



# Komparasi Algoritme K-NN, Naïve Bayes, dan Cart untuk Memprediksi Penerima Beasiswa

Ali Nur Ikhsan<sup>1\*</sup>, Pungkas Subarkah<sup>2</sup>, Raditya Sani Alifian<sup>3</sup>



<sup>1,2,3</sup> Informatika, Universitas Amikom Purwokerto, Purwokerto, Indonesia

## ARTICLE INFO

### Article history:

Received August 26, 2022

Revised September 11, 2022

Accepted May 13, 2023

Available online July 25, 2023

### Kata Kunci:

Beasiswa, Algoritme, K-NN, Naïve Bayes, CART

### Keywords:

Scholarships, Algorithms, K-NN, Naïve Bayes, CART



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2023 by Author. Published by Universitas Pendidikan Ganesha.

## ABSTRAK

Persebaran penerima beasiswa di tanah air Indonesia terdapat masalah salah satunya yaitu tidak tepat sasaran. Pemerintah Indonesia memberikan beasiswa kepada peserta didik di Indonesia sebagai contoh yaitu Program Indonesia Pintar dan, Program Indonesia Pintar Pendidikan Dasar dan Pendidikan menengah. Pemberian beasiswa diperlukan adanya klasifikasi dalam mengambil keputusan penerima beasiswa tersebut untuk meminimalisir salah sasaran. Prediksi secara dini harus dilakukan untuk mengantisipasi kesalahan dalam penerima bantuan beasiswa, salah satunya menggunakan teknik data mining. Tujuan penelitian ini untuk menganalisis Komparasi Algoritme K-NN, Naïve Bayes, Dan CART untuk Memprediksi Penerima Beasiswa bagi pengelola di SMA. Penelitian yang dilakukan menggunakan data mining terhadap dataset penerima beasiswa dengan memanfaatkan aplikasi Weka dalam mengolah data. Dataset yang digunakan dalam penelitian ini yaitu data penerima beasiswa di salah satu SMA dengan jumlah dataset yaitu 948 data dan memiliki 6 atribut (5 atribut dan 1 target atribut). Metode yang digunakan dalam penelitian ini yaitu Confusion matrix dan K-fold 10 Cross Validation. Komparasi Algoritme K-NN, Naïve Bayes, Dan CART untuk Memprediksi Penerima Beasiswa. Dari ketiga Algoritme yang digunakan dalam penelitian diperoleh kesimpulan Algoritme CART merupakan Algoritme dengan hasil akurasi yang paling tinggi sebesar 91.3502% untuk memprediksi penerima beasiswa dengan kategori Good Classification.

## ABSTRACT

The distribution of scholarships recipients in Indonesia has problems, one of which is not on target. The Indonesian government provides scholarships to students in Indonesia, for example the Smart Indonesia Program and the Smart Indonesia Program for Primary and Secondary Education. In granting scholarships, it is necessary to classify scholarship recipients to minimize mistargeting. Initial predictions must be made to anticipate mistakes in scholarship recipients, one of which is by using data mining techniques. The research was conducted using data mining on the dataset of scholarship recipients by utilizing the Weka application in processing data. The dataset used in this study is data on scholarship recipients in one of the high schools in Banjarnegara Regency, Central Java with a total dataset of 948 data and has 6 attributes (5 attributes and 1 target attribute). The method used in this research is Confusion matrix and K-fold 10 Cross Validation. Researchers compared the K-NN, Naïve Bayes, and CART algorithms to predict scholarship recipients. Of the three algorithms used in the study, it was concluded that the CART algorithm was the algorithm with the highest accuracy of 91.3502% for predicting scholarship recipients in the Good Classification category.

## 1. PENDAHULUAN

Dampak pandemi covid 19 sangat berpengaruh terhadap kehidupan masyarakat salah satunya di bidang Pendidikan. Mendapatkan dan melanjutkan pendidikan yang layak dibutuhkan oleh setiap masyarakat di Indonesia (Arianto, 2021; Fitria, Linda., Neviyarni., Netrawati., & Karneli, 2020). Ada beberapa cara mendapatkan pendidikan yang baik salah satunya dengan memanfaatkan beasiswa yang berasal dari pemerintah (Alwi & Munirah, 2022; Kidd et al., 2020; Straub & Vilsmaier, 2020). Berdasarkan hasil observasi data dari pengelola Beasiswa Peningkatan Prestasi Akademik (PPA) dan Kartu Indonesia Pintar (KIP-K) di SMA yang berada di Kabupaten Banjarnegara, Pendaftar beasiswa PPA dan KIP-K setiap tahunnya mengalami peningkatan, namun dengan penyeleksiannya masih manual dengan bantuan pengolahan aplikasi Microsoft Excel. Hal ini dapat menyebabkan proses penyeleksian manual kurang efektif dan tepat sasaran dalam segi mencapai tujuan maupun dalam segi waktu rekap penyeleksian. Hal ini tentunya membutuhkan prediksi dini mengenai penerima beasiswa. Prediksi secara dini harus dilakukan, salah satunya menggunakan teknik data mining, yang bertujuan untuk menggambarkan penemuan baru pada prediksi mengenai beasiswa sekolah. Dimana strategi penambangan informasi menggambarkan siklus yang menggunakan prosedur Artificial Intelligence (AI) (Pungkas Subarkah et al.,

\*Corresponding author.

