	2	Tidak
...
Sepeda motor	Tidak	...	1	3	6	Tidak
Jalan kaki	Tidak	...	4	4	1	Tidak
Angkutan umum/bus/pete-pete	Tidak	...	1	1	3	Tidak
Sepeda motor	Tidak	...	2	4	5	Tidak

3.2 Hasil Praproses Data

Pada data awal terdapat 520 baris data yang mengandung nilai Blank (kosong), baris data tersebut kemudian dihapus sehingga tersisa 1000 data. Dari sisa data tersebut terdapat 269 baris data dengan label Yes (berhak menerima PIP) dan 731 baris data dengan label No (tidak berhak menerima PIP). Sampel data pembersihan dapat dilihat pada Tabel 5.

Tabel 5. Sampel Hasil Pembersihan Data

A1	A2	A3	A4	...	A9	A10	A11	Status
3	2	1	3	...	1	0	0	Ya
9	2	12	3	...	1	2	4	Tidak
9	1	10	3	...	3	3	2	Ya
9	2	10	4	...	1	2	2	Tidak
...
9	2	3	4	...	1	3	6	Tidak
3	2	1	3	...	4	4	1	Tidak
2	2	3	2	...	1	1	3	Tidak
9	2	6	4	...	2	4	5	Tidak
6	2	10	3	...	2	2	0	Tidak

Setelah melakukan pembersihan data dilakukan normalisasi *min-max*. Sampel data normalisasi dapat dilihat pada Tabel 6.

Tabel 6. Sampel Hasil Normalisasi

A1	A2	A3	A4	...	A9	A10	A11	Status
0.222	1	0	0.333	...	0	0	0	Ya
0.889	1	0.786	0.333	...	0	0.222	0.200	Tidak
0.889	0	0.643	0.333	...	0.250	0.333	0.100	Ya
0.889	1	0.643	0.500	...	0	0.222	0.100	Tidak
...
0.889	1	0.143	0.500	...	0	0.333	0.300	Tidak
0.222	1	0	0.333	...	0.375	0.444	0.050	Tidak
0.111	1	0.143	0.167	...	0	0.111	0.150	Tidak
0.889	1	0.357	0.500	...	0.125	0.444	0.250	Tidak

3.3 Hasil Pemodelan Klasifikasi

3.3.1 K-Nearest Neighbor

Berikut adalah hasil pengujian menggunakan k=10 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Pengujian Nilai k

Perfoma k-NN				
No	K	Accuracy	Precision	Recall
1	3	83,20%	88,55%	97,95%
2	5	77,80%	79,42%	93,98%
3	7	83,20%	88,55%	97,95%



4	9	78,30%	78,37%	97,13%
5	11	78,60%	81,61%	97,81%
6	13	78,20%	84,93%	98,50%
7	15	77,70%	83,82%	98,00%
8	17	84,20%	80,50%	82,39%
9	19	78,00%	87,69%	98,91%
10	21	77,80%	89,83%	99,18%

Berdasarkan Tabel 7 dapat dilihat bahwa akurasi tertinggi terjadi ketika nilai K pada K-NN sama dengan 21, yaitu dengan nilai akurasi 99,18%. Presisi tertinggi terjadi ketika K pada K-NN sama dengan 21, yaitu dengan nilai presisi 89,83%, sedangkan recall tertinggi ketika nilai K pada K-NN sama dengan 17, yaitu dengan nilai recall 84,20%. Karena akurasi, presisi dan recall tertinggi berada pada nilai K yang berbeda, maka penelitian ini cenderung memilih K-NN dengan nilai K pada recall terbaik yaitu 21. Hal ini karena pada model klasifikasi K=21 bisa meminimalkan kesalahan calon yang berhak menerima menjadi tidak berhak menerima. Hal tersebut dapat terlihat dari nilai recall yang tinggi. Hasil confusion matrix dari K=21 dapat dilihat pada Tabel 8.

Tabel 8. Confusion Matrix k-NN

Confusion Matrix		
	true No	true Yes
pred. No	34	8
Pres. Yes	88	275

3.3.2 Naïve Bayes Classifier

Hasil performa dari pengujian model NBC dapat dilihat pada Tabel 9 dan *confusion matrix* dapat dilihat pada Tabel 10.

