

IMPLEMENTASI ALGORITMA C4.5 DAN NAÏVE BAYES DALAM PENGAMBILAN KEPUTUSAN UNTUK PROGRAM INDONESIA PINTAR (PIP) DI SEKOLAH DASAR NEGERI 04 MAJALANGU

Ana amalia, Ade Irma Ade Irma Purnamasari, Irfan Ali

Teknik Informatika, STMIK IKMI Cirebon

Rekaya perangkat Lunak, STMIK IKMI Cirebon

Jalan Raya perjuangan No 10 B kota Cirebon, Indonesia

Amaliaa835@gmail.com

ABSTRAK

Pendidikan di Indonesia menghadapi tantangan terkait dengan kemiskinan, di mana prestasi siswa dinilai setiap akhir semester. Program Indonesia Pintar memberikan dukungan keuangan untuk biaya sekolah dan bantuan belajar siswa. Penelitian ini bertujuan membantu SD N 04 Majalangu dalam menyeleksi penerima Program Indonesia Pintar yang lebih akurat dan efisien. Metode Naïve Bayes adalah metode klasifikasi dari machine learning yang memiliki keunggulan yaitu menggunakan sampel data pelatihan untuk mengestimasi parameter yang terlibat dalam proses klasifikasi dapat tersaji secara cepat, serta memperoleh akurasi tinggi [1]. Algoritma C4.5 adalah algoritma yang digunakan untuk membentuk pohon keputusan, yang merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal [2]. Metode Decision Tree dengan algoritma C4.5 dan Naïve Bayes digunakan untuk pemilihan yang lebih objektif. Penelitian ini memastikan bantuan tepat sasaran, mengurangi beban biaya pendidikan, dan meningkatkan motivasi belajar siswa. Data mining memastikan keputusan seleksi didasarkan pada bukti yang kuat, mengurangi potensi kesalahan. Hasil pengujian menunjukkan akurasi model algoritma C4.5 sebesar 64,42%, dan algoritma Naïve Bayes sebesar 100%. Penelitian ini memberikan wawasan tentang faktor-faktor kontribusi dalam menentukan kelayakan penerima PIP di SD N 04 Majalangu, mendukung efisiensi dan objektivitas dalam pemilihan penerima bantuan pendidikan.

Kata kunci : Klasifikasi, Siswa , Pip, Algoritma C.45 , Naïve Bayes

1. PENDAHULUAN

Kemajuan dari suatu negara tergantung dari kemajuan aspek Pendidikan guna menciptakan pembangunan nasional yang lebih baik. Pendidikan mampu mencerdaskan anak bangsa sehingga akan tercapai sebuah kesejahteraan rakyat (Agnesita Widi Larasati, 2019).[3] Penelitian ini membahas kemajuan pendidikan di Indonesia dan kesadaran meningkat terhadap pentingnya pendidikan. Tantangan muncul bagi masyarakat ekonomi rendah yang ingin melanjutkan pendidikan, namun mengalami kendala finansial. Meskipun minat pendaftaran siswa baru meningkat, tantangan terkait seleksi Program Indonesia Pintar (PIP) tetap ada.

Proses seleksi saat ini terbatas pada verifikasi manual tanpa analisis data yang efektif, mengakibatkan hasil yang suboptimal. Oleh karena itu, penelitian ini bertujuan menerapkan teknik data mining dengan algoritma C4.5 dan Naïve Bayes untuk mengidentifikasi dan mengelompokkan data secara akurat dalam seleksi penerima PIP. Tujuannya adalah meningkatkan efisiensi dan keakuratan proses seleksi beasiswa, membangun model data mining untuk mengidentifikasi calon penerima secara efisien dan objektif. Penelitian ini memberikan kontribusi penting pada bidang Informatika dengan menerapkan teknik data mining canggih. Analisis menggunakan algoritma Naïve Bayes menunjukkan akurasi sebesar 90%, mengindikasikan potensi penggunaannya dalam seleksi penerima beasiswa. Temuan penelitian juga mencakup perbandingan algoritma C4.5 dengan K-NN

dalam penentuan penerima beasiswa di STMIK Bani Saleh, dengan C4.5 dinilai lebih baik. Tujuan utama penelitian ini adalah meningkatkan efisiensi dan akurasi dalam seleksi penerima beasiswa PIP dengan menerapkan algoritma data mining, mencari hasil terbaik dari perbandingan menggunakan algoritma C4.5 dan *naïve bayes* Diharapkan dapat mengenali pola kompleks dalam data, termasuk informasi sosioekonomi dan pendidikan, untuk meramalkan calon penerima beasiswa PIP secara lebih akurat. Signifikansi penelitian ini terletak pada kontribusinya terhadap peningkatan kualitas kebijakan pemberian beasiswa dan pengambilan keputusan berbasis data.

