

ANALISIS PENERAPAN ALGORITMA C4.5 DALAM PENENTUAN SISWA PENERIMA BEASISWA KARAWANG CERDAS (STUDI KASUS : SMK PGRI CIKAMPEK)

Aditya Nugraha, Iqbal Maulana, Purwantoro, Agung Susilo Yuda Irawan, Apriade Voutama

Informatika, Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

1910631170152@student.unsika.ac.id

ABSTRAK

Program beasiswa Karawang Cerdas memberikan bantuan pendidikan bagi masyarakat khususnya daerah Kabupaten Karawang mulai dari tingkat sekolah dasar hingga perguruan tinggi. Penerima beasiswa tersebut belum tepat sasaran yang hanya berfokus pada satu kategori yaitu Keluarga Tidak Mampu (KTM), namun banyak siswa yang memiliki potensi sebagai penerima beasiswa dari kategori lain seperti siswa berprestasi maupun orang tua yang memiliki latar belakang ASN/TNI/POLRI. Penelitian ini menggunakan metode *data mining* dengan algoritma C4.5 untuk mengklasifikasikan siswa SMK PGRI Cikampek sebagai penerima beasiswa, melalui pendekatan *Knowledge discovery in database* (KDD). Data yang digunakan pada penelitian ini merupakan data demografis siswa kelas X, XI, dan XII tahun ajaran 2022/2023 yang mencakup atribut seperti tempat lahir, alat transportasi, pekerjaan orang tua, penghasilan per golongan, penerima KIP, jumlah tanggungan, status kepemilikan rumah, prestasi, dan status kelayakan. Metodologi penelitian mencakup seleksi data, pra-pemrosesan data, transformasi data, *data mining*, dan evaluasi. Seleksi data memilih 10 atribut relevan dari total 19. Atribut Pra-pemrosesan melibatkan pembersihan data tidak relevan, penanganan *missing values*, dan konversi atribut kategorikal menjadi numerik. *Dataset* dibagi menjadi tiga skenario untuk *training* dan *testing* yaitu skenario A (90:10), skenario B (80:20), dan skenario C (70:30). Hasil akurasi tertinggi diperoleh skenario A mencapai 96.08%. Evaluasi menggunakan *Confusion matrix* dan *Classification Report* menunjukkan rata-rata akurasi 94.31% melalui *K-Fold Cross Validation*. Penelitian ini berhasil merancang pola seleksi penerima beasiswa dengan akurasi tinggi, berkontribusi penting bagi proses seleksi beasiswa di SMK PGRI Cikampek.

Kata kunci : *Beasiswa Karawang Cerdas, Algoritma C4.5, Knowledge discovery in database, Data mining*

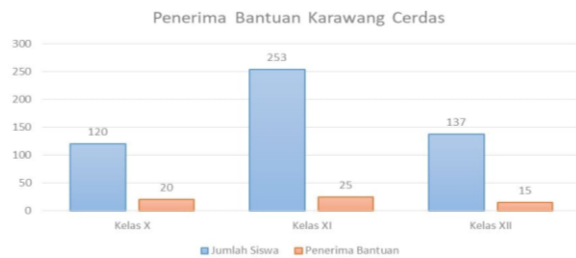
1. PENDAHULUAN

Pendidikan memerlukan peran dari berbagai pihak salah satunya peran pemerintah dalam penyebaran dan peningkatan kualitas pendidikan sesuai dengan konstitusi yang berlaku, sebagaimana termaktub pada UUD 1945 pasal 31 ayat (3) dan (4), yaitu pemerintah berkewajiban menjamin terselenggaranya pendidikan bagi masyarakat dengan mengalokasikan 20% dari Anggaran Pendapatan dan Belanja Nasional/Daerah (APBN/APBD) untuk mendukung pembiayaan pendidikan [1]. Sebagai respon terhadap hal tersebut, pemerintahan Kabupaten Karawang mengadakan program beasiswa Karawang Cerdas guna memberikan dukungan serta bantuan berupa beasiswa sekolah kepada masyarakat di Kabupaten Karawang [2].

Pemerintah Daerah Kabupaten Karawang, melalui Dinas Pendidikan Pemuda dan Olahraga membentuk sebuah kebijakan publik berupa Program Beasiswa Karawang Cerdas untuk meningkatkan kualitas sumber daya manusia di Kabupaten Karawang, sebagaimana yang tertuang dalam peraturan bupati (PERBUP) Nomor. 66 Tahun 2018 tentang Program Karawang Cerdas, serta dikeluarkannya Surat Keputusan Kepala Dinas Pendidikan Pemuda dan Olahraga Nomor 420 tentang Petunjuk Teknis (Juknis) Beasiswa Karawang Cerdas [3]. Program Karawang Cerdas sudah berjalan sejak

tahun 2018 hingga saat ini, namun masih ditemukan permasalahan yang timbul yaitu Jumlah penduduk Kabupaten Karawang yang terus bertambah tetapi pemahaman masyarakat mengenai pendidikan masih kurang, sehingga banyak masyarakat karawang mengalami kesulitan memahami juknis yang ada, serta indikasi bahwa program Beasiswa Program Karawang Cerdas yang belum tepat sasaran. Hal tersebut dapat mengurangi efektifitas program beasiswa dalam meningkatkan daya serap siswa yang membutuhkan bantuan secara finansial dan siswa yang berprestasi disekolah [4].

SMK PGRI Cikampek menjadi salah satu sekolah yang menerima program bantuan Karawang Cerdas. SMK PGRI berlokasi di Jl. Ir. H Juanda No. 402 Kp. Sukamulya Ds. Jomin Barat Kec. Kotabaru Kab. Karawang. Dengan memiliki tiga bidang keahlian diantaranya Teknik Permeriharaan Mekanik Industri, Teknik Kendaraan Ringan, dan Rekayasa Perangkat Lunak. Total keseluruhan siswa di SMK PGRI Cikampek kelas X berjumlah 120 siswa, kelas XI berjumlah 253 siswa, dan kelas XII berjumlah 137 siswa. Sedangkan data siswa penerima beasiswa Karawang Cerdas di SMK PGRI Cikampek dapat dilihat pada Gambar 1.



