Jurnal Komputer Antartika
Volume 3 Nomor 1 2025 Hal. 18-25
ISSN (Online) 2988-5272
DOI: https://doi.org/10.70052/jka.v3i1.623

Perbandingan Akurasi Algoritma C4.5 dan K-NN Untuk Prediksi Kelulusan Mahasiswa Penerima Beasiswa

Andrian Fakih[1*], Mohammad Alfi Hamzami[2], Muhammad Raihan Hadianto[3], Najdah Ibtisamah Shafira Alifah[4]

Universitas Bina Sarana Informatika
Jl. Kramat Raya No 98, Jakarta Pusat, Indonesia

e-mail korespondensi: andriannfakih@gmail.com

Submit: 10-12-2024 | Revisi: 06-01-2025 | Terima: 10-01-2025 | Terbit online: 25-01-2025

Abstrak – Penelitian ini bertujuan untuk menganalisis permasalahan yang dihadapi oleh Administrator Universitas dalam pengambilan keputusan penerima beasiswa. Metode yang digunakan melibatkan pemrograman dinamis untuk menggantikan pendekatan tradisional yang masih sering digunakan. Penelitian ini berfokus pada perbandingan performa antara algoritma Decision Tree (C4.5) dan K-Nearest Neighbor (K-NN) dalam mengklasifikasikan data penerima beasiswa. Data yang diperoleh adalah sekunder, yaitu data historis penerima beasiswa yang telah dikumpulkan oleh pihak universitas. Tahapan penelitian dimulai dari pengumpulan dan pra-pemrosesan data, diikuti dengan penerapan algoritma C4.5 dan K-NN, serta evaluasi performa algoritma menggunakan metrik seperti akurasi, presisi, recall, dan AUC. Berdasarkan hasil pengujian, algoritma Decision Tree (C4.5) menunjukkan akurasi sebesar 74,95%, presisi 20,9%, recall 77,1%, dan AUC sebesar 0,752. Sementara itu, algoritma K-Nearest Neighbor hanya mencapai akurasi 71,79%, presisi 39,0%, recall 45,3%, dan AUC sebesar 0,734. Dengan demikian, algoritma Decision Tree (C4.5) memiliki performa yang lebih baik dalam menyelesaikan permasalahan klasifikasi ini dibandingkan algoritma K-Nearest Neighbor.

Kata Kunci : Beasiswa, Klasifikasi, C4.5, K-Nearest Neighbors

Abstracts *- This research aims to analyze the problems faced by University Administrators in making decisions about scholarship recipients. The method used involves dynamic programming to replace the traditional approach which is still often used. This research focuses on the performance comparison between the Decision Tree (C4.5) and K-Nearest Neighbor (K-NN) algorithms in classifying scholarship recipient data. Data was obtained from secondary sources, specifically historical records of scholarship recipients. Based on test results, the Decision Tree algorithm (C4.5) shows an accuracy of 74.95%, precision of 20.9%, recall of 77.1%, and AUC of 0.752. Meanwhile, the K-Nearest Neighbor algorithm only achieved 71.79% accuracy, 39.0% precision, 45.3% recall, and an AUC of 0.734. Thus, the Decision Tree algorithm (C4.5) performs better in solving this classification problem than the K-Nearest Neighbor algorithm.*

Keywords : Scholarship, Classification, C4.5, K-Nearest Neighbors

1. Pendahuluan

Salah satu kebutuhan untuk memperbaiki kehidupan masyarakat saat ini adalah pendidikan tinggi dikarenakan adanya globalisasi untuk menuntut adanya perbaikan kualitas pelayanan masyarakat di berbagai bidang. Dalam upaya meningkatkan kualitas sumber daya yang baik tentunya harus mendukung peningkatan sumber daya manusia. Peningkatan sumber daya manusia yang berkualitas dapat dilakukan melalui pendidikan, pendidikan di Indonesia dilakukan secara sistematis, berkelanjutan dan berjenjang dimulai dari pendidikan dasar, menengah hingga perguruan tinggi. Perguruan tinggi berperan penting dalam pengembangan pengetahuan dan keterampilan mahasiswa, berupaya menciptakan lulusan yang mampu berkontribusi bagi masyarakat. Untuk mendukung proses tersebut, beasiswa adalah salah satu sarana yang penting untuk mengakses ke pendidikan yang lebih tinggi, dari bantuan keuangan yang diberikan kepada individu dengan tujuan membantu mereka membiayai pendidikan. Organisasi atau yayasan menawarkan berbagai macam beasiswa yang disediakan dan umumnya diberikan kepada mahasiswa yang termasuk dalam kategori yang telah ditentukan [1].