Penelitian ini relevan bagi praktisi di lembaga-lembaga pendidikan dan pemerintah, dapat meningkatkan efisiensi dalam pemilihan penerima PIP, mendukung alokasi sumber daya yang tepat, dan bahkan menjadi dasar untuk pengembangan sistem otomatis di masa mendatang. Bagi peneliti di bidang Informatika, penelitian ini memberikan wawasan tentang pemanfaatan teknik data mining dalam pengambilan keputusan sosial. Keseluruhannya, penelitian ini berpotensi memberikan dampak luas, dari aplikasi praktis dalam seleksi penerima PIP hingga kontribusi pada pemahaman teknologi data mining dalam konteks sosial.

2. TINJAUAN PUSTAKA

Pada penelitian-penelitian sebelumnya yang berkaitan dengan kepuasan dan penggunaan algoritma C4.5 dan Naïve Bayes, dapat dilihat pada analisis

kepuasan penumpang pada maskapai penerbangan menggunakan algoritma C4.5 dan Naïve Bayes. memperoleh hasil akurasi 81,19% dan 76,24% untuk algoritma Naïve Bayes [8]. Hasil yang tidak berbeda didapat dalam perbandingan metode klasifikasi C4.5 dan Naïve Bayes untuk mengukur kepuasan pelanggan hasil yang diperoleh adalah C4.5 mempunyai nilai accuracy sebesar 94, 17%, lebih besar dibandingkan Naïve Bayes dengan nilai accuracy 85, 83% [9]. Namun pada penelitian mengenai perbandingan algoritma C4.5 dan Naive Bayes dalam klasifikasi tingkat kepuasan mahasiswa terhadap pembelajaran daring. Pada hasil perbandingan kedua algoritma bahwa tingkat akurasi dari algoritma Naïve Bayes lebih baik dibandingkan dengan C4.5 dengan selisih 11.77%. Nilai akurasi Algoritma Naïve Bayes dengan nilai 70.59%. Sedangkan tingkat akurasi dari algoritma C4.5 adalah 58.82% [4]

2.1. Data Mining

Data mining merupakan pencarian pada pola atau aturan dari sebuah data yang kapasitasnya sangat besar untuk menemukan informasi baru. Data mining dapat diartikan sebagai ilmu penggalian informasi yang berguna dari dataset atau database yang berukuran besar [5]. Data mining juga dapat disebut sebagai proses yang digunakan untuk menentukan sebuah struktur data Karakteristik dari data mining yaitu : (i) berkaitan dengan penemuan sesuatu yang baru yang tidak diketahui sebelumnya terkait pola data tertentu; (ii) data yang digunakan merupakan data yang besar untuk membuat hasil yang besar; (iii) dapat digunakan untuk penentuan strategi dan keputusan yang kritis. [6]

2.2. Decision Tree

Decision Tree akan memperlihatkan faktor-faktor kemungkinan (probabilitas) yang akan mempengaruhi alternatif/alternatif prestasi belajar siswa, disertai dengan prediksi hasil akhir yang akan didapat bila faktor-faktor dalam Decision Tree terpenuhi. Decision Tree akan mengubah data kedalam bentuk visual berupa diagram pohon dan aturanaturan keputusan. Data dalam Decision Tree dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Salah satu atribut yang merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan instance. Model Decision Tree merupakan salah satu metode klasifikasi dalam data mining yang menghasilkan klasifikasi berupa pohon keputusan berdasarkan sampel data dapat berupa data predictor univariat ataupun data multivariat [1] Alur proses analisis dalam decision tree adalah mengubah bentuk data (table) menjadi model tree, mengubah model tree menjadi rule dan menyederhanakan rule (pruning). Data yang diambil dalam penelitian ini adalah populasi sejumlah 416 siswa akan digunakan untuk membuat model prediksi Decision Tree. Model yang telah dibuat

kemudian akan dihitung tingkat akurasi prediksinya. [7]

2.3. Naïve Bayes` s

Naive Bayes merupakan sebuah pengklasifikasian atau penggolongan data yang menghitung kemungkinan dari dataset yang tersedia. Sedangkan menurut Bustami (2013) Naive Bayes merupakan pengklasifikasian untuk memprediksi peluang masa depan dengan metode probabilitas dan statistik sesuai dengan pengalaman di waktu sebelumnya. Kelebihan dari algoritma Naive Bayes adalah data yang di butuhkan untuk menetapkan perkiraan parameter dalam proses penggolongan dalam menggunakan metode ini hanya memerlukan jumlah data pelatihan yang kecil . Sedangkan menurut Syarli kelebihan Naive Bayes yaitu mudah diimplementasikan dan pada banyak kasus memberikan hasil yang baik, kemudian kekurangannya yaitu tidak terkaitnya antar fitur atau bersifat independent, sedangkan pada kenyataannya keterkaitan itu harus ada dan tidak dapat dimodelkan oleh Naive Bayesian Classifier. [6] Jadhav et all (2016) menyatakan bahwa Naïve Bayes Classifier adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema Bayes. [8]