Gambar 1. Data Perbandingan Siswa Penerima Beasiswa Karawang Cerdas (Sumber: Staff Administrasi SMK PGRI Cikampek)

Pemerintah telah berupaya memberikan program bantuan beasiswa Karawang Cerdas, namun masih banyak indikasi bahwa bantuan Beasiswa Program Karawang Cerdas yang masih belum tepat sasaran, seharusnya banyak siswa memiliki potensi sebagai penerima beasiswa dari kategori lain seperti siswa berprestasi maupun orang tua yang memiliki latar belakang ASN/TNI/POLRI.

Kasus tersebut terdapat di SMK PGRI Cikampek, di mana hal ini tersaji pada gambar 1.1 yang jelas terlihat bahwa penyerapan siswa penerima beasiswa Karawang Cerdas hanya sedikit dibandingkan dengan jumlah siswa aktif yang berada di SMK PGRI Cikampek. Hal ini disebabkan oleh pihak sekolah yang belum memaksimalkan adanya pendaftaran beasiswa Karawang Cerdas selain dari kategori Keluarga Tidak Mampu (KTM), sehingga hasil dari siswa yang berkesempatan mendapatkan beasiswa belum optimal.

Penelitian ini mencoba untuk merancang pola penyeleksian beasiswa karawang cerdas di SMK PGRI Cikampek, sistem yang dibuat dianalisis menggunakan *data mining* khususnya metode klasifikasi dengan pendekatan algoritma C4.5.

Berdasarkan beberapa penelitian sebelumnya yang dilakukan [5] dengan judul “Pohon Keputusan C4.5 Algoritma Untuk Klasifikasi Program Bantuan Belajar”, melakukan proses pengklasifikasian dalam jumlah 124 kasus menggunakan algoritma pohon Keputusan C4.5 untuk menampilkan pola keputusan yang dapat menentukan mahasiswa tersebut diterima atau tidak sebagai penerima beasiswa, dari penelitian tersebut menunjukkan model klasifikasi yang cukup baik dengan tingkat akurasi, presisi dan recall 87% untuk semua bagian. Artinya, model tersebut berkinerja baik untuk diimplementasikan ke dalam sebuah sistem. Penelitian lainnya yang dilakukan [6] “Penerapan *Data mining* Klasifikasi C4.5 Pada Penerima Beasiswa di SMK Swasta Anak Bangsa”, penelitian tersebut melakukan proses klasifikasi algoritma C4.5 menggunakan bantuan software Rapidminer untuk menghasilkan model keputusan penentuan kelayakan penerima beasiswa.

Dari hasil pemaparan latar belakang sebelumnya, maka penulis mengusulkan penelitian yang berjudul “Analisis Penerapan Algoritma C4.5 dalam Penentuan Siswa Penerima Beasiswa Karawang Cerdas (Studi Kasus: SMK PGRI Cikampek)”. Penelitian ini bertujuan untuk melakukan penerapan algoritma

C4.5 dalam proses klasifikasi siswa penerima beasiswa di SMK PGRI Cikampek, serta memberikan kontribusi positif bagi sistem seleksi beasiswa yang lebih efisien dan akurat.

2. TINJAUAN PUSTAKA

2.1. Pendidikan

Pendidikan dapat mendorong kehidupan masyarakat dan pembangunan berkelanjutan. Namun dalam mencapai pendidikan yang berkualitas memerlukan pengeluaran dana yang besar [7]. Mastur [8] berpendapat bahwa keberadaan program beasiswa yang ditawarkan baik oleh pemerintah maupun swasta dapat meringankan permasalahan anggaran.

2.2. Beasiswa

Beasiswa merupakan salah satu bentuk apresiasi yang diberikan kepada seseorang agar dapat melanjutkan ke tahap pendidikan yang lebih tinggi, baik berupa kesempatan untuk bergabung dengan suatu instansi atau bantuan finansial yang diberikan untuk mendukung biaya Pendidikan [9].

Beasiswa sebagaimana didefinisikan oleh Suweleh, dkk [2] merupakan program bantuan keuangan yang diberikan dari pemerintah maupun swasta kepada perorangan, pelajar atau mahasiswa untuk membantu memenuhi biaya pendidikannya.

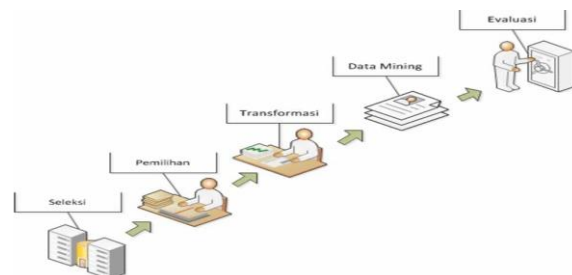
2.3. Data mining

Data mining merupakan teknik menemukan informasi tersembunyi dari sekumpulan data dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk menentukan informasi yang berpotensi sebagai ilmu pengetahuan [10].

Data mining menurut Risnawati & Rahmadani, [11] adalah proses menganalisis sekumpulan data untuk mengidentifikasi pola-pola penting dan menarik kesimpulan yang sebelumnya tidak diketahui agar dapat dipahami dan berguna bagi pemilik data. *Data mining* terbagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukannya [12].

2.4. Knowledge discovery in database (KDD)

Data mining merupakan bagian dari *Knowledge discovery in database* (KDD) yang merupakan proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya dan tersembunyi dari data [13].



Gambar 2. Knowledge discovery in database (KDD)

Proses *Knowledge discovery in database* melibatkan hasil dari *data mining* (proses ekstraksi pola dari data), yang kemudian mengonversi hasil tersebut menjadi informasi yang mudah dipahami secara akurat. Tahapan teknik *data mining Knowledge discovery in database* (KDD) pada gambar 2.