Dalam hal ini, panitia perlu menyeleksi untuk memberikan bantuan beasiswa terhadap orang yang tepat, jika terdapat kesalahan dan juga tidak tepat sasaran maka hal tersebut menyebabkan penyalahgunaan bantuan beasiswa yang diberikan. Hal tersebut tentu saja membutuhkan berbagai macam ketentuan-ketentuan atribut di dalamnya seperti pendapatan, tanggungan dan prestasi akademik maupun non akademik. Dari atribut yang

bermacam-macam tersebut digunakan dalam proses seleksi untuk menentukan calon mahasiswa/I yang layak mendapatkan bantuan beasiswa [2].
Data mining adalah proses ekstraksi dan pemrosesan informasi dari data besar yang telah dikumpulkan. Salah satu metode yang digunakan dalam data mining adalah algoritma C4.5, yang membangun pohon keputusan untuk klasifikasi. Algoritma ini memiliki keunggulan dalam tingkat akurasi dan efisiensi dalam menangani atribut bertipe diskret maupun numerik. Namun, kelemahan utama algoritma C4.5 adalah keterbatasannya dalam hal skalabilitas, karena seluruh data harus dimuat ke dalam memori untuk diproses secara bersamaan [3].

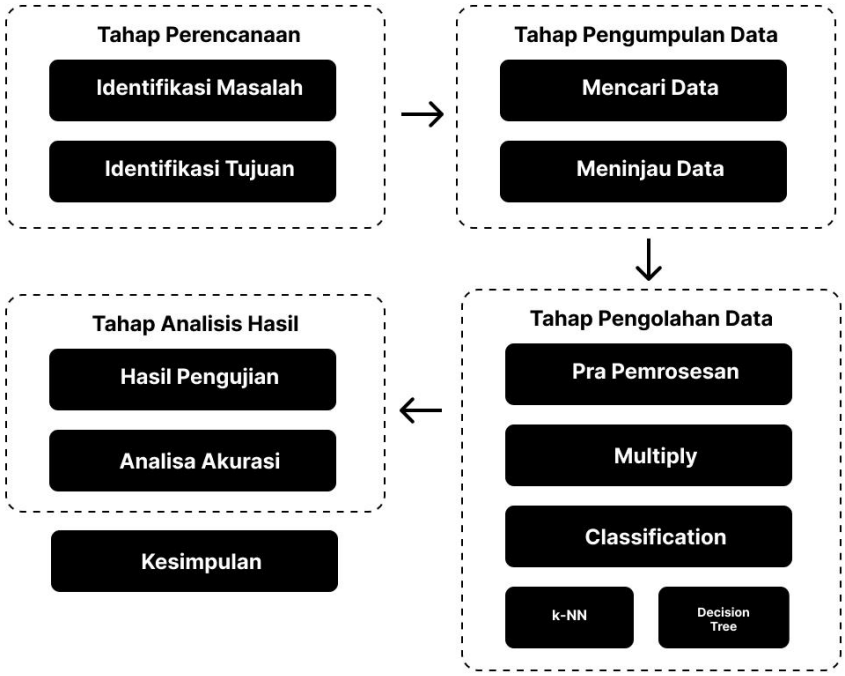
Sebagai alternatif, algoritma K-Nearest Neighbor (K-NN) digunakan untuk klasifikasi data berdasarkan jarak terdekat. Metode ini menghitung jarak antara data baru dan data yang sudah diklasifikasikan untuk menentukan kelasnya [4].

Penelitian ini akan memanfaatkan algoritma C4.5 dan K-NN untuk membantu proses seleksi penerima beasiswa secara lebih efektif dan akurat. Penelitian ini bertujuan untuk mengembangkan sistem yang mampu mengolah data atribut mahasiswa dengan lebih efisien, sehingga dapat membantu panitia seleksi dalam menentukan penerima beasiswa yang tepat sasaran.