E-mail addresses: [alinurikhsan@amikompurwokerto.ac.id](mailto:alinurikhsan@amikompurwokerto.ac.id) (Ali Nur Ikhsan)

2021). Prediksi ini bertujuan guna mengetahui nilai akurasi dari hasil yang didapatkan (Coussement et al., 2020; Pallathadka et al., 2021). Sekolah Menengah Atas (SMA) di Kabupaten Banjarnegara, tiap tahun akademik mendapatkan bantuan Beasiswa, baik Beasiswa yang diperoleh dari lembaga pemerintah, perusahaan, ataupun yayasan, dengan bentuk berupa bantuan biaya pendidikan bagi mahasiswa kurang mampu secara ekonomi tetapi memiliki potensi akademik. Namun jumlah alokasi Beasiswa yang disalurkan setiap tahunnya terbatas dan tidak sebanding dengan jumlah mahasiswa yang mengajukan Beasiswa, sehingga tersebut mengakibatkan harus adanya proses penyeleksian yang ketat oleh bagian pengelola beasiswa, agar penyaluran Beasiswasesuai dan tepat sasaran. Pemberian beasiswa mempunyai tujuan meringankan biaya sekolah siswa, meningkatkan optimisme siswa untuk bersekolah lagi. Mendorong aagr siswa semangat belajar agar bisa meraih cita-cita yang diinginkan (Laajaj et al., 2022; Mittal, P., & Gill, 2014). *Machine Learning* (ML) adalah salah satu metodologi cerdas yang telah menunjukkan hasil yang menjanjikan dalam domain klasifikasi dan prediksi. Salah satu tugas ML secara umum melibatkan memprediksi variabel target dalam data yang sebelumnya tidak terlihat, adalah klasifikasi (Darmayanti et al., 2021). Tujuan klasifikasi adalah untuk memprediksi variabel target (kelas) dengan membangun model klasifikasi berdasarkan pada dataset pelatihan, dan kemudian menggunakan model itu untuk memprediksi nilai kelas data uji (Bunker & Thabtah, 2019).

Algoritme K-Nearest Neighbor (K-NN) merupakan salah satu metode klasifikasi dalam data mining yang termasuk ke dalam *supervised learning* (Lu et al., 2021; Okfalisa et al., 2018; Wang & Mao, 2020). Pengklasifikasian yang dilakukan berdasarkan atribut dan data training, sehingga proses pengklasifikasian data baru dilakukan berdasarkan perbandingan kemiripan mayoritas pada data training. K-NN metode nonparametrik untuk mengklasifikasikan data baru yang kelasnya belum diketahui dan memilih data sebanyak k yang letaknya paling dekat dari data baru (Nurdiawan et al., 2021). Dalam kNN nilai jarak ditentukan dengan pengujian data testing terhadap data training kemudian menggunakan nilai terkecil dari nilai ketetanggaan terdekat (Lu et al., 2021; Permana et al., 2021). K-NN termasuk kelompok *instance based learning*. Algoritma ini juga merupakan salah satu teknik lazy learning. K-NN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data *testing* (Bermejo-Martín, Rodríguez-Monroy, Núñez-Guerrero, et al., 2021; Mustakim & Oktaviani F, 2016; Nawaz et al., 2021; Wazery et al., 2021). Algoritme K-Nearest Neighbor (KNN) dikenal karena kemudahannya dalam melakukan prediksi, bekerja dengan cara mencari sejumlah k pola yang terdekat dengan pola masukan, kemudian menentukan kelas keputusan berdasarkan jumlah pola terbanyak diantara k pola tersebut (Gao et al., 2021; Sisodia & Sisodia, 2021). Sedangkan pada algoritme naive bayes adalah suatu algoritma yang digunakan untuk klasifikasi, teorema bayes dan berasumsi bahwa nilai antar variabel saling bebas (independen) pada suatu nilai *output* (Han et al., 2012; Idris, 2019). Algoritme naive bayes diterapkan pada data asli untuk mendapatkan data pelatihan baru dan berkualitas tinggi dengan tujuan meningkatkan akurasi deteksi dan mengurangi tingkat *noise* (Gu & Lu, 2021; Karo et al., 2022). Algoritme *Classification And Regression Trees* (CART) CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. Metode ini merupakan metode yang biasa diterapkan suatu himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan bine CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. Metode ini merupakan metode yang biasa diterapkan suatu himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan bine (Helena et al., 2019). Pada klasifikasi algoritme CART (Classification and Regression Tress), sebuah record akan diklasifikasikan ke dalam salah satu dari sekian klasifikasi yang tersedia pada variabel tujuan berdasarkan nilai-nilai variabel prediktornya (Susanto & Suryadi, 2010). Secara umum, algoritme baik digunakan untuk *dataset numeric* (Bermejo-Martín, Rodríguez-Monroy, & Núñez-Guerrero, 2021; P. Subarkah et al., 2018).