3. METODE PENELITIAN

Metode penelitian merupakan suatu faktor yang sangat penting untuk memecahkan suatu masalah dan turut menuntukan keberhasilan suatu penelitian. Metode penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu Penelitian yang peneliti gunakan peneliti ambil sesuai dengan judul yang merupakan metode klasifikasi Decision Tree data mining dengan algoritma C4.5, klasifikasi Decision Tree merupakan suatu metode pengelompokan atau classification data yang sangat mudah dimengerti Ada banyak metode data mining yang digunakan untuk tujuan yang berbeda. Metode klasifikasi digunakan untuk membantu memahami pengelompokan data. Pada penelitian ini peneliti juga menggunakan metode algoritma C4.5, Algoritma C4.5 merupakan salah satu algoritma yang secara luas digunakan, khususnya di area machine learning yang memiliki beberapa perbaikan dari algoritma sebelumnya yaitu ID3. Algoritma C4.5 dan ID3 merupakan model yang tak terpisahkan, karena membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Menjelaskan kronologis penelitian, termasuk desain penelitian, prosedur penelitian (dalam bentuk algoritma, Pseudocode atau lainnya), bagaimana untuk menguji dan akuisisi data [6]

Metode yang digunakan dalam penelitian ini untuk Memprediksi penerima PIP SD N 04 Majalangu adalah jenis penelitian Kuantitatif. Penelitian kuantitatif mengandalkan proses pengumpulan dan analisis data kuantitatif, dimana data ini dapat diukur dalam bentuk angka dan dapat dianalisis melalui

pendekatan statistik. Penelitian yang dilakukan oleh penulis termasuk dalam penelitian Kuantitatif. Menurut Creswell 2009 Penelitian kuantitatif adalah metode - metode untuk menguji teori - teori tertentu dengan cara meneliti hubungan antar Variabel. Variabel biasanya diukur dengan instrumen penelitian sehingga data yang terdiri dari angka - angka dapat dianalisis berdasarkan prosedur - prosedur statistik.

3.1. Deskripsi Aktifitas Metode Penelitian

3.1.1. Perencanaan

Menentukan jenis data yang akan Anda butuhkan penelitian ini. Ini dapat mencakup data demografis, data ekonomi, data akademik, atau variabel-variabel lain yang relevan. Juga, tentukan sumber data tersebut. Data mining merupakan langkah analisis terhadap proses pengetahuan didalam basis knowledge discovery in database (KDD).[2]

3.1.2. Pengumpulan Data

Mengumpulkan data, termasuk metode survei, wawancara, pengumpulan data sekunder, atau penggunaan sistem informasi sekolah, Memiliki izin dan persetujuan yang diperlukan untuk mengakses data dan mulai mengambil data di lapangan.

3.1.3. Pengolahan Data

Pengolahan Data awal . initial Data Processing, Mempertimbangkan bagaimana data awal akan diolah sebelum digunakan, dalam analisis. Ini mungkin mencakup pemisahan data, perubahan format, atau identifikasi masalah data awal.

3.1.4. Seleksi Atribut

Perancangan Eksperimen, Pemilihan Tools dan Perangkat Lunak, Perangkat lunak atau tools yang digunakan untuk menerapkan algoritma C4.5 dan analisis data.

3.1.5. Pembagian Data

Perangkat lunak atau tools yang digunakan untuk menerapkan algoritma C4.5 dan analisis data. Pembagian data menjadi Training set, data test set kemudian memisahkan label yang ingin di prediksi. Memisahkan fitur atau variabel input yang berperah dalam membuat prediksi atau mengidentifikasi pola, Cross-Validation (Opsional).

Membagi data menjadi subset dan melakukan pelatihan serta evaluasi secara bergantian untuk menghindari overfitting. Data yang telah melalui tahapan ini siap untuk digunakan dalam pelatihan model pada tahap selanjutnya dari siklus KDD, dengan beberapa tahap bersifat opsional tergantung pada karakteristik data dan tujuan analisis.

3.2. Sumber Data

Dalam penelitian ini, digunakan sumber data sekunder yang diperoleh melalui kesepakatan antara penulis dan komunitas. Data yang diambil berasal dari komunitas yang setuju untuk berbagi data. Data yang

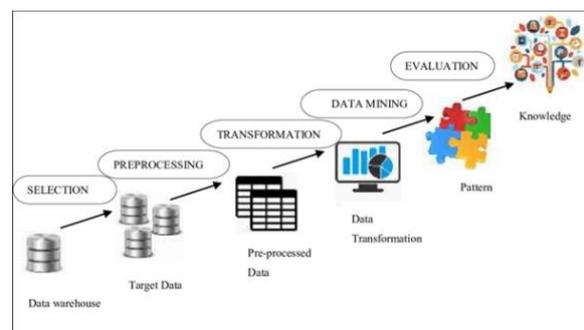
digunakan termasuk dalam kategori data sekunder dan mencakup dataset sebanyak 400 sampel data, melibatkan Data siswa dan orangtua siswa dari SD N 04 Majalangu.