2.5. Supervised Learning Dan Unsupervised Learning

Supervised learning merupakan penggunaan data yang sudah diannotasi, di mana setiap sampel data dikaitkan dengan hasil yang diketahui sebelumnya. *Supervised learning* memiliki tujuan untuk melatih model komputer yang dapat mempelajari pola dalam data dan melakukan dugaan akurat terhadap data yang belum diketahui [15].

Unsupervised learning merupakan teknik yang digunakan dalam pembuatan Artificial Intelligence. Model pada pendekatan *Unsupervised Learning* tidak Algoritma *Unsupervised Learning* memiliki sifat deskriptif yang berguna dalam mengkategorikan data [16].

2.6. Klasifikasi

Budiarto dkk, [17] menyatakan klasifikasi adalah proses mengklasifikasikan objek atau konsep yang didasarkan dari objek atau konsep yang bersangkutan. Jadi dapat ditarik kesimpulan bahwa klasifikasi adalah pengumpulan sejumlah parameter ke dalam salah satu kelas yang sudah ditentukan sebelumnya.

2.7. Algoritma C4.5

Algoritma C4.5 merupakan teknik klasifikasi yang digunakan untuk membuat pohon keputusan dan aturan yang dapat meningkatkan akurasi prediksi, sehingga algoritma C4.5 mudah diaplikasikan dan mudah dimengerti [18]. Algoritma C4.5 adalah perkembangan dari algoritma ID3. Algoritma C4.5 menjadi pondasi dari pohon keputusan yang bersumber pada *dataset* training berisikan tupel dalam basis data [19].

2.8. Confusion matrix

Table khusus yang disebut *Confusion matrix* digunakan untuk menunjukkan seberapa baik suatu algoritma bekerja, setiap baris mewakili instance di kelas sebenarnya dan setiap kolom mewakili instance kelas prediksi [20]. *Confusion matrix* menurut Riyyan & Firdaus [21] merupakan teknik penjumlahan data yang digunakan dalam menghitung jumlah observasi pada setiap kelas secara akurat kemudian menampilkan hasilnya dalam model klasifikasi.

Confusion matrix mempunyai empat komponen penyusun yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Komponen penyusun dilakukan perhitungan untuk mendapatkan nilai *accuracy*, *recall*, *precision*, dan *F1-Score* [22].

2.9. Python

Python merupakan bahasa pemrograman yang bersifat interpretatif, berorientasi objek, dan memiliki stuktur kode yang dinamis. Python pertama kali diperkenalkan pada tahun 1991, dengan seiring berkembangnya bahasa pemrograman python saat ini menyediakan stuktur data tingkat tinggi, serta mendukung *dynamic typing* dan *dynamic binding* sas [23]. Menurut Mulyadien [24], Python memiliki banyak fungsi didalamnya, seperti membuat web, analisis data, scripting, serta didukung dengan adanya berbagai library untuk mempermudah proses *data mining* mulai dari pengumpulan data, *preprocessing* data, eksplorasi data, pemodelan hingga evaluasi.

2.10. Google Colaboratory

Google Colab merupakan perangkat yang sangat berguna bagi pemula dan pakar di bidang *data science*, *machine learning*, dan penanganan data secara umum. Google Colab telah menjadi pilihan populer untuk banyak proyek dan tugas komputasi berbasis Python karena sumber daya komputasi yang luas, kemudahan berbagi dan lingkungan yang interaktif [25].

Menurut pendapat Mulyadien [24], Google Collaboratory, adalah editor kode berbasis cloud yang tersedia secara gratis untuk tujuan penelitian. Dibangun dengan elemen – elemen dari jupyter, google colab menyediakan hampir semua library yang diperlukan untuk penelitian, termasuk *data mining*.

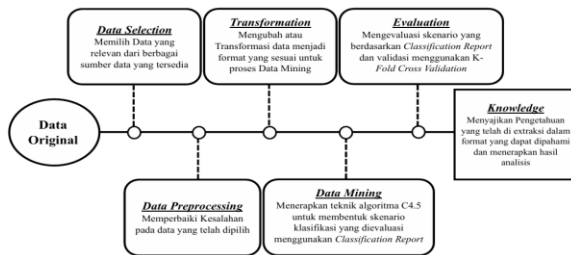
2.11. K-Fold Cross Validation

K-Fold Cross Validation digunakan dalam mengevaluasi rata-rata keberhasilan suatu sistem yang melibatkan pengulangan pengujian dengan mengacak atribut masukan, sehingga sistem dapat diuji dengan berbagai atribut untuk diinput secara acak [26]. Adapun pendapat lain yang disampaikan oleh Apriyani & Kurniati [27]a bahwa K-Fold Cross Validation adalah teknik validasi pengujian untuk mengevaluasi kinerja proses algoritma dan digunakan untuk mengurangi waktu proses klasifikasi dengan menjaga nilai akurasi dengan cara mengacak dan mengumpulkan data sebanyak K-Fold Cross Validation agar data uji tidak overlapping pada data testing.

3. METODE PENELITIAN

Penelitian ini melakukan pengklasifikasian dengan pendekatan *data mining* terhadap siswa SMK PGRI Cikampek sebagai penerima beasiswa. Objek yang akan diteliti yaitu data demografis siswa yang mencakup kelas X berjumlah 120 siswa, kelas XI berjumlah 253 siswa, dan kelas XII berjumlah 137 siswa pada rentang waktu tahun ajaran 2022/2023. Metodologi yang digunakan pada penelitian kali ini yaitu *Knowledge discovery in database* (KDD) dengan menggunakan algoritma C4.5 yang telah terstruktur dalam 5 tahapan yaitu yaitu *Data Selection*, *Data*

Preprocessing, Transformation, Data mining, dan Evaluation.