Tabel 9. Hasil Performa NBC

Perfoma NBC			
	Accuracy	Precision	Recall
NBC	82,00%	83,48%	93,98%

Tabel 10. Confusion Matrix NBC

Confusion Matrix		
	true No	true Yes
pred. No	133	44
Pres. Yes	136	687

Berdasarkan pada Tabel 9 dapat dilihat algoritma *naïve bayes* akurasi sebesar 82%, presisi sebesar 83,48% dan recall sebesar 93,98%.

3.3.3 Algoritma C4.5

Hasil performa dari model C4.5 dapat dilihat pada Tabel 11 dan *confusion matrix* dapat dilihat pada Tabel 12. Pohon keputusan dari algoritma C4.5 dapat dilihat pada Gambar 2.

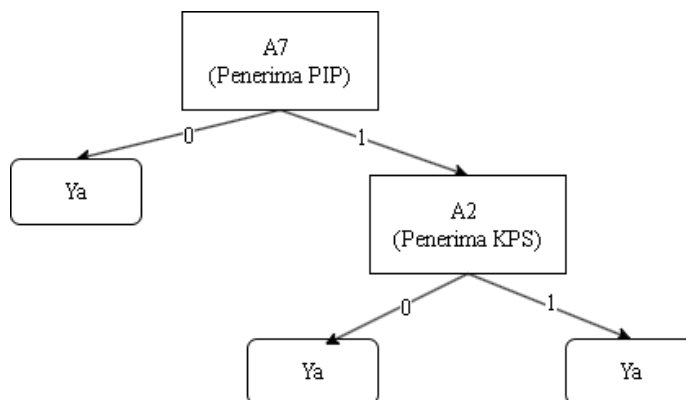
Tabel 11. Hasil Performa C4.5

Perfoma C4.5			
	Accuracy	Precision	Recall
C4.5	83,20%	88,55%	97,95%

Tabel 12. Confusion Matrix C4.5

Confusion Matrix		
	true No	true Yes
pred. No	116	15
Pres. Yes	153	716

Berdasarkan Tabel 11 dapat dilihat bahwa algoritma C4.5 memiliki akurasi sebesar 83,20%, nilai presisi sebesar 88,55% dan nilai recall sebesar 97,95%.

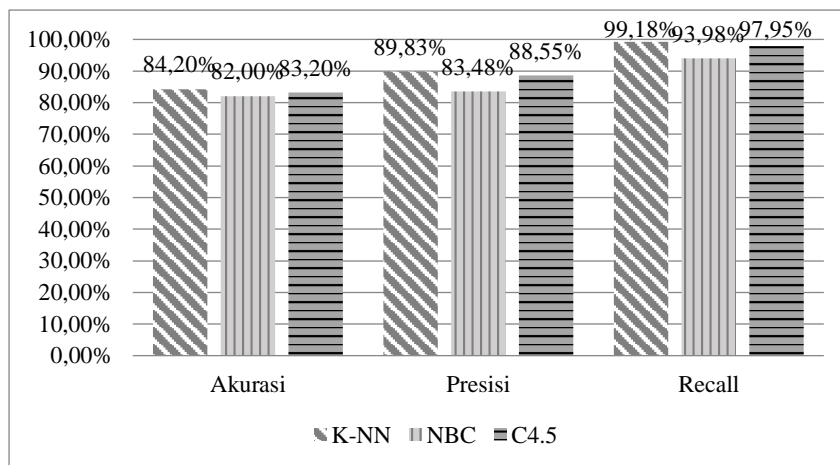


Gambar 2. Pohon Keputusan

Berdasarkan Gambar 2, terlihat bahwa meskipun awalnya terdapat 11 atribut yang dianalisis, setelah penerapan algoritma C4.5, hanya 2 atribut yang terbukti signifikan, yaitu "Penerima PIP" dan "Penerima KPS". Pengurangan jumlah atribut ini secara otomatis mengurangi kompleksitas model yang dihasilkan oleh algoritma C4.5, sehingga menghasilkan model yang lebih sederhana dan efisien dalam proses analisis data. Dengan jumlah atribut yang lebih sedikit, model menjadi lebih mudah dipahami dan diinterpretasikan.

3.4 Hasil Perbandingan Perfoma Model Klasifikasi

Hasil perbandingan algoritma klasifikasi dapat dilihat pada Gambar 3.