Gap penelitian ini adalah belum adanya kajian yang mengintegrasikan algoritma C4.5 dan K-NN secara optimal dalam proses seleksi beasiswa. Oleh karena itu, penelitian ini akan mengeksplorasi integrasi kedua algoritma tersebut untuk meningkatkan akurasi dan efisiensi seleksi penerima beasiswa. Sistem yang dikembangkan diharapkan dapat menjadi solusi inovatif dalam proses seleksi beasiswa di masa depan.

2. Metode Penelitian

Pada temuan penelitian ini, digunakan pendekatan kuantitatif dengan desain eksperimental untuk menemukan perbandingan kinerja algoritma K-Nearest Neighbors dan Decision Tree, yang merupakan algoritma machine learning. Metodologi yang digunakan ditunjukan secara spesifik pada gambar 1.



Gambar 1. Alur Penelitian

2.1 Alur Penelitian

Tahap perencanaan, diawali melalui proses identifikasi masalah dan identifikasi tujuan. Diputuskan topik penerimaan beasiswa sebagai data yang akan digunakan untuk menentukan akurasi kedua algoritma machine learning. Tahap pengumpulan data, dimana pada penelitian ini data bersumber dari situs Kaggle. Kolom dan baris pada data tersebut kemudian diperiksa untuk melihat kesesuaian antara masalah yang ditentukan pada proses identifikasi.

Tahap ketiga yaitu tahap pengolahan data, dimana dataset melalui sejumlah tahapan pra-pemrosesan atau data cleaning sebelum diproses melalui operator dan algoritma pada aplikasi Rapid Miner. Pada tahap ini, dilakukan penghapusan nilai yang kosong atau hilang (missing values), membuang duplikasi, dan memperbaiki inkonsistensi pada data [5] [6]. Penetapan atribut kelas atau label pada data juga dilakukan pada tahap ini. Data

yang telah dibersihkan kemudian diproses dengan operasi multiply agar sejumlah operator yang berhubungan dapat dijalankan secara bersamaan [7]. Klasifikasi pada tahap ini berfungsi sebagai pemisah antara penggunaan algoritma k-NN dan C4.5 melalui operator cross validation. Teknik cross validation memungkinkan evaluasi yang lebih objektif terhadap kemampuan generalisasi model pada berbagai subset data yang belum pernah diproses sebelumnya [8].

Pada tahap keempat, evaluasi kinerja model dilakukan dengan menganalisis tingkat akurasi yang dicapai oleh algoritma k-NN dan C4.5, di mana performa algoritma K-NN dan C4.5 dalam memprediksi kelas akan dibandingkan berdasarkan tingkat akurasinya.

2.2 Decision tree (C4.5)

Decision tree atau Pohon keputusan adalah struktur hierarkis yang digunakan dalam data mining untuk mengklasifikasikan item dengan membagi data menjadi beberapa subset berdasarkan variabel input [9]. Terdapat berbagai jenis pohon keputusan, seperti ID3, CART, dan C4.5. Algoritma ID3, yang dikembangkan oleh Quinlan, kemudian disempurnakan menjadi algoritma C4.5 [10]. ID3 cocok untuk data kategorikal, sedangkan C4.5 dapat menangani data kategorikal dan numerik. Metode ID3 menggunakan *information gain* untuk pemilihan atribut, sedangkan C4.5 menggunakan *gain ratio* [11]. Tahapan yang terlibat dalam algoritma Pohon Keputusan adalah sebagai berikut [12]:

1. Siapkan dataset pelatihan.
2. Tentukan akar dari pohon keputusan.
3. Tentukan fitur akar dengan menghitung nilai *gain*. Fitur dengan nilai *gain* tertinggi dipilih sebagai akar. Nilai *gain* dihitung menggunakan Persamaan 1.

   Gain (S, A) = Entropy(S) - $\sum_{I=1}^{N} \frac{|si|}{|s|}$ x Entropy (Si) (1)
4. Ulangi langkah 2 untuk setiap cabang yang terbentuk. Hitung nilai entropi menggunakan rumus yang sesuai, seperti yang ditunjukkan dalam Persamaan 2.

   Entropy (S) = $\sum_{i=1}^{N} - \pi\, x\, log 2\pi$ (2)
5. Proses pembentukan pohon keputusan berakhir ketika semua cabang dari suatu node memiliki kelas yang sama.