Beberapa penelitian terdahulu terkait rekomendasi penerima beasiswa menggunakan algoritma K-Nearest Neighbor (K-Nn) Dan Algoritma C4.5 dengan hasil algoritma k-NN memiliki tingkat akurasi yang lebih tinggi dibandingkan algoritma C4.5 (Noviana et al., 2019). Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up diperoleh hasil algoritma Decision Tree merupakan algoritma yang paling cocok untuk digunakan di antara algoritma kNN dan Naive Bayes (Permana et al., 2021). Klasifikasi Calon penerima bidikmisi dengan menggunakan algoritma K-Nearest Neighbor dengan hasil perhitungan dengan menggunakan algoritma K-NN, diperoleh hasil klasifikasi sudah sangat bagus (Zerlinda et al., 2019). Prediksi kelompok UKT mahasiswa pada program studi S1 Sistem Informasi FMIPA Universitas Riau menggunakan algoritme K-Nearest Neighbor. Hasil yang didapatkan pada penelitian ini yaitu nilai akurasi sebesar 84,21% (Sukamto et al., 2020). Kontribusi dari penelitian yang dilakukan yaitu Komparasi Algoritme K-NN, Naive Bayes, Dan CART Untuk Memprediksi

Penerima Beasiswa. Tujuan penelitian ini untuk menganalisis komparasi algoritme K-NN, Naïve Bayes, dan CART untuk memprediksi penerima beasiswa bagi pengelola di SMA Kabupaten Banjarnegara.

## 2. METODE

Jenis penelitian ini merupakan penelitian kualitatif. Tahap penelitian terdiri dari tahap identifikasi masalah, tahap pre processing, dan penarikan kesimpulan. Pada tahap identifikasi masalah tim mencari dan mengidentifikasi masalah yang akan dijadikan penelitian. Objek yang dijadikan penelitian yaitu terkait penerimaan beasiswa dan untuk subjek penelitian difokuskan dengan menggunakan data SMA yang bekerja sama dengan peneliti untuk diambil dataset penerima beasiswa. Pada penelitian ini, menggunakan teknik *data mining*, yaitu komparasi algoritme Naive Bayes, KNN dan CART. Dengan menggunakan 3 komparasi algoritme diharapkan dapat memperoleh hasil perbandingan yang paling baik untuk memprediksi penerima beasiswa. Penelitian ini dilakukan di salah satu SMA di Kabupaten Banjarnegara dengan *dataset* yang digunakan juga bersumber dari SMA tersebut. *Dataset* dalam penelitian ini terdiri dari Jenis Kelamin, Usia, Alat Transportasi, Penghasilan Orang tua, KPS, Layak KIP dan Penerima KIP (Kelas Target). Pada tahap pengumpulan data dilakukan dengan pengumpulan data berupa jurnal dan buku yang berkaitan dengan penelitian. *Dataset* yang digunakan pada penelitian diperoleh secara langsung dari pihak SMA. Tahap *pre-processing*, *dataset* yang diperoleh masih berupa data Excel, data tersebut diolah terlebih dahulu ke dalam bentuk file dengan jenis Arff yang dapat dibaca oleh aplikasi Weka. Penelitian melakukan konversi data dengan menyesuaikan data yang akan digunakan dalam aplikasi weka untuk dilakukan mining. Pada tahap ini, melakukan seleksi data yang akan digunakan sebagai data penelitian. Data diidentifikasi dan dikelompokkan sesuai dengan atributnya (*attribute identification and selection*), kemudian proses penindakan nilai atribut yang hilang atau tidak lengkap (*handling missing values*) dan proses diskritisasi nilai. Metode yang digunakan pada penelitian ini menggunakan Algoritme K-NN, Naïve Bayes, Dan CART. Pada tahap ini dilakukan penghitungan nilai akurasi melalui *confusion matrix*. Tahap *validasi dan evaluasi* dilakukan dengan tujuan mengetahui hasil yang diperoleh dari penelitian sudah akurat dan dapat dilanjutkan dengan penarikan kesimpulan. Penarikan kesimpulan dilakukan dengan melihat hasil perbandingan atau komparasi Algoritme K-NN, Naïve Bayes, dan CART yang telah dilakukan dengan *output* berupa nilai akurasi dari ketiga algoritme, jurnal nasional dan laporan penelitian.