Gambar 3. Algoritma C4.5

Berikut Penjelasan Lengkap tentang tahapan penelitian yang akan dilakukan:

3.1. Data Selection

Tahap awal yaitu dilakukan pemilihan data yang menjadi target dengan menitik beratkan pada sample data. Data yang digunakan sebagai bahan penelitian bersumber dari staff tata usaha (TU) atau bagian administrasi SMK PGRI Cikampek.

3.2. Data Preprocessing

Tahap kedua yaitu dilakukannya data preprocessing dengan menggunakan bahasa pemrograman python untuk menghapuskan beberapa point data yang tidak relevan, karena pada umumnya data yang didapatkan penulis dari sumber terkait belum bisa dikatakan sempurna.

3.3. Data Transformation

Proses ini juga mencakup normalisasi atau standarisasi, standarisasi dilakukan terhadap seluruh atribut sehingga pada setiap atribut mempunyai kontribusi yang proporsional untuk hasil akhir pada proses data mining.

3.4. Data mining

Langkah berikutnya yaitu data mining yang merupakan tahapan pencarian pola atau pengalihan informasi dari data yang dipilih menggunakan algoritma C4.5. Proses data mining ini menggunakan bahasa pemrograman python memanfaatkan tahap Google Colaboratory. Terdapat tiga skenario yang akan dijalankan untuk memperoleh hasil terbaik.

Dataset dibagi menjadi set data training dan data testing dengan perbandingan tertentu, dimana Skenario A memakai perbandingan 90% data training dan 10% data testing, Skenario B memakai perbandingan 80% data training dan 20% data testing, dan Skenario C memakai perbandingan 70% data training dan 30% data testing.

Penetapan tiga skenario dapat memangkas waktu komputasi, sehingga semakin bervariasi skenario yang diteliti, maka semakin lama waktu yang dibutuhkan dalam menguji setiap modelnya.

3.5. Evaluation

Pada tahap ini, dilakukan evaluasi terhadap proses data mining dengan menganalisis hasil classification report dari masing-masing skenario. Evaluasi ini melibatkan perbandingan tingkat akurasi untuk menentukan skenario yang memiliki performa tertinggi. Skenario dengan tingkat akurasi tertinggi selanjutnya divalidasi menggunakan K-Fold cross validation.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian yang telah dilakukan adalah merancang sebuah pola penyeleksian terhadap siswa di SMK PGRI Cikampek sebagai penerima beasiswa Karawang Cerdas menggunakan analisis data mining khususnya teknik klasifikasi menggunakan pendekatan algoritma C4.5. Objek yang akan diteliti adalah data demografis siswa aktif kelas X sebanyak 120 siswa, kelas XI sebanyak 253 siswa, dan kelas XII sebanyak 137 siswa pada tahun ajaran 2022/2023. Metode yang digunakan ialah pendekatan Knowledge discovery in database (KDD).

4.1. Data Selection

Data yang digunakan berasal dari staf tata usaha (TU) atau bagian administrasi SMK PGRI Cikampek. Dalam proses pemilihan ini, hanya 19 atribut yang dipertimbangkan. Atribut terpilih dapat ditunjukkan pada Tabel 1.

Tabel 1. Atribut Terpilih

No.	Atribut	Keterangan
1.	Tempat Lahir	Atribut Utama
2.	Alat Transportasi	Atribut Utama
3.	Pekerjaan Orang Tua / Wali	Atribut Utama
4.	Penghasilan per Golongan	Atribut Utama
5.	Penerima KIP	Atribut Utama
6.	Jumlah Tanggungan	Atribut Utama
7.	Status Kepemilikan Rumah	Atribut Utama
8.	Prestasi	Atribut Utama
9.	PNS/TNI/POLRI	Atribut Utama
10.	Status	Kelas Target

4.2. Data Preprocessing

Tahapan selanjutnya yaitu Data Preprocessing sering juga disebut data cleaning atau pembersihan data.

a. Tipe Dataset

Tipe dataset yang tepat penting dilakukan karena metode analisis dan pemrosesan data yang digunakan berbeda tergantung dari tipe data yang disajikan.

b. Missing Value

Pada data yang digunakan tidak terdapat missing value, maka dari itu tidak ada penanganan dan melanjutkan ke proses selanjutnya.

c. Data Duplicate

Data duplicate digunakan untuk memastikan kualitas data yang diperlukan integrasi data tidak berisi duplikasi data.

4.3. Data Transformation

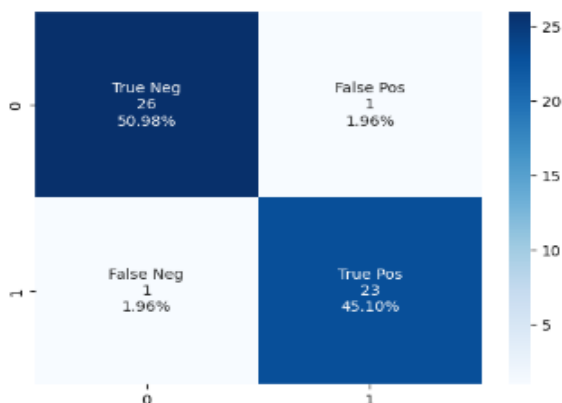
Tahap *Data Transformation* adalah proses merubah data kompleks menjadi sederhana agar mudah dalam proses pengelolaan. Pada tahap sebelumnya, terdapat data kategorikal yang perlu diubah kebentuk numerik untuk memudahkan proses selanjutnya.

4.4. Data mining

Pada tahap *data mining*, strategi pengujian menggunakan *dataset* yang telah disiapkan sebelumnya. *Dataset* tersebut terdiri dari 510 entri dengan 10 kolom, yang dibagi menjadi set pelatihan dan pengujian dengan rasio yang berbeda dalam skenario A, B, dan C. Pemilihan tiga skenario ini bertujuan untuk mengoptimalkan waktu komputasi. Proses pembagian *dataset* menggunakan fungsi `train_test_split` dari library `sklearn` untuk memastikan efisiensi waktu dalam melatih dan menguji model.