3.3. Teknik Pengumpulan Data

- Data mengenai siswa atau siswi SD N 04 Majalangu diperoleh melalui persetujuan yang dibuat selama penelitian di SD N 04 Majalangu.
- Berdasarkan kesepakatan tersebut, dataset dapat digunakan untuk penelitian dalam bentuk file Excel dan kemudian disimpan ke dalam aplikasi Rapid Miner sebagai dataset.
- Data studi dokumentasi, seperti jurnal, makalah, prosiding, dan artikel lainnya, diperoleh berdasarkan keterkaitannya dengan penelitian yang sedang dilakukan

3.4. Teknik Analisis Data

Penelitian ini mengadopsi metode Knowledge Discovery In Database (KDD) yang melibatkan serangkaian tahapan dan proses sebagai berikut :



Gambar 1. Alur Penelitian Proses KDD

3.4.1. Selection

Tidak semua data diperlukan dalam proses data mining. Proses seleksi dilakukan pada data Pemilihan Penerima Program Indonesia Pintar (PIP) di SD N 04 Majalangu pada tahun 2021 dan 2022, mencakup 400 rekaman dengan 8 atribut, diantaranya nama siswa, tahun, nama ayah kandung, nama ibu kandung, jumlah pendapatan, jumlah tanggungan, pekerjaan, Jenis kelamin. Pengolahan data pada tahap ini menggunakan Microsoft Excel dan hasilnya dimasukkan ke aplikasi Rapid Miner.

3.4.2. Preprocessing

Tahap berikutnya adalah pre-processing, yang mencakup pembersihan data (Data Cleaning) untuk menangani data yang tidak valid atau duplikat, memastikan kualitas data yang baik. Data yang digunakan harus melewati data preprocessing, karena sumber data yang diperoleh langsung dari sumber masih bersifat kotor, artinya terdiri dari beberapa data yang kurang lengkap, data hilang atau kosong, kekurangan atribut tertentu atau atribut yang sesuai, data noise, dan data yang tidak konsisten.[9]

3.4.3. Transformation

Pada tahap transformasi data, kategori data diubah sesuai dengan jenis data. Transformasi data melibatkan pengubahan format, skala, atau struktur data agar sesuai dengan kebutuhan analisis atau model yang akan digunakan.

3.4.4. Data Mining

Di tahap data mining, analisis dan perhitungan dilakukan menggunakan algoritme seperti *Naive Bayes* dan *Decision Tree*. Dataset dibagi menjadi data training dan data testing.

3.4.5. Interpretasi / Evaluasi.

Tahap ini melibatkan evaluasi dari aturan yang ditemukan sesuai dengan pola dan tahapan yang telah ditentukan. Aplikasi Rapid Miner digunakan untuk menentukan hasil evaluasi menggunakan algoritme *Naive Bayes* dan *Decision Tree*.