## 3. HASIL DAN PEMBAHASAN

### Hasil

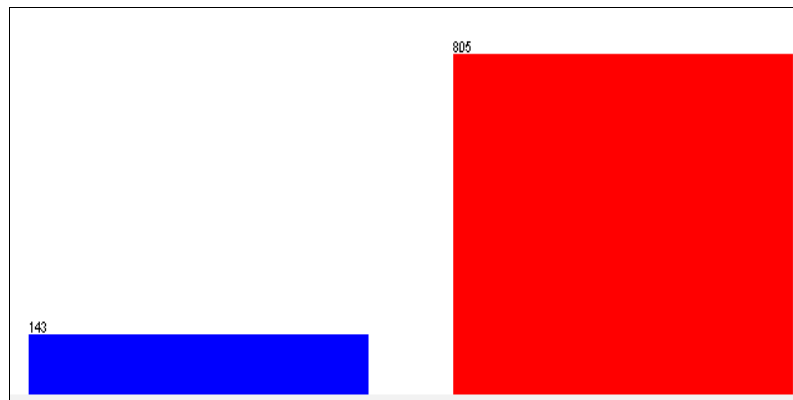
Penelitian ini menggunakan *dataset* yang berasal dari SMA di Kabupaten Banjarnegara dengan rincian 948 data dan memiliki 6 atribut (5 atribut dan 1 target atribut). Ya (siswa tersebut putus sekolah) dan Tidak (siswa tersebut tidak putus sekolah). Data terlebih dahulu diolah melalui tahapan *pre-processing*, bertujuan untuk menyesuaikan atribut-atribut yang akan diolah menggunakan aplikasi weka, serta untuk menemukan nilai yang kosong pada atribut (*missing values*). Atribut- atribut pada dataset terdiri dari atribut 0, Atribut pertama ialah jenis kelamin siswa. Atribut 1 usia siswa yang bersekolah. Atribut 2 Alat transportasi yang digunakan saat bersekolah. Atribut 3 jumlah penghasilan orang tua. Atribut 4 Kartu Perlindungan Sosial (KPS). Atribut 5 Layak KIP. Atribut 6 penerima KIP (*Class Target*). Selanjutnya ialah tahapan *pre-processing* yang berguna untuk penyeleksian dari dataset agar diperoleh data yang siap digunakan. *Dataset* yang telah dilakukan penyesuaian atribut untuk digunakan pada aplikasi weka disajikan pada [Tabel 1](#).

**Tabel 1.** Data *Pre-Processing*

Data Orisinil	Data <i>Pre-Processing</i>	Keterangan
P	P	Jenis Kelamin
15	15	Usia
5	5	Alat Transportasi
1	1	Penghasilan Orang tua
Tidak	Tidak	KPS
Ya	Ya	Layak KIP
Tidak	Tidak	Penerima KIP (Kelas Target)

Berdasarkan [Tabel 1](#) bahwa hasil data *pre-processing* dataset Penerima Beasiswa yang telah dilakukan tidak terdapat *missing value*. Tahap selanjutnya setelah dilakukan tahap *pre-processing* yaitu mengolah dataset dengan menggunakan aplikasi Weka. Tahap ini bertujuan untuk mengetahui hasil

*confusion Matrix* berdasarkan metode evaluasi *10-fold cross validation*. Dataset yang diolah dibagi menjadi 7 subsets dengan 6 subsets sebagai *training sets* dan 1 subsets sebagai *testing set* kemudian dilakukan perulangan sebanyak 10 kali. *Classifier* yang digunakan pada dataset yaitu IBk/K-NN, NAÏVE BAYES, dan CART. Visualisasi informasi dataset yang telah melalui tahap *pre-processing*.disajikan pada [Gambar 1](#).



**Gambar 2.** Visualisasi data keseluruhan

Berdasarkan [Gambar 2](#). dapat dilihat penerima beasiswa pada grafik warna biru yaitu dan bukan penerima beasiswa pada grafik warna merah. Pada algoritme KNN diperoleh nilai akurasi sebesar 90.0844% yang diperoleh dari nilai *precision*, *recall*, dan *F-Measure*. *Confusion* Kelas “0” merupakan penjabaran dari nilai *confusion matrix* yang dihasilkan dari *Classifier Output* KNN. Pada *Confusion* Kelas “0” diperoleh nilai *True Positive*/ TP = 85, *False Positive*/FP = 36, *False Negatif*/FN=58, dan *True Negative*/TN = 769. *Confusion* Kelas “1” merupakan penjabaran dari nilai *confusion matrix* yang dihasilkan dari *Classifier Output* KNN. Pada *Confusion* Kelas “1” diperoleh nilai *True Positive*/ TP = 769, *False Positive*/FP = 58, *False Negatif*/FN=36, dan *True Negative*/TN = 85. Dari perhitungan yang sudah dilakukan nilai akurasi berdasarkan *Confusion Matrix* Algoritme KNN dapat dilihat pada [Tabel 2](#).