4.5. Skenario A

Pada Skenario A dibentuk dengan menggunakan data training sebesar 90% dan data testing sebesar 10% dari total data yang tersedia. Pengujian menggunakan *confusion matrix* dalam skenario ini menghasilkan prediksi angka untuk setiap kelas, baik kelas positif (P) maupun kelas negatif (N). Hasil pengujian terlihat pada Gambar 4.3, di mana dari total 51 entri data yang merupakan bagian dari data testing, angka prediksi berhasil didapatkan.



Gambar 5. *Confusion matrix* pada Skenario A

Berdasarkan gambar 5 *Confusion matrix* pada Skenario A, dapat dijelaskan sebagai berikut:

- True Negative (TN):** 26, menunjukkan bahwa terdapat 26 data yang sebenarnya termasuk dalam kelas negatif dan berhasil diprediksi dengan benar sebagai kelas negatif.
- False Positive (FP):** 1, artinya hanya ada 1 data yang seharusnya kelas negatif namun salah diprediksi sebagai kelas positif.
- False Negative (FN):** 1, yaitu terdapat 1 data yang sebenarnya kelas positif namun keliru diprediksi sebagai kelas negatif.

- True Positive (TP):** 23, menandakan bahwa ada 23 data yang benar-benar kelas positif dan berhasil diprediksi sebagai kelas positif.

Berdasarkan *Confusion matrix* tersebut, dapat dihitung beberapa matrik evaluasi performa skenario klasifikasi, sebagaimana pada Gambar 6.

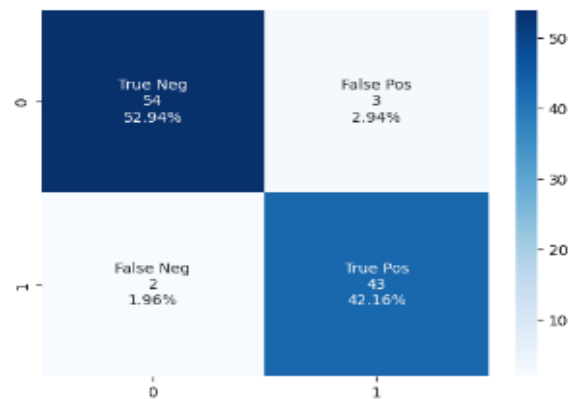
Classification Report 1:

	precision	recall	f1-score	support
0	0.96	0.96	0.96	24
1	0.96	0.96	0.96	27
accuracy			0.96	51
macro avg	0.96	0.96	0.96	51
weighted avg	0.96	0.96	0.96	51

Gambar 6. *Classification Report* Skenario A

4.6. Skenario B

Pengujian menggunakan *Confusion matrix* pada skenario ini menghasilkan prediksi untuk setiap kelas, baik kelas positif (P) maupun kelas negatif (N). Dari total 102 entri data yang digunakan sebagai data testing, hasil prediksinya terdokumentasi seperti yang terlihat pada Gambar 7.



Gambar 7. *Confusion matrix* Skenario B

Berdasarkan Gambar 7 *Confusion matrix* Skenario B, dapat dijelaskan sebagai berikut:

- True Negative (TN):** 54, menunjukkan bahwa terdapat 54 data yang sebenarnya termasuk dalam kelas negatif dan berhasil diprediksi dengan benar sebagai kelas negatif.
- False Positive (FP):** 3, artinya hanya ada 3 data yang seharusnya kelas negatif namun salah diprediksi sebagai kelas positif.
- False Negative (FN):** 2, yaitu terdapat 2 data yang sebenarnya kelas positif namun keliru diprediksi sebagai kelas negatif.
- True Positive (TP):** 43, menandakan bahwa ada 43 data yang benar-benar kelas positif dan berhasil diprediksi sebagai kelas positif.

Berdasarkan *confussion matrix* tersebut, dapat dihitung beberapa matrik evaluasi performa model klasifikasi, sebagaimana pada Gambar 8.

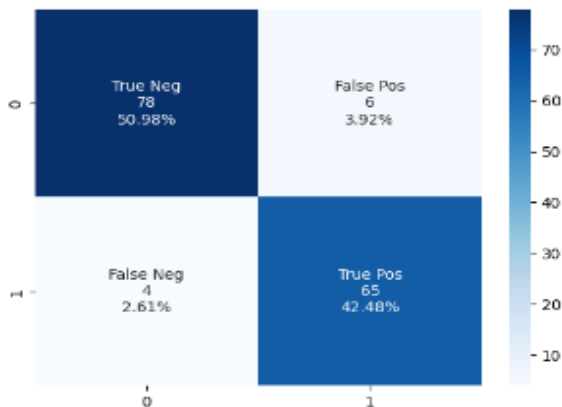
Classification Report 2:

	precision	recall	f1-score	support
0	0.93	0.96	0.95	45
1	0.96	0.95	0.96	57
accuracy			0.95	102
macro avg	0.95	0.95	0.95	102
weighted avg	0.95	0.95	0.95	102

Gambar 8. *Clasification Report* Skenario B

4.7. Skenario C

Pengujian menggunakan *Confusion matrix* pada skenario ini menghasilkan prediksi untuk setiap kelas, baik kelas Positif (P) maupun kelas Negatif (N). Dari total 153 entri data yang digunakan sebagai data testing, hasil prediksinya terdokumentasi seperti yang terlihat pada Gambar 9.

Gambar 9. *Confusion matrix* pada Skenario C

Berdasarkan Gambar 9 *Confusion matrix* pada Skenario C, dapat dijelaskan sebagai berikut:

- True Negative* (TN): 78, menunjukkan bahwa terdapat 78 data yang sebenarnya termasuk dalam kelas negatif dan berhasil diprediksi dengan benar sebagai kelas negatif.
- False Positive* (FP): 6, artinya hanya ada 6 data yang seharusnya kelas negatif namun salah diprediksi sebagai kelas positif.
- False Negative* (FN): 4, yaitu terdapat 4 data yang sebenarnya kelas positif namun keliru diprediksi sebagai kelas negatif.
- True Positive* (TP): 65, menandakan bahwa ada 63 data yang benar-benar kelas positif dan berhasil diprediksi sebagai kelas positif.