2.3 K-Nearest Neighbors

Algoritma K-NN (*K-Nearest Neighbor*) adalah teknik klasifikasi data yang menentukan kelompok mana yang paling mungkin dimasuki oleh suatu titik data untuk memperkirakan kemungkinan titik data tersebut bergabung dengan kelompok tersebut [13]. K-NN merupakan contoh algoritma *lazy-learning*, di mana perhitungan hanya dilakukan secara lokal dan tidak menyelesaikan semua proses hingga tahap klasifikasi [14]. Algoritma ini bekerja dengan mencari k item dalam data pelatihan yang paling mirip dengan objek dalam data uji atau data baru [15]. Langkah-langkah dalam menghitung algoritma K-Nearest Neighbor adalah sebagai berikut [14] :

1. Menentukan parameter K.
2. Menghitung jarak antara data pelatihan dan data uji.
   Perhitungan jarak yang paling umum digunakan dalam algoritma KNN adalah jarak Euclidean. Rumusnya adalah sebagai beri kut:

   $$euc = \sqrt{(\sum(pi - qi)\ 2)\ n\ i = 1} \quad (3)$$

   Keterangan:
   - pi : Data pelatihan dan sampel
   - qi : Data uji
   - i : Variabel data
   - n : Dimensi data
3. *Mengurutkan jarak yang telah dihitung.*
4. Menentukan K jarak terdekat.
5. Menentukan kelas yang sesuai.
6. Menghitung jumlah kelas pada tetangga terdekat dan menetapkan kelas tersebut sebagai kelas data yang dievaluasi.

2.4 Evaluasi Kerja

a. K-fold Cross Validation

K-fold cross validation merupakan teknik validasi model yang digunakan untuk mengevaluasi kemampuan hasil analisis statistik dalam menggeneralisasi pada kumpulan data yang independen. Teknik ini terutama diterapkan untuk memprediksi performa model dan memperkirakan tingkat akurasi model prediktif saat digunakan dalam situasi nyata. Salah satu metode yang sering digunakan dalam validasi silang adalah *k-fold*

cross-validation, di mana dataset dibagi menjadi *k* bagian dengan ukuran yang sama. Metode ini membantu mengurangi bias pada data dengan melatih dan menguji model sebanyak *k* kali [16]. Pada penelitian ini, metode k-fold cross validation diterapkan untuk mengevaluasi kemampuan algoritma C4.5 dan K-NN dalam menggeneralisasi data baru, sekaligus meminimalkan risiko *overfitting*.

b. Confusion matrix

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa suatu metode klasifikasi. Confusion matrix pada dasarnya berisi informasi yang membandingkan hasil klasifikasi yang dihasilkan oleh sistem dengan hasil klasifikasi yang seharusnya. Tabel ini menunjukkan jumlah prediksi yang tepat (True Positive dan True Negative) serta jumlah prediksi yang salah (False Positive dan False Negative). True Positive (TP) adalah kumpulan data di mana tupel positif diklasifikasikan dengan benar sebagai positif. False Positive (FP) terjadi ketika tupel positif dalam data diklasifikasikan secara keliru sebagai negatif. Sebaliknya, False Negative (FN) adalah kondisi ketika tupel negatif diklasifikasikan secara salah sebagai positif. Terakhir, True Negative (TN) adalah kumpulan data di mana tupel negatif diklasifikasikan dengan benar sebagai negatif [17] . Berikut rumus-rumus matrix yang digunakan:

$$\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{P} = \frac{TP}{TP+Fp} \quad (5)$$

$$Sn = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Sp} = \frac{TN}{TN+FP} \quad (7)$$

c. Kurva ROC

Kurva ROC (*Receiver Operating Characteristic*) sering digunakan oleh peneliti untuk mengevaluasi hasil prediksi. Kurva ini menunjukkan kinerja klasifikasi tanpa mempertimbangkan distribusi kelas atau jenis kesalahan. Sumbu vertikal pada kurva merepresentasikan nilai positif (True Positive/TP), sementara sumbu horizontal menggambarkan nilai negatif (False Positive/FP).