Gambar 3. Perfoma Hasil Pengujian Algoritma K-NN, NBC dan C4.5

Pada gambar tersebut diketahui akurasi, presisi, dan recall tertinggi pada model klasifikasi yang dihasilkan oleh algoritma k-NN. Hal ini menandakan bahwa KNN lebih efektif dalam mengidentifikasi siswa yang berhak menerima Program Indonesia Pintar (PIP). Tingginya recall pada KNN berarti algoritma ini dapat dengan baik menangkap banyak siswa yang seharusnya mendapatkan bantuan, sehingga meminimalkan risiko ketidakberdayaan siswa yang sebenarnya memenuhi syarat namun terlewatkan. Dengan demikian, keputusan yang dihasilkan lebih akurat dan dapat diandalkan, karena KNN mampu memprediksi dengan lebih baik siswa mana yang perlu dibantu. Dalam hal ini, penggunaan KNN sebagai algoritma pemodelan dapat mengurangi kesalahan dalam proses penerimaan PIP, memastikan bahwa siswa yang berhak mendapatkan dukungan finansial tidak terabaikan, yang pada gilirannya berkontribusi pada keberhasilan program tersebut dalam mendukung pendidikan di Indonesia.

Dalam penerimaan PIP, akurasi yang tinggi berarti bahwa jumlah siswa yang benar-benar layak dan tidak layak yang teridentifikasi oleh model adalah tepat, sehingga mengurangi kesalahan dalam pengambilan keputusan. Presisi mengukur seberapa banyak dari siswa yang diidentifikasi sebagai penerima PIP yang benar-benar layak menerima bantuan. Jika KNN memiliki presisi tinggi, ini berarti bahwa ketika model memprediksi siswa sebagai penerima PIP, kemungkinan besar prediksi tersebut benar. Hal ini mengurangi risiko penerima yang tidak tepat, sehingga lebih banyak siswa dapat dialokasikan secara efisien kepada siswa yang benar-benar membutuhkan.

Meskipun begitu algoritma C4.5 memiliki keunggulan dari segi kesederhanaan kompleksitas perhitungan. Hal ini dikarenakan model klasifikasi yang dihasilkan oleh algoritma C4.5 berhasil mereduksi jumlah atribut dari 11 menjadi 2 atribut.

4. KESIMPULAN

Indonesia Pintar (PIP) menggunakan tiga algoritma, yaitu K-Nearest Neighbor (KNN), Naive Bayes Classifier (NBC), dan C4.5, telah berhasil dibandingkan. Hasil evaluasi menunjukkan bahwa algoritma KNN memiliki performa terbaik dengan akurasi tertinggi sebesar 84.20%, presisi sebesar 89.83%, dan recall sebesar 99.18%. Algoritma ini dinilai lebih efektif dalam mengidentifikasi siswa yang layak menerima bantuan PIP karena mampu meminimalkan kesalahan klasifikasi. Meskipun algoritma C4.5 memiliki keunggulan dalam hal kesederhanaan model dengan hanya menggunakan dua atribut signifikan, yaitu "Penerima PIP" dan "Penerima KPS", performanya masih berada di bawah KNN. Sementara itu, NBC juga menunjukkan kinerja yang baik dengan akurasi sebesar 82%, namun masih kurang dibandingkan KNN. Penggunaan metode data mining ini diharapkan dapat membantu pihak sekolah dalam membuat keputusan yang lebih akurat dalam penyaluran bantuan, sehingga program PIP dapat berjalan lebih efektif dan tepat sasaran, mendukung pendidikan bagi siswa yang membutuhkan.