4. HASIL DAN PEMBAHASAN

4.1. Data

No Tahun	Nama_siswa	JK	Nama_ibu_kandung	Nama_Ayah	pekerjaan	Jumlah_tanggungan	Pendapatan	Penetapan
1	2022	ALFARREL ZIGHY P	L	AL ROHMAH	ARIS MUJIARTO	PETANI	SEDANG	CUKUP DITERIMA
2	2022	IFTI MEYLINA PUTI	P	RIYATI	DEI PURNOMO	PENGAGANG	BANYAK	CUKUP DITERIMA
3	2022	MUHAMMAD ROZ	L	TURAHMI	TARMONO	WIRASWASTA	BANYAK	CUKUP DITERIMA
4	2022	AZKA ANUROFIQ	P	TASKIAH	TUKUMAN	WIRASWASTA	SEDANG	RATA-RATA DITERIMA
5	2022	IZZA NAFISATUL AL	L	MASRUROH	ANOTO	PRANOTO	SEDKIT	CUKUP DITERIMA
6	2022	MUHAMMAD SOH	L	MWIT MURTINI	CASSMO	WIRASWASTA	SEDANG	CUKUP DITERIMA
7	2022	AZRI NUBUHH	P	ISKANTON	TRAFUR	BURUH	SEDKIT	CUKUP DITERIMA
8	2022	HUMAMAH DHALL	L	ROKHATIN	KODMAN	BURUH	SEDKIT	RATA-RATA TIDAK DITERIMA
9	2022	LUTHFI ADITYA FIR	L	SRI ASIH	SUGIVANTO	WIRASWASTA	SEDKIT	TINGGI DITERIMA
10	2022	ATIF MUSRIFAH	P	RIKHATIN	SUPRIYO	PENGUSAHA	BANYAK	RATA-RATA DITERIMA
11	2022	ULFAN IBNU MARI	L	SITI JUHROH	RASWANTO	PENGUSAHA	SEDKIT	RATA-RATA TIDAK DITERIMA
12	2022	DIFA PRISTIA ME	P	PRHATIN	ISHAQI	WIRASWASTA	BANYAK	CUKUP DITERIMA
13	2022	BAYHAN ATHIF ISA	L	SATUN NUUR KHASAN	RISYANTO	PETANI	BANYAK	CUKUP DITERIMA
14	2022	HANFATUL ISTIQ	P	WARISNI	SODIH	PETANI	BANYAK	CUKUP DITERIMA
15	2022	DAFFA RIDZY RAM	L	IKHAMIDAH	JOKO HARIYADI	PETANI	BANYAK	CUKUP DITERIMA
16	2022	MIZANUN AFKAR	L	ASTITI	SISWANTO	PETANI	SEDKIT	RATA-RATA DITERIMA
17	2022	ALFARREL ZIGHY P	L	AL ROHMAH	ARIS MUJIARTO	PETANI	SEDANG	TINGGI DITERIMA
18	2022	ADI PRASETYO	L	TURNATI	SONI	TUKANG	SEDANG	CUKUP DITERIMA
400	2023	LUTHI SAKHI AZIDA	L	SUDIRAH	PRANOTO	BURUH	SEDANG	RATA-RATA TIDAK DITERIMA

Gambar 2. Data Set

Dalam penelitian ini, dataset yang digunakan mencakup informasi mengenai siswa atau siswi SD N 04 Majalangu. Data ini diperoleh melalui kesepakatan yang terjadi selama penelitian di SD N 04 Majalangu. Data yang diperoleh melalui Perizinan dari dengan Kepala Sekolah SD N 04 Majalangu terdiri dari 400 entri data, dengan 8 atribut termasuk tahun, nama siswa, nama ibu kandung, nama ayah kandung, jumlah penghasilan, pekerjaan, jumlah tanggungan, dan jenis kelamin

4.2. Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) merujuk pada proses mengekstraksi pengetahuan yang bermanfaat, pola, dan informasi tersembunyi dari sekumpulan data yang besar dan kompleks. KDD melibatkan serangkaian langkah atau tahapan untuk mengubah data mentah menjadi pengetahuan yang dapat digunakan untuk pengambilan keputusan.

Berikut adalah ringkasan dari langkah-langkah umum dalam proses KDD:

- Pemahaman Tujuan: Menetapkan tujuan analisis dan kebutuhan bisnis yang ingin dicapai.
- Pemahaman Data: Mengumpulkan data yang relevan untuk analisis, memahami karakteristiknya, serta menentukan format dan struktur data.

- Pembersihan Data: Menangani data yang tidak lengkap, tidak akurat, atau tidak konsisten, dan melakukan pre-processing untuk mempersiapkan data.
- Integrasi Data: Menggabungkan data dari berbagai sumber jika diperlukan untuk menciptakan satu set data yang utuh.
- Reduksi Data: Mengurangi kompleksitas data dengan memilih fitur-fitur yang paling relevan atau menggunakan teknik reduksi dimensi.
- Transformasi Data: Melakukan transformasi pada data untuk mendukung proses analisis, seperti normalisasi atau pengkodean ulang.
- Pengembangan Model: Menerapkan teknik-teknik pembelajaran mesin atau statistik untuk mengembangkan model yang dapat mengekstraksi pola atau melakukan prediksi.
- Evaluasi Model: Mengukur kinerja model dan memvalidasi apakah hasil yang diperoleh sesuai dengan tujuan awal.[3]
- Interpretasi dan Pemahaman: Mengevaluasi dan menginterpretasikan temuan untuk mendapatkan pemahaman yang lebih dalam, serta menerapkan pengetahuan tersebut ke dalam konteks bisnis.
- Penggunaan Pengetahuan: Menggunakan pengetahuan yang ditemukan untuk mendukung pengambilan keputusan dan meningkatkan kinerja organisasi.

Secara umum, KDD melibatkan kerja sama antara peneliti, analis data, dan domain expert untuk menghasilkan informasi yang bernilai dari data yang tersedia.

4.2.1. Data Selection

Seleksi data merupakan langkah awal dalam proses Knowledge Discovery in Database (KDD), di mana dataset dipilih sesuai dengan tujuan penelitian. Pada tahap ini, ditentukan target atau variabel yang akan diprediksi oleh model Algoritma yang sedang digunakan.