Salah satu cara umum untuk mengukur performa kurva ROC adalah dengan menghitung luas di bawah kurva tersebut, yang dikenal sebagai *Area Under Curve* (AUC). Nilai AUC selalu berada dalam rentang 0,0 hingga 1,0. Luas di atas 0,5 dianggap relevan, di mana semakin besar luasnya, semakin baik kinerja klasifikasi. Berikut panduan klasifikasi berdasarkan nilai AUC [17]:

- **0.9–1.0**: Klasifikasi sangat baik
- **0.8–0.9**: Klasifikasi baik
- **0.7–0.8**: Klasifikasi rata-rata
- **0.6–0.7**: Klasifikasi rendah
- **0.5–0.6**: Kegagalan

2.5 Perbandingan pada Evaluasi Kerja

Setelah melalui proses evaluasi kinerja menggunakan k-fold cross validation, confusion matrix, dan kurva ROC, hasil dari kedua algoritma yaitu K-Nearest Neighbors (K-NN) dan C4.5 akan dibandingkan. Perbandingan akan difokuskan pada metrik-metrik evaluasi yang telah dihitung, seperti akurasi, precision, recall, F1-score, dan Area Under Curve (AUC). Metrik-metrik ini akan memberikan gambaran mengenai optimasi kinerja masing-masing algoritma dalam mengklasifikasikan data.

Dengan membandingkan nilai-nilai metrik tersebut, jenis algoritma yang memiliki kinerja terbaik dapat diidentifikasi untuk kebutuhan memprediksi kelulusan mahasiswa penerima beasiswa.

3. Hasil dan Pembahasan

3.1. Dataset

Pengelolahan data dibagi menjadi data training dan data testing jumlah 1042 record yang terdiri dari 13 atribut, Data tersebut bisa dilihat di website kaggle dengan title 1000-dataset-beasiswa dibuat oleh Hilman Abdurrohim Azis. Dataset beassiswa ditampilkan dalam gambar 2.

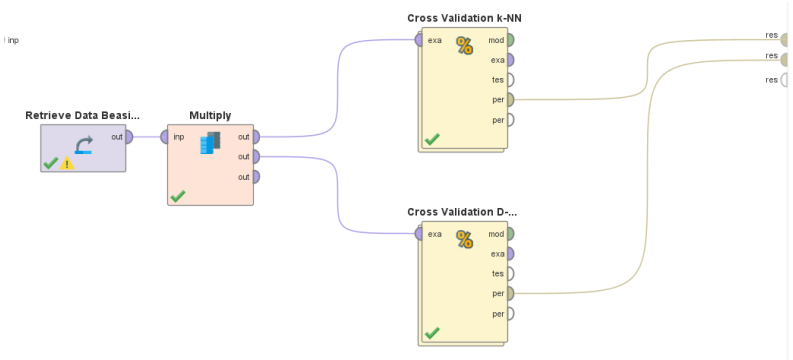
	Ikut Organis... polynominal	Ikut UKM polynominal	IPK real	Pekerjaan O... polynominal	Penghasilan polynominal	Tanggungan integer	Status Beasi... polynominal label
1	Ikut	Ikut	3.570	Wiraswasta	Sedang	4	Terima
2	Tidak	Ikut	2.950	Buruh	Sedang	2	Tidak
3	Tidak	Ikut	3.670	Petani	Sedang	4	Terima
4	Tidak	Ikut	3.190	Wiraswasta	Tinggi	2	Tidak
5	Tidak	Ikut	3.190	Wiraswasta	Sedang	2	Tidak
6	Tidak	Ikut	3.780	Wiraswasta	Rendah	3	Terima
7	Tidak	Ikut	3.330	Petani	Sedang	3	Tidak
8	Tidak	Ikut	3.290	Petani	Rendah	1	Tidak
9	Tidak	Ikut	3.480	Wiraswasta	Sedang	3	Tidak
10	Tidak	Ikut	3.380	Pedagang	Sedang	1	Tidak
11	Tidak	Ikut	3.520	Petani	Rendah	2	Tidak
12	Tidak	Ikut	2.810	Pedagang	Rendah	2	Tidak
13	Tidak	Ikut	3.700	Karyawan Swasta	Sedang	1	Tidak
14	Tidak	Ikut	3.710	Petani	Rendah	2	Tidak
15	Tidak	Ikut	3.380	Wiraswasta	Tinggi	1	Tidak
16	Tidak	Ikut	3.570	Petani	Rendah	2	Tidak
17	Tidak	Ikut	3.290	Pedagang	Rendah	2	Tidak
18	Tidak	Ikut	3.000	Wiraswasta	Rendah	1	Tidak

Gambar 2. Dataset

3.2. Pembahasan

3.2.1 Eksperimen dan Pengujian

Pada tahap implementasi algoritma ini, dilakukan proses pengujian. Dalam pengujian, dari total 1043 data hasil preprocessing, data dibagi menjadi dua bagian, yaitu data latih (training) dan data uji (testing). Pada pengujian ini, digunakan metode validasi silang(Cross-validation). Metode ini akan memisahkan data dengan membagi keseluruhan data menjadi data pelatihan dan data pengujian. Proses ini dilakukan menggunakan RapidMiner untuk mengklasifikasikan penerima beasiswa dengan menerapkan algoritma Decision Tree dan K-Nearest Neighbor dapat dilihat pada Gambar 4.