**Tabel 2.** Nilai Akurasi Berdasarkan *Confusion Matrix* Algoritme KNN

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F-Measure</b>
0	0.702	0.594	0.644
1	0.930	0.955	0.942
<i>Weighted Avg</i>	0.896	0.901	0.897

Dengan menggunakan aplikasi Weka 3.8.6, pengujian dataset Penerima Beasiswa diperoleh hasil klasifikasi dengan waktu 0 seconds. Pada algoritme Naïve Bayes diperoleh nilai akurasi sebesar 90.8228% yang diperoleh dari nilai *precision*, *recall*, dan *F-Measure*. *Confusion* Kelas “0” merupakan penjabaran dari nilai *confusion matrix* yang dihasilkan dari *Classifier Output* Naïve Bayes. Pada *Confusion* Kelas “0” diperoleh nilai *True Positive*/ TP = 90, *False Positive*/FP = 34, *False Negatif*/FN=53, dan *True Negative*/TN = 771. *Confusion* Kelas “1” merupakan penjabaran dari nilai *confusion matrix* yang dihasilkan dari *Classifier Output* Naïve Bayes. Pada *Confusion* Kelas “1” diperoleh nilai *True Positive*/ TP = 771, *False Positive*/FP = 53, *False Negatif*/FN=34, dan *True Negative*/TN = 90. Dari perhitungan yang sudah dilakukan nilai akurasi berdasarkan *Confusion Matrix* Algoritme Naïve Bayes dapat dilihat pada [Tabel 3](#).

**Tabel 3.** Nilai Akurasi Berdasarkan *Confusion Matrix* Algoritme Naïve Bayes

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F-Measure</b>
0	0.726	0.629	0.674
1	0.936	0.958	0.947
<i>Weighted Avg</i>	0.904	0.908	0.905

Dengan menggunakan aplikasi Weka 3.8.6, pengujian dataset Penerima Beasiswa diperoleh hasil klasifikasi dengan waktu 0.03 seconds. Pada algoritme CART diperoleh nilai akurasi sebesar 91.3502% yang diperoleh dari nilai *precision*, *recall*, dan *F-Measure*.*Confusion* Kelas “0” merupakan penjabaran dari nilai *confusion matrix* yang dihasilkan dari *Classifier Output* CART. Pada *Confusion* Kelas “0” diperoleh nilai *True Positive*/ TP = 97, *False Positive*/FP = 36, *False Negatif*/FN=46, dan *True Negative*/TN = 76. *Confusion*

Kelas “1” merupakan penjabaran dari nilai *confusion matrix* yang dihasilkan dari *Classifier Output* CART. Pada *Confusion* Kelas “1” diperoleh nilai *True Positive/ TP* = 769, *False Positive/FP* = 46, *False Negatif/FN*=36, dan *True Negative/TN* = 97. Dari perhitungan yang sudah dilakukan nilai akurasi berdasarkan *Confusion Matrix* Algoritme CART dapat dilihat pada [Tabel 4](#).

**Tabel 4.** Nilai Akurasi Berdasarkan *Confusion Matrix* Algoritme CART

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
0	0.729	0.678	0.703
1	0.944	0.955	0.949
<i>Weighted Avg</i>	0.911	0.914	0.912

Dengan menggunakan aplikasi Weka 3.8.6, pengujian dataset Penerima Beasiswa diperoleh hasil klasifikasi dengan waktu 0.03 seconds. Hasil perhitungan yang sudah dilakukan dengan menggunakan IBk/K-NN, NAÏVE BAYES, dan CART diperoleh perbandingan hasil akurasi yang dapat dilihat pada [Tabel 5](#). Algoritme CART mempunyai hasil akurasi paling tinggi dibandingkan algoritme K-NN dan NAÏVE BAYES.

**Tabel 5.** Perbandingan Hasil Akurasi Algoritme IBk/K-NN, NAÏVE BAYES, dan CART

<i>Algoritme</i>	<i>Hasil Akurasi</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>Waktu (Seconds)</i>
KNN	90.0844 %	0.896	0.901	0.897	0
NAÏVE BAYES	90.8228 %	0.904	0.908	0.905	0.03
CART	91.3502 %	0.911	0.914	0.912	0.03

Dari penelitian yang telah dilakukan terkait komparasi Algoritme K-NN, Naïve Bayes dan CART untuk memprediksi penerima beasiswa, Dari [Tabel 5](#). pada algoritme KNN diperoleh nilai akurasi sebesar 90.0844%, pada algoritme Naïve Bayes diperoleh nilai akurasi sebesar 90.8228%, dan pada algoritme CART diperoleh nilai akurasi sebesar 91.3502% diperoleh kesimpulan bahwa nilai akurasi CART lebih baik dibandingkan dengan KNN dan NAÏVE BAYES dalam memprediksi penerima beasiswa dan rekomendasi untuk penggunaan Algoritme yang digunakan dalam memprediksi penerima beasiswa adalah menggunakan algoritme CART. Upaya mengukur tingkat akurasi dari Algoritme KNN, NAÏVE BAYES dan CART digunakan *confusion matrix* yang diperoleh dari proses pengolahan data di aplikasi Weka.