Atribut yang tidak relevan dalam proses data mining adalah atribut "Nama" "Tahun" "Jenis Kelamin" karena tidak memiliki pengaruh pada proses data mining. Biasanya, atribut nama tidak menyumbang informasi yang signifikan secara statistik dan tidak dapat digunakan untuk membuat Keputusan. Atribut nama ini cenderung tidak memberikan informasi berharga dalam analisis yang dilakukan.

Atribut Jenis Pekerjaan termasuk kategori karena terdapat beberapa pilihan, Atribut Jumlah pendapatan Bukan Tipe numerik karena di dalam data set terdapat 4 pilihan diantaranya "kurang", "Cukup", "sedang", "Tinggi". Atribut Jumlah Tanggungan bertipe kategori karena terdapat beapa pilihan seperti "Sedikit", "Sedang", "Banyak".

untuk mengenali pola yang signifikan tanpa membuatnya menjadi terlalu rumit. Melakukan eksperimen dengan berbagai nilai dan mengevaluasi kinerja model merupakan metode yang efektif untuk menemukan nilai yang paling sesuai.

Pemilihan nilai confidence dalam Decision Tree, seperti nilai 0.1, mempengaruhi keputusan yang dibuat oleh model dan struktur pohon keputusan yang dikonstruksi. Nilai confidence ini terkait dengan tingkat keyakinan yang diperlukan agar suatu keputusan dianggap sah.

Rumus yang digunakan untuk menghitung nilai accuracy :

a) C4.5 Decision Tree

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN}$$

Sumber : [11]

True Positive (TP) : Jumlah kasus yang diprediksi sebagai "diterima" dan benar-benar "diterima." (TP = 0), True Negative (TN): Jumlah kasus yang diprediksi sebagai "tidak diterima" dan benar-benar "tidak diterima." (TN = 134), False Positive (FP): Jumlah kasus yang diprediksi sebagai "diterima" tetapi sebenarnya "tidak diterima." (FP = 0), False Negative (FN): Jumlah kasus yang diprediksi sebagai "tidak diterima" tetapi sebenarnya "diterima." (FN = 74).

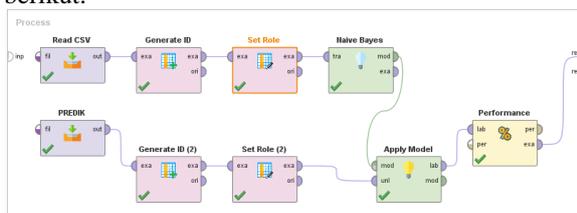
Rumus perhitungan nilai accuracy menggunakan Algoritme Decision Tree :

$$Accuracy = \frac{134 + 0}{134 + 0 + 0 + 74} = 0,644$$

Setelah evaluasi dilaksanakan menggunakan algoritma Decision tree, hasilnya menunjukkan bahwa model dapat menghasilkan data dalam sesuai dengan hasil accuracy pada rapid miner, yang berisi pola prediksi terkait Penetapan Penerima Program Indonesia Pintar (PIP) untuk periode tahun 2023. Selanjutnya, dalam penelitian ini, akan dilakukan analisis terhadap hasil prediksi, yang mengindikasikan bahwa terdapat 70 siswa yang diterima dan 138 siswa yang tidak diterima pada tahun 2023.

b) Naïve Bayes :

Proses data mining dengan menggunakan algoritma Naïve Bayes memiliki urutan sebagai berikut:



Gambar 8. Proses Naïve Bayes

Setelah menyesuaikan pengaturan dan menjalankan langkah-langkah dalam proses di atas, diperoleh hasil evaluasi kinerja dari model Naive Bayes. Evaluasi ini didasarkan pada data yang digunakan dalam dataset tersebut. Hasil evaluasi ini memberikan informasi tentang seberapa baik model *Naive Bayes* dapat memprediksi atau mengklasifikasikan data berdasarkan karakteristik dari dataset yang telah diolah.

Rumus yang digunakan untuk menghitung nilai accuracy Naïve Bayes :

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN}$$

Sumber : [11]

True Positive (TP): Jumlah kasus yang diprediksi sebagai "diterima" dan benar-benar "diterima." TP=69. True Negative (TN): Jumlah kasus yang diprediksi sebagai "tidak diterima" dan benar-benar "tidak diterima." TN=139. False Positive (FP): Jumlah kasus yang diprediksi sebagai "diterima" tetapi sebenarnya "tidak diterima." FP=0. False Negative (FN): Jumlah kasus yang diprediksi sebagai "tidak diterima" tetapi sebenarnya "diterima." FN=0.