Gambar 3. Skema pengujian pada RapidMiner

3.2.2 Evaluasi dan Validasi Hasil

1. Hasil Pengujian Algoritma k-NN

Pada tahap ini, peneliti menerapkan algoritma k-NN untuk mengolah data yang telah melalui proses preprocessing atau pembersihan data menggunakan aplikasi RapidMiner. Berdasarkan pengujian yang dilakukan dengan aplikasi tersebut, diperoleh hasil bahwa algoritma k-NN mencapai tingkat akurasi 71.79% yang ditampilkan melalui gambar 4.

accuracy: 71.79% +/- 2.66% (micro average: 71.79%)

	true Terima	true Tidak	class precision
pred. Terima	106	128	45.30%
pred. Tidak	166	642	79.46%
class recall	38.97%	83.38%	

Gambar 4. Confusion Matrix k-NN

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} = \frac{106 + 642}{106 + 642 + 166 + 128} = \frac{748}{1042} = 71.79\%$$

2. Hasil Pengujian Algoritma C4.5

Pada tahap ini, peneliti menerapkan algoritma C4.5 untuk mengolah data yang telah melalui proses preprocessing atau pembersihan data menggunakan aplikasi RapidMiner. Berdasarkan pengujian yang dilakukan dengan aplikasi tersebut, diperoleh hasil bahwa algoritma C4.5 mencapai tingkat akurasi 74.95% yang ditampilkan

pada gambar 5.

accuracy: 74.95% +/- 2.31% (micro average: 74.95%)

	true Terima	true Tidak	class precision
pred. Terima	57	46	55.34%
pred. Tidak	215	724	77.10%
class recall	20.96%	94.03%	

Gambar 5. Confusion Matrix C4.5

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} = \frac{57 + 724}{57 + 724 + 215 + 46} = \frac{781}{1042} = 74.95\%$$

Hasil pengujian mencakup beberapa metrik evaluasi, yaitu akurasi, precision, dan recall. Akurasi menggambarkan tingkat kesesuaian antara nilai derajat yang diukur dengan nilai sebenarnya. Precision menunjukkan sejauh mana sistem mampu memilih dokumen teks yang relevan dari seluruh dokumen yang tersedia, sedangkan recall mengukur proporsi dokumen teks relevan yang berhasil ditemukan dari keseluruhan dokumen relevan dalam koleksi. Perbandingan hasil pengujian antara Algoritma C4.5 dan K-Nearest Neighbor dapat dilihat pada Tabel 1.

Tabel 1 Perbandingan Kinerja

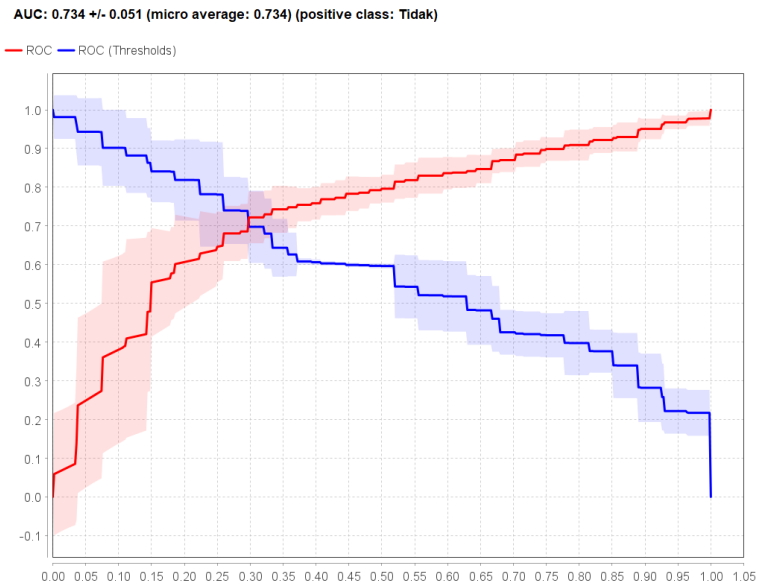
Model	Akurasi	Presisi	Recall
K-Nearest Neighbor	71.79%	39.0%	45.30%
Decision Tree (C4.5)	74.95%	20.9%	77.10%