REFERENCES

- [1] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 452–458, 2023.
- [2] T. Siregar, R. Ardian, and A. Arisman, "STUDI KASUS SMA N 1 SINUNUKAN: IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PENERIMA BEASISWA PROGRAM INDONESIA PINTAR (PIP)," *Jurnal Cermatika*, vol. 4, no. 1, pp. 9–25, 2024.
- [3] I. Priyanto, E. M. Dewanti, T. Tundo, M. Nurdin, and R. Kasiono, "PENERAPAN ALGORITMA METODE NAÏVE BAYES UNTUK PENENTUAN PENERIMAAN BANTUAN PROGRAM INDONESIA PINTAR (PIP)," *Jurnal Manajemen Informatika Jayakarta*, vol. 4, no. 2, pp. 162–172, 2024.
- [4] C. Yanasari and T. Arifin, "Implementasi Algoritma K-Nearest Neighbor Untuk Klasifikasi Penerimaan Beasiswa Program Indonesia Pintar," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 1, no. 4, pp. 178–194, 2023.
- [5] I. W. Lorenza, R. Umilasari, and M. Dasuki, "Klasifikasi Penerima Bantuan Beasiswa Menggunakan Algoritma K-Nearest Neighbour Dengan Seleksi Fitur Backward Elimination," *Jurnal Aplikasi Sistem Informasi dan Elektronika*, vol. 3, no. 1, pp. 26–31, 2021.
- [6] D. J. Sitanggang, E. Buulolo, and S. Aripin, "Normalisasi Data Dengan Model Min Max Untuk Klasifikasi Calon Mahasiswa Yang Layak Memperoleh KIP Kuliah Dengan Algoritma K-Nearest Neighbor," *Informasi dan Teknologi Ilmiah (INTI)*, vol. 11, no. 3, pp. 88–95, 2024.
- [7] Z. Arifin, W. J. Shudiq, and S. Maghfiroh, "Penerapan Metode Knn (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerimaan Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql," *NJCA (Nusantara Journal of Computers and Its Applications)*, vol. 4, no. 1, pp. 27–34, 2019.
- [8] M. Hidayat, A. N. Fuadi, D. P. Utomo, E. D. Astuti, and D. Asmarajati, "Studi Komparasi Algoritma Naïve Bayes Dan K-Nn Untuk Klasifikasi Penerimaan Beasiswa Di Mi Al-Islamiah Karangsawah," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 4, pp. 172–180, 2023.
- [9] E. Budiarto, R. Rino, S. Hariyanto, and D. Susilawati, "Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4. 5," *ALGOR*, vol. 3, no. 2, pp. 23–34, 2022.
- [10] A. Amalia, A. I. Purnamasari, and I. Ali, "IMPLEMENTASI ALGORITMA C4. 5 DAN NAÏVE BAYES DALAM PENGAMBILAN KEPUTUSAN UNTUK PROGRAM INDONESIA PINTAR (PIP) DI SEKOLAH DASAR NEGERI 04 MAJALANGU," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1889–1896, 2024.
- [11] W. R. S. O. Ningse, S. Sumarno, and Z. M. Nasution, "Klasifikasi Algoritma C4. 5 untuk Penentuan Penerima Program Indonesia Pintar pada MIS Al-Khoirot C4. 5 Algorithm Classification for Determining Smart Indonesia Program Recipients at MIS Al-Khoirot," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 1, pp. 65–76, 2022, doi: 10.55123/jomlai.v1i1.165x.
- [12] F. D. Pratama, I. Zufria, and T. Triase, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 1, pp. 77–84, 2022.
- [13] H. Hamria, "K-Nearest Neighbor Berbasis Seleksi Atribut Chi Square Untuk Klasifikasi Penerima Beasiswa," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 1, pp. 39–48, 2023.
- [14] A. Hadi and I. Ali, "Menentukan Kelayakan Penerima Kip Menggunakan Klasifikasi Dengan Metode Algoritma Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 366–372, 2023.
- [15] A. Purwanto and H. W. Nugroho, "Analisa Perbandingan Kinerja Algoritma C4. 5 Dan Algoritma K-Nearest Neighbors Untuk Klasifikasi Penerima Beasiswa," *Jurnal Teknoinfo*, vol. 17, no. 1, pp. 236–243, 2023.
- [16] A. Sumiah and N. Mirantika, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan," *Buffer Informatika*, vol. 6, no. 1, pp. 1–14, 2020.
- [17] H. Noviyanto and B. Mukti, "Implementasi Algoritme Naive Bayes untuk Menentukan Kelayakan Calon Penerima Beasiswa," *Jurnal Informatika Dan Tekonologi Komputer (JITEK)*, vol. 1, no. 2, pp. 7–12, 2021.
- [18] Q. A'yuniyah et al., "Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 72, Sep. 2022, doi: 10.30865/json.v4i1.4781.
- [19] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-Sakti)*, vol. 5, no. 2, pp. 697–711, 2021.
- [20] K. P. Siwilopo and H. Marcos, "MEMBANDINGKAN KLASIFIKASI PADA BUAH JERUK MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DAN K-NEAREST NEIGHBOR," *KOMPUTA: Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, 2023.