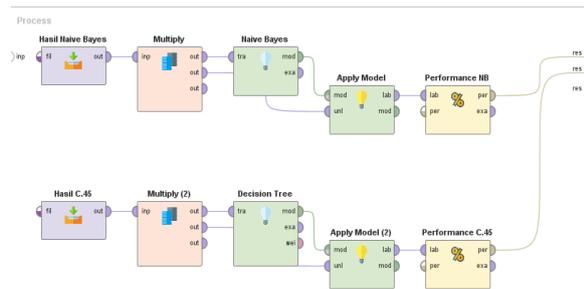
Rumus yang digunakan untuk menghitung nilai accuracy Naïve Bayes :

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN}$$

$$Accuracy = \frac{69 + 139}{69 + 0 + 0 + 139} = \frac{208}{208} = 1$$

Setelah evaluasi dilaksanakan menggunakan algoritma *Naïve Bayes*, hasilnya menunjukkan bahwa model dapat menghasilkan data dalam bentuk tabel prediksi yaitu 1, atau 100% sesuai dengan hasil accuracy pada rapid miner

Setelah melakukan prediksi dalam proses data mining dan mendapatkan hasil prediksi tahun 2023, langkah selanjutnya adalah melakukan evaluasi klasifikasi untuk melihat sejauh mana model yang dibangun dapat menghasilkan prediksi yang akurat. Berikut adalah langkah-langkah umum yang dapat diambil setelah prediksi untuk mengevaluasi klasifikasi dan mengukur akurasi:



Gambar 9. Proses Klasifikasi data hasil prediksi

Evaluasi performa model klasifikasi merupakan langkah krusial dalam mengklasifikasikan data hasil prediksi. Setelah model melakukan prediksi terhadap dataset uji atau data baru, Hal ini bertujuan untuk menghasilkan nilai akurasi yang mencerminkan sejauh mana algoritma telah berhasil dalam proses klasifikasi tersebut.

a. Hasil Klasifikasi Accuracy C4.5 Decision Tree

Metode	Accuracy	Precision	Recall
Decision Tree	64,42%	64,42%	100%

accuracy: 64.42%

	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	0	0	0.00%
pred. TIDAK DITERIMA	74	134	64.42%
class recall	0.00%	100.00%	

Gambar 10. Hasil Accuracy C4.5

Pada bagian result, terdapat hasil dengan menampilkan akurasi sebesar 64.42%, Class precision pada prediksi diterima memiliki nilai 0.00% sedangkan pada prediksi tidak diterima memiliki nilai 64.00%. class recal pada true diterima memiliki nilai 0.00%, sedangkan pada true tidak diterima memiliki nilai 100.00%

Dengan keterangan Sebagai berikut :

- Jumlah prediksi diterima dan ternyata true diterima adalah sebanyak 0 record
- Jumlah prediksi diterima dan ternyata true tidak diterima adalah sebanyak 0 record.
- Jumlah prediksi tidak diterima dan ternyata true diterima adalah sebanyak 74 record
- Jumlah prediksi tidak diterima dan ternyata true tidak diterima adalah sebanyak 134 record.

b. Hasil Klasifikasi Accuracy Naive Bayes

accuracy: 100.00%

	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	69	0	100.00%
pred. TIDAK DITERIMA	0	139	100.00%
class recall	100.00%	100.00%	

Gambar 11. Hasil Accuracy Naive Bayes

Pada bagian result, terdapat hasil dengan menampilkan akurasi sebesar 100.00%, Class precision pada prediksi diterima memiliki nilai 100.00% sedangkan pada prediksi tidak diterima memiliki nilai 100.00%. class recal pada true diterima memiliki nilai 100.00%, sedangkan pada true tidak diterima memiliki nilai 100.00%

Dengan keterangan Sebagai berikut :

- Jumlah prediksi diterima dan ternyata true diterima adalah sebanyak 69 record
- Jumlah prediksi diterima dan ternyata true tidak diterima adalah sebanyak 0 record.
- Jumlah prediksi tidak diterima dan ternyata true diterima adalah sebanyak 0 record.
- Jumlah prediksi tidak diterima dan ternyata true tidak diterima adalah sebanyak 139 record.

Hasil pengujian terhadap model klasifikasi menunjukkan tingkat akurasi dari model algoritma C4.5 sebesar 100% dan tingkat akurasi model algoritma Naive Bayes sebesar 64,42% Perbandingan tingkat akurasi sejalan dengan hasil penelitian sejenis di tempat lain dimana tingkat akurasi model algoritma C4.5 lebih tinggi dari model algoritma Naive Bayes.