Selain menggunakan Confusion Matrix untuk mengevaluasi kinerja pengujian, kami juga mengandalkan kurva ROC/AUC (Area Under Curve) yang dihasilkan. Perbandingan hasil kurva AUC antara algoritma C4.5 dan K-Nearest Neighbor dapat dilihat pada Tabel 2 berikut ini.

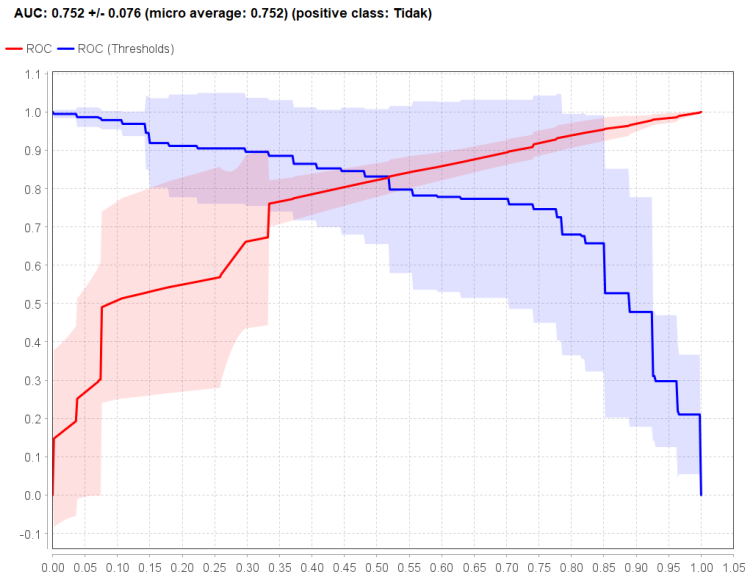
Tabel 2 Perbandingan hasil AUC

Model	AUC
K-Nearest Neighbor	0.734
Decision Tree (C4.5)	0.752

Gambar 6 adalah grafik yang menunjukkan perbandingan hasil kinerja kedua algoritma dari tabel 2:



Gambar 6. AUC K-Nearest Neighbor



Gambar 7. AUC C4.5

Pada gambar-gambar 7 menampilkan kurva ROC (Receiver Operating Characteristic) untuk dua algoritma, yaitu K-Nearest Neighbor (K-NN) dan Decision Tree (C4.5), yang digunakan untuk klasifikasi data. Berikut adalah penjelasan dari kedua gambar:

1. Gambar 6 (AUC K-Nearest Neighbor):
    - Kurva ROC untuk algoritma K-Nearest Neighbor (K-NN) menunjukkan performa model berdasarkan kemampuan memisahkan kelas positif dan negatif.
    - Nilai AUC (Area Under the Curve) sebesar **0.734** menunjukkan bahwa model memiliki kemampuan klasifikasi yang moderat, dengan performa yang tidak terlalu baik namun cukup dapat diterima.
    - Semakin dekat kurva ke sudut kiri atas, semakin baik performa model.

2. Gambar 7 (AUC Decision Tree C4.5)
    - Kurva ROC untuk algoritma Decision Tree (C4.5) menunjukkan performa model dengan nilai AUC sebesar **0.752**.
    - Nilai ini lebih tinggi dibandingkan K-NN, menunjukkan bahwa Decision Tree (C4.5) memiliki kemampuan yang lebih baik dalam mengklasifikasikan data.
    - Interval kepercayaan (± 0.076) pada kurva ini sedikit lebih besar daripada K-NN, menunjukkan bahwa performa model C4.5 mungkin lebih bervariasi tergantung pada data uji.
    - Kurva yang mendekati sudut kiri atas menegaskan bahwa model ini lebih akurat dibandingkan K-NN.