**Tabel 6.** *Confusion Matrix* Algoritme KNN

<i>Class</i>	<b>0</b>	<b>1</b>
0	85	58
1	36	769
948	121	827

Pada [Tabel 6](#), jumlah data hasil bentukan *rule* yang diprediksi menerima beasiswa dengan data testing yang juga menerima beasiswa sebanyak 85. Jumlah data hasil bentukan *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi menerima beasiswa sebanyak 58, kemudian jumlah data hasil *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi menerima beasiswa sebanyak 36. jumlah data hasil *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi tidak menerima beasiswa sebanyak 769.

**Tabel 7.** *Confusion Matrix* Algoritme NAÏVE BAYES

<i>Class</i>	<b>0</b>	<b>1</b>
0	90	53
1	34	771
948	124	824

Pada [Tabel 7](#), jumlah data hasil bentukan *rule* yang diprediksi menerima beasiswa dengan data testing yang juga menerima beasiswa sebanyak 90. Jumlah data hasil bentukan *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi menerima beasiswa sebanyak 53, kemudian jumlah data hasil *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi

menerima beasiswa sebanyak 34. jumlah data hasil rule yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi tidak menerima beasiswa sebanyak 771.

**Tabel 8.** Confusion Matrix Algoritme CART

Class	0	1
0	97	46
1	36	769
948	133	815

Pada **Tabel 8**, jumlah data hasil bentukan *rule* yang diprediksi menerima beasiswa dengan data testing yang juga menerima beasiswa sebanyak 97. Jumlah data hasil bentukan *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi menerima beasiswa sebanyak 46, kemudian jumlah data hasil *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi menerima beasiswa sebanyak 36. jumlah data hasil *rule* yang diprediksi tidak menerima beasiswa dengan data testing yang diprediksi tidak menerima beasiswa sebanyak 769.

### Pembahasan

Hasil perbandingan algoritme K-NN, Naïve Bayes dan CART mendapatkan nilai hasil yaitu algoritme K-NN sebesar 90.0844%, algoritme Naive Bayes sebesar 90.8228%, dan algoritme CART sebesar 91.3502%. Melalui pengujian algoritme tersebut, instansi sekolah dapat menggunakan sebagai media pertimbangan dalam menentukan penerima beasiswa, sehingga memudahkan dalam proses seleksi hingga penetapan penerima beasiswa (Pujianto et al., 2018). Nilai akurasi pada penentuan penerima beasiswa dengan menggunakan 948 data, dapat diketahui bahwa dari ketiga algoritme menunjukkan bahwa algoritme CART yang memperoleh nilai akurasi tertinggi dari kedua algoritme yang digunakan. Didasarkan pada nilai *Precision*, *Recall* dan *F-Measure* pada algoritme CART. Algoritme CART memang cocok digunakan untuk data non parametrik dan jumlahnya besar (P. Subarkah et al., 2018). Beberapa penelitian yang menggunakan dan mengimplementasikan algoritme CART telah berhasil dan mampu memberikan nilai akurasi yang baik dan tinggi pada klasifikasi dan prediksi permasalahan yang ada (García et al., 2019; Khan et al., 2022; Shen et al., 2022). Tingkat akurasi Algoritme CART yang tinggi ditunjukkan pada penelitian Klasifikasi Bike Share Daily Dataset menggunakan berdasarkan gambar pada Global Feature Extraction untuk Akurasi tertinggi diperoleh pada CART dengan hasil akurasi sebesar 96% (Purnamawati et al., 2022). Pada penelitian Perbandingan Akurasi Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa, algoritma CART memberikan akurasi lebih baik daripada C4.5, yaitu 63,16% berbanding 61,54% (Alverina et al., 2018). Penelitian Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN, algoritme yang termasuk dalam eager learner yang salah satunya yaitu algoritme CART memiliki performa yang lebih baik dibandingkan lazy learner (k-NN) (Praningki & Budi, 2018).

### 4. SIMPULAN

Hasil perhitungan yang telah dilakukan pada KNN, NAÏVE BAYES dan CART pada penentuan penerima beasiswa dapat diperoleh kesimpulan Algoritme CART merupakan Algoritme dengan hasil akurasi yang paling tinggi untuk memprediksi penerima beasiswa dan berkategori *Good Classification*.

### 5. DAFTAR PUSTAKA

- Alverina, D., Chrismanto, A. R., & Santosa, R. G. (2018). Perbandingan Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa. *Jurnal Teknologi Dan Sistem Komputer*, 6(2), 76–83. <https://doi.org/10.14710/jtsiskom.6.2.2018.76-83>.
- Alwi, A., & Munirah. (2022). The Concept Of Naive Bayes And Its Simple Use For Prediction Final Score Of Student Examination Using R Language. *Jurnal Teknik Informatika (JUTIF)*, 3(1), 133–140. <https://doi.org/10.20884/1.jutif.2022.3.1.139>.
- Arianto, B. (2021). Dampak Pandemi COVID-19 terhadap Perekonomian Dunia. *Jurnal Ekonomi Perjuangan*, 2(2), 212–224. <https://doi.org/10.36423/jumper.v2i2.665>.
- Bermejo-Martín, G., Rodríguez-Monroy, C., & Núñez-Guerrero, Y. M. (2021). Water consumption range prediction in Huelva's households using classification and regression trees. *Water (Switzerland)*, 13(4). <https://doi.org/10.3390/w13040506>.