Berdasarkan *Confusion matrix* tersebut, dapat dihitung beberapa matrik evaluasi performa skenario klasifikasi, sebagaimana pada Gambar 10.

Classification Report 3:

	precision	recall	f1-score	support
0	0.92	0.94	0.93	69
1	0.95	0.93	0.94	84
accuracy			0.93	153
macro avg	0.93	0.94	0.93	153
weighted avg	0.94	0.93	0.93	153

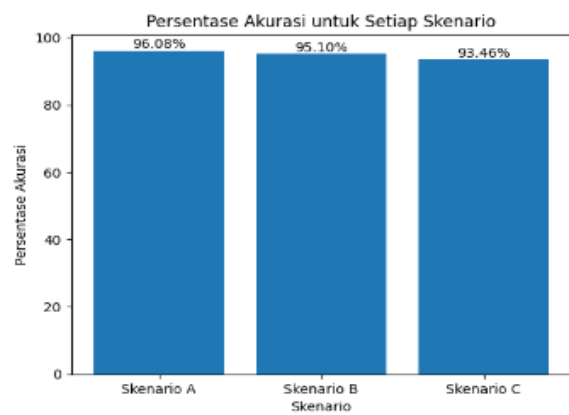
Gambar 10. *Clasification Report* pada Skenario C

4.8. Evaluation

Proses evaluasi bertujuan untuk mengevaluasi performa dan kualitas skenario dalam memprediksi atau mengklasifikasikan data. Dengan membandingkan hasil akurasi dari berbagai skenario yang dibangun, dapat diidentifikasi skenario mana yang memiliki performa yang lebih baik dan sesuai dengan tujuan atau kebutuhan penelitian. Penilaian skenario dilakukan dengan membandingkan hasil prediksi skenario dengan nilai sebenarnya dari data uji atau data validasi.

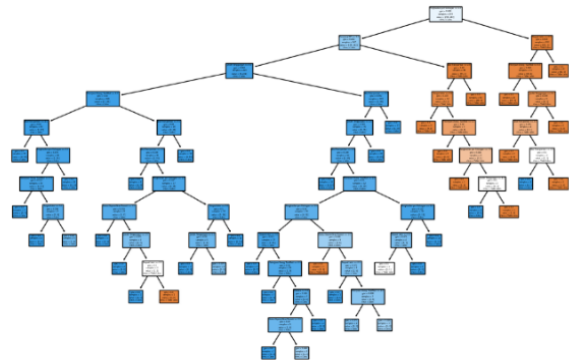
4.9. Akurasi Skenario

Berdasarkan hasil penilaian skenario A, skenario B dan skenario C yang memiliki performa yang baik persentase diatas 90% seperti pada gambar 11.



Gambar 11. Persentase Akurasi untuk Setiap Skenario

Dari gambar 11 diketahui nilai skenario dengan tingkat akurasi tertinggi yaitu 96.08% pada skenario A, dengan demikian decision tree dibuat menggunakan skenario A. Decision Tree dapat dilihat pada gambar 12.



Gambar 12. Decision Tree Skenario A

4.10. K-Fold Cross Validation

Untuk mengecek hasil Performa skenario A yang terbaik sebelumnya, maka akan dilakukan evaluasi melalui Fold Cross Validation, seperti yang ditunjukkan dalam Tabel 4.

Tabel 4. Hasil K-Fold Cross Validation

Fold ke -	Score	Persentasi
1	0.9411764705882353	94.12%
2	0.9411764705882353	94.12%
3	0.9607843137254902	96.08%
4	0.9411764705882353	94.12%
5	0.8627450980392157	86.27%
6	0.9215686274509803	92.16%
7	0.9607843137254902	96.08%
8	0.9411764705882353	94.12%
9	0.9607843137254902	96.08%
10	1.0	100.00%
Rata-Rata Score	0.9431372549019608	94.31%

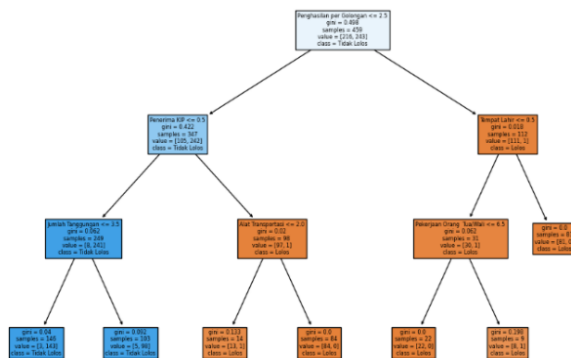
Hasil rata-rata 10 - Fold Cross Validation diperoleh sebesar 94.31% menunjukkan bahwa skenario memiliki performa yang baik dalam memprediksi nilai target.

4.11. Pembahasan

Dengan akurasi tertinggi sebesar 96%, skenario A menjadi pilihan utama untuk membentuk pohon keputusan dalam analisis ini. Pohon keputusan ini memiliki tingkat kompleksitas yang tinggi, menyulitkan perumusan aturan (rules). Untuk mengatasi hal ini, diperlukan proses pruning atau pemangkasan yang bertujuan untuk menyederhanakan struktur pohon keputusan yang kompleks sehingga interpretasinya menjadi lebih mudah.

4.12. Pruning

Penelitian ini menggunakan teknik post-pruning dengan mengatur nilai max_depth pada pohon keputusan sebagai strategi untuk mengontrol kedalaman maksimum dari pohon tersebut. Dalam konteks ini, max_depth memungkinkan pembentukan pohon yang lebih sederhana atau lebih kompleks tergantung pada nilai yang dipilih. Dengan memilih max_depth 3, penelitian ini mencapai titik tengah yang memungkinkan pembentukan pohon yang cukup kompleks untuk memodelkan data dengan efisien tanpa menjadi terlalu rumit. Hasil dari penerapan teknik max depth tergambar dalam Gambar 13.