4. Kesimpulan

Penelitian yang telah dilakukan untuk mengembangkan skema penerimaan beasiswa yang optimal dan pemerataan bagi universitas. Metode yang digunakan bisa diterapkan dengan harapan dapat membantu administrator universitas agar dapat menyeleksi dengan mudah. Metode tradisional yang digunakan sebelumnya memakan banyak waktu. Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwasanya algoritma Decision Tree atau C4.5 memiliki performansi yang lebih baik dibanding algoritma K-Nearest Neighbor dikarenakan hasil yang didapatkan oleh algoritma C4.5 atau Decision Tree dengan tingkat akurasi 74.95%, presisi sebesar 20.9%, recall sebesar 77.10% dan AUC 0.752, sedangkan jika menggunakan algoritma K-Nearest Neighbor tingkat akurasi hanya mendapatkan 71.79%, presisi sebesar 39.0%, recall sebesar 45.30% dengan hasil AUC 0.734.

Referensi

[1] B. Harpad, M. Fahmi, P. Pahrudin, and R. Andrea, “Pengelompokan Siswa Layak Penerima Beasiswa dengan Menerapkan Algoritma K-Means Clustering Data Mining,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 913, 2024, doi: 10.30865/mib.v8i2.7394.

[2] I. Arfyanti, M. Fahmi, and P. Adytia, “Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1196–1201, 2022, doi: 10.47065/bits.v4i3.2275.

[3] A. Purwanto, “Analisa Perbandingan Kinerja Algoritma C4.5 Dan Algoritma K-Nearest Neighbors Untuk

Klasifikasi Penerima Beasiswa," *J. Teknoinfo*, vol. 14, pp. 48–58, 2020.
[4] D. H. H. S. Kamagi, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Proc. - 2016 IEEE Int. Power Electron. Motion Control Conf. PEMC 2016*, vol. VI, no. 1, pp. 254–260, 2014, doi: 10.1109/EPEPEMC.2016.7752007.
[5] R. Hadi, I. G. A. Desi Saryanti, and I. G. N. Ady Kusuma, "Implementasi Aplikasi Sederhana Dengan Klasifikasi Penjualan Berbasis K-Nearest Neighbor Pada Aplikasi Desktop," *J. Innov. Res. Knowl.*, vol. 1, no. 9, pp. 1–8, 2022.
[6] J. C. Mestika, M. O. Selan, and M. I. Qadafi, "Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python," *Bul. Ilm. Ilmu Komput. dan Multimed.*, vol. 1, no. 1, pp. 216–219, 2023.
[7] N. Khansa and Z. Fatah, "Gudang Jurnal Multidisiplin Ilmu Analisis Perbandingan Algoritma Machine Learning Dalam Klasifikasi Gangguan Tidur," vol. 2, no. November, pp. 76–81, 2024.
[8] I. H. Witten and E. Frank, *Data Mining Practical Machine Learning Tools and Techniques*. 2005.
[9] M. R. Anugrah, N. A. Al-Qadr, N. Nazira, and N. Ihza, "Implementation of C4.5 and Support Vector Machine (SVM) Algorithm for Classification of Coronary Heart Disease," *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, pp. 20–25, 2023, doi: 10.57152/predatecs.v1i1.805.
[10] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
[11] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698–703, 2019, doi: 10.1016/j.promfg.2019.06.011.
[12] A. Cherfi, K. Nouira, and A. Ferchichi, "Very Fast C4.5 Decision Tree Algorithm," *Appl. Artif. Intell.*, vol. 32, no. 2, pp. 119–137, 2018, doi: 10.1080/08839514.2018.1447479.
[13] F. Musa, F. Basaky, and O. E.O, "Obesity prediction using machine learning techniques," *J. Appl. Artif. Intell.*, vol. 3, no. 1, pp. 24–33, 2022, doi: 10.48185/jaai.v3i1.470.
[14] H. Rajaguru and S. R. Sannasi Chakravarthy, "Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer," *Asian Pacific J. Cancer Prev.*, vol. 20, no. 12, pp. 3777–3781, 2019, doi: 10.31557/APJCP.2019.20.12.3777.
[15] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.
[16] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 577–584, 2018, doi: 10.25126/jtiik.201855983.
[17] A. Sucipto, "Credit Prediction With Neural Network Algorithm," *Pros. Semin. Nas. Multi Disiplin Ilmu Call Pap. Unisbank*, vol. 978-979–36, no. 15, pp. 1–10, 2012.