Hasil dari kedua algoritma yang diterapkan menunjukkan tingkat akurasi Naive Bayes dan Decision Tree. Masing-masing mencapai akurasi, yaitu 100% untuk Naive Bayes dan 64,42% untuk Decision Tree. Dengan demikian, dapat disimpulkan dari hasil tersebut bahwa

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dalam bidang data mining, perbandingan klasifikasi antara Algoritma Naive Bayes dan Algoritma Decision Tree pada dataset pemilihan penerima Program Indonesia Pintar telah diuji menggunakan 400 data. Hasil pengujian tersebut menghasilkan penemuan bahwa kinerja algoritma Naive Bayes lebih unggul daripada Algoritma C4.5 Decision Tree dengan tingkat akurasi masing-masing sebesar 100% dan 64,42%. Temuan ini sejalan dengan penelitian serupa di lokasi lain. Untuk menentukan model klasifikasi terbaik di antara algoritma-algoritma yang akan diimplementasikan, disarankan merujuk pada penelitian lain yang telah membandingkan kinerja berbagai model algoritma yang mungkin dipilih. Sementara itu, untuk membandingkan kinerja dari segi waktu proses (execution time), disarankan menggunakan dataset yang mencakup jumlah data yang memadai guna mengidentifikasi waktu proses dengan lebih akurat. Dalam penelitian berikutnya, terdapat beberapa rekomendasi yang dapat diajukan, yakni optimalisasi pemilihan Algoritma dengan menggunakan Algoritma Decision Tree sebagai pilihan terbaik untuk klasifikasi, serta melakukan lebih banyak eksperimen dengan Algoritma selain Naive Bayes dan Decision Tree yang digunakan, C4.5 secara alami mendukung data kategorikal. Pastikan untuk menangani data

kategorikal dengan benar, misalnya, dengan mengubahnya menjadi data numerik atau menggunakan teknik seperti one-hot encoding. Jika data numerik digunakan dalam model Naive Bayes, pertimbangkan transformasi data atau pengelompokan interval untuk memfasilitasi asumsi probabilitas. mengingat setiap Algoritma memiliki keunikan dan karakteristik yang berbeda-beda.

DAFTAR PUSTAKA

- [1] A. Supriyadi, "Perbandingan Algoritma Naive Bayes dan Decision Tree(C4.5) dalam Klasifikasi Dosen Berprestasi," *Gener. J.*, vol. 7, no. 1, pp. 39–49, 2023, doi: 10.29407/gj.v7i1.19797.
- [2] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 2, pp. 191–196, 2021, doi: 10.31294/p.v23i2.11395.
- [3] M. R. Fahdia, N. Asnonik, and S. A. P. Putri, "Implementasi Metode Naive Bayes Pada Analisa Penerimaan Kartu Jakarta Pintar Di Smk Al Kautsar Jakarta," *INTI Nusa Mandiri*, vol. 14, no. 2, pp. 231–236, 2020, doi: 10.33480/inti.v14i2.1779.
- [4] D. Toresa *et al.*, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Mengukur Tingkat Kepuasan Mahasiswa Dalam Penggunaan Edlink," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 3, pp. 250–256, 2023, doi: 10.47233/jteksis.v5i3.855.
- [5] S. Chodijah and M. Iqbal, "Penentuan Penerima Beasiswa Di Stmik Bani Saleh Dengan Perbandingan Metode Algoritma C4.5 Dan Knearest Neighbors," *J. Inf. dan Komput.*, vol. 10, no. 1, pp. 107–114, 2022, doi: 10.35959/jik.v10i1.295.
- [6] F. Solikhah, M. Febianah, A. L. Kamil, W. A. Arifin, and Shelly Janu Setyaning Tyas, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *Tematik*, vol. 8, no. 1, pp. 96–103, 2021, doi: 10.38204/tematik.v8i1.576.
- [7] J. Simada, S. Informasi, B. Data, and V. No, "PERBANDINGAN PENERAPAN ALGORITMA DECISION TREE C . 45 DAN NAÏVE BAYES DALAM ANALISA KELULUSAN SISWA PADA SMK SWADHIPA 2 NATAR," vol. 6, no. 2, pp. 117–123, 2023.
- [8] N. Nanni and A. Sudransyah, "Perbandingan Kinerja Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Harga Pangan," *PROtek J. Ilm. Tek. Elektro*, vol. 7, no. 1, pp. 20–24, 2020, doi: 10.33387/protk.v7i1.1710.
- [9] R. M. Debora *et al.*, "Application of C . 45 Algorithm for Potential Students Classification Drop Out at Budi Luhur University," vol. 2, no. April, pp. 316–325, 2023.
- [10] K. Cirebon, H. Rudi, and M. Modification, "Untuk Menentukan Introvert Dan Ekstrovert Pada," vol. 6, no. 1, pp. 74–83, 2021.
- [11] A. H. Wicaksono *et al.*, "Klasifikasi Siswa Slow Learner Untuk Mendukung Sekolah Algoritma Naïve Bayes Classification of Slow Learner Students To Support Schools in Improving Student Understanding Using the Naïve Bayes," vol. 9, no. 3, pp. 589–596, 2022, doi: 10.25126/jtiik.202295509.