Gambar 13. Pohon Keputusan Setelah Proses Pruning

Gambar 13 menampilkan pohon keputusan setelah proses pruning dengan Dengan menetapkan max_depth pada nilai 3, penelitian ini mencapai keseimbangan yang sesuai pada skenario A Decision Tree sehingga lebih efisien dan tidak terlalu kompleks. Pohon keputusan yang disajikan telah disederhanakan, hanya menampilkan cabang dengan kedalaman maksimum 3. Dari pohon keputusan yang telah disederhanakan tersebut, dapat dihasilkan beberapa interpretasi dari pohon keputusan sebagai berikut:

a. Aturan pertama mengenai penghasilan per golongan kurang dari atau sama dengan Rp.2,999,999:

- Jika penghasilan per golongan seseorang tidak melebihi Rp.2,999,999, status kelayakan mereka akan dipengaruhi oleh status penerimaan KIP. Jika mereka bukan penerima KIP, status kelayakan ditentukan berdasarkan jumlah tanggungan yang mereka miliki. Jumlah tanggungan yang kurang dari atau sama dengan 3.5 akan mengarahkan mereka pada status "Tidak Lolos", sementara jumlah tanggungan lebih dari 3.5 status mereka pun akan tetap dinyatakan "Tidak Lolos". Dari pernyataan berikut dapat diartikan bahwa atribut pada Jumlah Tanggungan, tidak begitu berpengaruh pada proses penyeleksian penerima beasiswa.

- Jika seseorang adalah penerima KIP, status kelayakan mereka tidak akan dipengaruhi oleh jumlah tanggungan, tetapi ditentukan oleh jenis alat transportasi yang mereka gunakan. Penggunaan sepeda motor atau berjalan kaki akan membuat status mereka menjadi "Lolos".

b. Aturan kedua mengenai penghasilan per golongan lebih dari Rp.2,999,999:

- Jika penghasilan per golongan seseorang melebihi Rp.2,999,999, status kelayakan mereka akan bergantung pada tempat lahir mereka. Jika tempat lahirnya berada di luar Karawang, status kelayakannya akan ditentukan oleh pekerjaan orang tua atau wali mereka. Jika orang tua atau wali mereka tidak bekerja atau bekerja sebagai buruh, mereka dianggap "Lolos". Namun, jika orang tua atau wali mereka bekerja sebagai wiraswasta, karyawan swasta, petani, PNS/TNI/POLRI, wirausaha, atau pensiunan, status mereka tetap "Lolos".

- Jika seseorang lahir di Karawang, status kelayakannya secara otomatis dianggap "Lolos" tanpa mempertimbangkan faktor-faktor lain.

Dengan demikian, penjelasan di atas menggambarkan bagaimana kedalaman dalam setiap aturan dan keputusan yang diambil dalam menentukan status kelayakan berdasarkan kondisi yang diberikan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan algoritma C4.5 untuk mengklasifikasikan siswa penerima beasiswa di SMK PGRI Cikampek. Menggunakan pendekatan Knowledge Discovery and Data mining (KDD), data demografis siswa kelas X, XI, dan XII diolah untuk membangun model klasifikasi yang optimal. Evaluasi menggunakan *confusion matrix* dan *classification report* menunjukkan bahwa tiga skenario pengujian menghasilkan akurasi tinggi antara 93% hingga 96%, dengan skenario A mencapai akurasi tertinggi 96%. Pohon keputusan yang dihasilkan memungkinkan perumusan aturan untuk menentukan kelayakan siswa sebagai penerima beasiswa berdasarkan atribut seperti penghasilan, status penerima KIP, jumlah tanggungan, pekerjaan orang tua, dan tempat lahir.