- Bermejo-Martín, G., Rodríguez-Monroy, C., Núñez-Guerrero, Y. M., S. S., Wang, H., Subarkah, P., Abdallah, M. M., Hidayah, S. O. N., Wazery, Y. M., Saber, E., Houssein, E. H., Ali, A. A., Amer, E., Lu, J., Qian, W., Li, S., Cui, R., Ye, H., Wu, P., ... Ali, H. A. (2021). Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. *IEEE Access*, 9(1), 121. <https://doi.org/10.1007/s12652-020-02883-2>.
- Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15(1), 27–33. <https://doi.org/10.1016/j.aci.2017.09.005>.
- Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, 135(May), 113325. <https://doi.org/10.1016/j.dss.2020.113325>.
- Darmayanti, I., Subarkah, P., Anunggilarsa, L. R., & Suhaman, J. (2021). Prediksi Potensi Siswa Putus Sekolah Akibat Pandemi Covid-19 Menggunakan Algoritme K-Nearest Neighbor. *Jurnal Sains Dan Teknologi*, 10(2), 230–238.
- Fitria, Linda., Neviyarni., Netrawati., & Karneli, Y. (2020). Cognitive Behavior Therapy Counseling Untuk Mengatasi Anxiety Dalam Masa Pandemi Covid-19. *Al-Irsyad: Jurnal Pendidikan Dan Konseling*, 10(1), 23–29.
- Gao, L., Li, D., Yao, L., & Gao, Y. (2021). Sensor drift fault diagnosis for chiller system using deep recurrent canonical correlation analysis and k-nearest neighbor classifier. *ISA Transactions*, xxx. <https://doi.org/10.1016/j.isatra.2021.04.037>.
- García, V. J., Márquez, C. O., Isenhardt, T. M., Rodríguez, M., Crespo, S. D., & Cifuentes, A. G. (2019). Evaluating the conservation state of the páramo ecosystem: An object-based image analysis and CART algorithm approach for central Ecuador. *Heliyon*, 5(10). <https://doi.org/10.1016/j.heliyon.2019.e02701>.
- Gu, J., & Lu, S. (2021). An effective intrusion detection approach using SVM with naïve Bayes feature embedding. *Computers and Security*, 103, 102158. <https://doi.org/10.1016/j.cose.2020.102158>.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques 3rd Edition*. Morgan Kauffman.
- Helena, K., Susanti, Y., & Respatiwan. (2019). Penerapan Metode Chi-Squared Automatic Interaction Detection (CHAID) dan Classification And Regression Trees (CART) pada Klasifikasi Status Kerja di Kabupaten Brebes. *Seminar Nasional Sains Dan Entrepreneurship VI*.
- Idris, M. (2019). Implementasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Angka Kelahiran. *Jurnal Pelita Informatika*, 7(3), 421–428.
- Karo, I. M. K., Fajari, M. Y., Fadhillah, N. U., & Wardani, W. Y. (2022). Benchmarking Naïve Bayes and ID3 Algorithm for Prediction Student Scholarship. *IOP Conference Series: Materials Science and Engineering*, 1232(1), 012002. <https://doi.org/10.1088/1757-899x/1232/1/012002>.
- Khan, A., Li, J. P., Hasan, M. K., Varish, N., Mansor, Z., Islam, S., Saeed, R. A., Alshammari, M., & Alhumyani, H. (2022). PackerRobo: Model-based robot vision self supervised learning in CART. *Alexandria Engineering Journal*, 61(12), 12549–12566. <https://doi.org/10.1016/j.aej.2022.05.043>.
- Kidd, D., Miner, J., Schein, M., Blauw, M., & Allen, D. (2020). Ethics across the curriculum: Detecting and describing emergent trends in ethics education. *Studies in Educational Evaluation*, 67(October 2019), 100914. <https://doi.org/10.1016/j.stueduc.2020.100914>.
- Laajaj, R., Moya, A., & Sánchez, F. (2022). Equality of opportunity and human capital accumulation: Motivational effect of a nationwide scholarship in Colombia. *Journal of Development Economics*, 154. <https://doi.org/10.1016/j.jdeveco.2021.102754>.
- Lu, J., Qian, W., Li, S., & Cui, R. (2021). Enhanced k-nearest neighbor for intelligent fault diagnosis of rotating machinery. *Applied Sciences (Switzerland)*, 11(3), 1–15. <https://doi.org/10.3390/app11030919>.
- Mittal, P., & Gill, N. S. (2014). Comparative Analysis Of Classification Techniques On Medical Data Sets. *IJRET: International Journal of Research in Engineering and Technology*, 3(6), 454–460.
- Mustakim, & Oktaviani F, G. (2016). *Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa*. 13(2), 195–202.
- Nawaz, M. S., Shoaib, B., & Ashraf, M. A. (2021). Intelligent Cardiovascular Disease Prediction Empowered with Gradient Descent Optimization. *Heliyon*, 7(5), e06948. <https://doi.org/10.1016/j.heliyon.2021.e06948>.
- Noviana, D., Susanti, Y., & Susanto, I. (2019). Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Algoritma C4.5. *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*, 79–87.
- Nurdiawan, O., Kurnia, D. A., Solihudin, D., Hartati, T., & Suprpti, T. (2021). Comparison of the K-Nearest Neighbor algorithm and the decision tree on moisture classification. *IOP Conference Series*:

- Materials Science and Engineering*, 1088(1), 012031. <https://doi.org/10.1088/1757-899x/1088/1/012031>.
- Okfalisa, Fitriani, R., & Vitriani, Y. (2018). The comparison of linear regression method and k-nearest neighbors in scholarship recipient. *Proceedings - 2018 IEEE/ACIS 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2018*, 194–199. <https://doi.org/10.1109/SNPD.2018.8441068>.
- Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2021). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*, xxx. <https://doi.org/10.1016/j.matpr.2021.07.382>.
- Permana, A. P., Ainiyah, K., & Holle, K. F. H. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 178–188. <https://doi.org/10.14421/jiska.2021.6.3.178-188>.
- Praningki, T., & Budi, I. (2018). Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN. *Creative Information Technology Journal*, 4(2), 83. <https://doi.org/10.24076/citec.2017v4i2.100>.
- Pujianto, A., Kusriani, K., & Sunyoto, A. (2018). Perancangan Sistem Pendukung Keputusan Untuk Prediksi Penerima Beasiswa Menggunakan Metode Neural Network Backpropagation. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(2), 157. <https://doi.org/10.25126/jtiik.201852631>.
- Purnamawati, A., Winnarto, M. N., & Mailasari, M. (2022). Analisis Cart (Classification and Regression Trees) Untuk Prediksi Pengguna Sepeda Berdasarkan Cuaca. *Jurnal Teknoinfo*, 16(1), 14. <https://doi.org/10.33365/jti.v16i1.1478>.
- Shen, Z., Yang, Y., Ai, L., Yu, C., & Su, M. (2022). A hybrid CART-GAMs model to evaluate benthic macroinvertebrate habitat suitability in the Pearl River Estuary, China. *Ecological Indicators*, 143(May), 109368. <https://doi.org/10.1016/j.ecolind.2022.109368>.
- Sisodia, D., & Sisodia, D. S. (2021). Engineering Science and Technology , an International Quad division prototype selection-based k-nearest neighbor classifier for click fraud detection from highly skewed user click dataset. *Engineering Science and Technology, an International Journal*, xxx. <https://doi.org/10.1016/j.jestch.2021.05.015>.
- Straub, R., & Vilsmaier, U. (2020). Pathways to educational change revisited– controversies and advances in the German teacher education system. *Teaching and Teacher Education*, 96, 103140. <https://doi.org/10.1016/j.tate.2020.103140>.
- Subarkah, P., Ikhsan, A. N., & Setyanto, A. (2018). The effect of the number of attributes on the selection of study program using classification and regression trees algorithms. *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*. <https://doi.org/10.1109/ICITISEE.2018.8721030>.
- Subarkah, Pungkas, Abdallah, M. M., & Hidayah, S. O. N. (2021). Komparasi Akurasi Algoritme CART Dan Neural Network Untuk Diagnosis Penyakit Diabetes Retinopathy. *CogITo Smart Journal*, 7(1), 121. <https://doi.org/10.31154/cogito.v7i1.304.121-134>.
- Sukamto, S., Adriyani, Y., & Aulia, R. (2020). Prediksi Kelompok UKT Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *JUITA: Jurnal Informatika*, 8(1), 121. <https://doi.org/10.30595/juita.v8i1.6267>.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining – Menggali Pengetahuan dari Bongkahan Data*. ANDI.
- Wang, B., & Mao, Z. (2020). A dynamic ensemble outlier detection model based on an adaptive k-nearest neighbor rule. *Information Fusion*, 63(May), 30–40. <https://doi.org/10.1016/j.inffus.2020.05.001>.
- Wazery, Y. M., Saber, E., Houssein, E. H., Ali, A. A., & Amer, E. (2021). An Efficient Slime Mould Algorithm Combined with K-Nearest Neighbor for Medical Classification Tasks. *IEEE Access*, 9, 113666–113682. <https://doi.org/10.1109/ACCESS.2021.3105485>.
- Zerlinda, H. N., Slamet, I., & Zukhronah, E. (2019). *Klasifikasi Calon Penerima Bidikmisi Dengan Menggunakan*. 88–93.