DAFTAR PUSTAKA

- [1] M. M. Zamzami and D. Kurniansyah, "Good Governance dalam Program Beasiswa Karawang Cerdas Kabupaten Karawang," *J. Ilm. Wahana Pendidik*, <https://jurnal.unibrah.ac.id/index.php/JIWP>, vol. 8, no. 3, pp. 178–183, 2022, doi: 10.5281/zenodo.5832067.
- [2] A. S. Suweleh, D. Susilowati, and Hairani, "Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5," *J. BITE*, vol. 2, no. 1, pp. 12–21, 2020, doi: 10.30812/bite.v2i1.798.
- [3] I. Sudirman, "Analisis Proses Pembentukan Kebijakan Publik: Studi Kasus Program Beasiswa Karawang Cerdas Tahun 2020," *Indones. J. Soc. Polit. Sci.*, vol. 2, no. 1, pp. 1–12, 2021, [Online]. Available: <http://journal.epistemikpress.id/index.php/Epistemik/article/download/44/35>
- [4] B. D. F. E. Yulyana, and L. A. Aryani, "Evaluasi Program Karawang Cerdas Oleh Dinas Pendidikan Pemuda Dan Olahraga Kabupaten Karawang," *J. Pemerintah. dan Polit.*, vol. 7, no. 2, pp. 1–7, 2022, doi: 10.36982/jpg.v7i2.2100.
- [5] P. R. Adinda, "Pohon Keputusan C4. 5 Algoritma Untuk Klasifikasi Program Bantuan Belajar," *J. Portal Data*, vol. 2, no. 9, pp. 1–13, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/215%0Ahttp://portaldata.org/index.php/portaldata/article/download/215/204>
- [6] M. Sari, A. Perdana Windarto, and H. Okprana, "Penerapan Data mining Klasifikasi C4.5 Pada Penerima Beasiswa di SMK Swasta Anak Bangsa," *BEES Bull. Electr. Electron. Eng.*, vol. 1, no. 3, pp. 115–121, 2021.
- [7] Sabrina Fitri Jasmine, "Pengaruh Beasiswa KIP-K Terhadap Prestasi Belajar Mahasiswa Manajemen Pendidikan Angkatan 2021 Universitas Negeri Surabaya," *J. Pendidikan, Bhs. dan Budaya*, vol. 2, no. 2, pp. 61–70, 2023, doi: 10.55606/jpbb.v2i2.1437.
- [8] M. Mastur, A. N. Dinda, P. Wulandari, and ..., "Pengaruh Pemberian Beasiswa Terhadap Motivasi Belajar Mahasiswa Universitas Darunnajah," *J. Pendidik. ...*, vol. 7, pp. 6647–6652, 2023, [Online]. Available: <https://mail.jptam.org/index.php/jptam/article/view/7274%0Ahttps://mail.jptam.org/index.php/jptam/article/download/7274/6016>
- [9] N. A. Manihuruk, M. Zarlis, E. Irawan, and H. S. Tambunan, "Penerapan Data mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 4, no. 1, pp. 29–34, 2020, doi: 10.30865/komik.v4i1.2575.
- [10] F. Nuraeni, D. Kurniadi, and G. Fauzian Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma K-Means++," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 437–443, 2023, doi: 10.32736/sisfokom.v11i3.1439.
- [11] N. Rahmadani and Risnawati, "Implementasi Metode K-Means Clustering Tunggalan Uang Kuliah Pada Stmik Royal Kisaran," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 118–124, 2022, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [12] A. Sumiah and N. Mirantika, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan," *Buffer Inform.*, vol. 6, no. 1, pp. 1–10, 2020.
- [13] A. Azahari and N. Nursobah, "Rekomendasi Penerimaan Beasiswa Yayasan Untuk Siswa Baru SMK TI Airlangga dengan Algoritma C4.5," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 609, 2021, doi: 10.30865/mib.v5i2.2943.
- [14] A. Afriansyah, R. Annisa, and Z. R. Mair, "Sistem Pakar Deteksi Kerusakan (Troubleshooting) Pada Smartphone Berbasis Mobile," *Teknol. Inf. dan Komput.*, vol. 9, no. 1, pp. 17–25, 2019, [Online]. Available: <https://jurnal.polsky.ac.id/index.php/index/article/view/186>
- [15] J. C. Mestika, M. O. Selan, and M. I. Qadafi, "Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python," *BIKMA Bul. Ilm. Ilmu Komput. dan Multimed.*, vol. 99, no. 99, pp. 216–219, 2022.
- [16] H. Abijono, P. Santoso, and N. L. Anggreini, "Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data," *J. Teknol. Terap. G-Tech*, vol. 4, no. 2, pp. 315–318, 2021, doi: 10.33379/gtech.v4i2.635.
- [17] 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, doi: 10.31253/algor.v3i2.1019.
- [18] Anggi Trifani, Agus Perdana Windarto, and Hendry Qurniawan, “Penerapan *Data mining* Klasifikasi C4.5 dalam Menentukan Tingkat Stres Mahasiswa Akhir,” *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 2, pp. 91–105, 2022, doi: 10.55606/jurritek.v1i2.414.
- [19] Z. Saputra, D. Sartika, and M. Haviz Irfani, “RESOLUSI : Rekayasa Teknik Informatika dan Informasi Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree,” *Jurnal Rekayasa Teknik Informatika dan Informasi*, vol. 43, no. 3, pp. 231–240, 2024. [Online]. Available: <https://docs.python.org/3.13/tutorial/index.html>
- [20] B. Baskoro, S. Sriyanto, and L. S. Rini, “Prediksi Penerima Beasiswa dengan Menggunakan Teknik *Data mining* di Universitas Muhammadiyah Pringsewu,” *Pros. Semin. Nas. Darmajaya*, vol. 1, no. 0, pp. 87–94, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2918>
- [21] M. Riyyan and H. Firdaus, “PERBANDINGAN ALGORITME NAÏVE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar),” *Jurnal Informatika dan Rekayasa Elektronik*, vol. 5, no. 1, pp. 1–10, 2022. doi: 10.36595/jire.v5i1.547.
- [22] N. I. Nella, N. Y. Setiawan, and D. E. Dian, “Klasifikasi Penerima Bantuan Program Keluarga Harapan menggunakan Algoritme Decision Tree C4. 5 (Studi Kasus: Desa Mlirip Kabupaten Mojokerto),” ... *Teknologi Informasi dan ...*, vol. 6, no. 3, pp. 1332–1339, 2022. [Online]. Available: <https://jptiik.multi.web.id/index.php/jptiik/article/view/10801%0Ahttps://jptiik.multi.web.id/index.php/jptiik/article/download/10801/4769>
- [23] R. A. Saputra, S. Wasiyanti, and D. Pribadi, “Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (Bpnt),” *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 4, no. 1, p. 25, 2021. doi: 10.21927/ijubi.v4i1.1757.
- [24] M. K. Mulyadien and U. Enri, “Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT),” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 12, pp. 198–210, 2022, [Online]. Available: <http://jurnal.peneliti.net/index.php/JIWP/article/view/2025%0Ahttp://jurnal.peneliti.net/index.php/JIWP/article/download/2025/1535>
- [25] A. T. Zy, A. T. Sasongko, and A. Z. Kamalia, “Penerapan Naïve Bayes Classifier, Support Vector *Machine*, dan Decision Tree untuk Meningkatkan Deteksi Ancaman Keamanan Jaringan,” *Media Online*, vol. 4, no. 1, pp. 610–617, 2023.
- [26] M. Faridj and A. Informatika, “Klasifikasi Lama Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *Teknologipintar.org*, vol. 2, no. 12, pp. 1–17, 2023, 2022.
- [27] H. Apriyani and K. Kurniati, “Perbandingan Metode Naïve Bayes Dan Support Vector *Machine* Dalam Klasifikasi Penyakit Diabetes Melitus,” *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issue3.year2020.page133-143.