

ANALYSIS OF C 4.5 ALGORITHM AND K-NEAREST NEIGHBOR TO DETERMINE SCHOLARSHIP RECEIPT AT SMK INDRA BANGSA

ANALISIS ALGORITMA C 4.5 DAN K-NEAREST NEIGHBOR UNTUK MENENTUKAN PENERIMAAN BEASISWA DI SMK INDRA BANGSA

Muhammad Fauzi Firdaus¹, Sugiyanti²

Program Studi Sistem Komputer, Ilmu Komputer dan Universitas Pamulang Kampus Serang

E-mail: dosen03039@unpam.ac.id¹, dosen03040@unpam.ac.id²

ARTICLE INFO

Correspondent:

Muhammad Fauzi Firdaus
dosen03039@unpam.ac.id

Key words:

Scholarships, Data Mining, C4.5 Algorithm, K-Nearest Neighbor

Website:

<https://idm.or.id/JSCR/index.php/JSCR>

Page: 885 - 896

ABSTRACT

Scholarships are one solution to overcome the problem of costs for those who are less fortunate. The provision of scholarships at SMK Indra Bangsa sometimes only focuses on orphans and for students who have good grades so they do not get scholarships, so this study aims to determine the classification of scholarship acceptance at SMK Indra Bangsa, for the data used as the research unit are SMK students from grades 10-12 totaling 56 students where later from that number 30% will be taken which will be used as testing data. With the data mining algorithm, the attribute data used for processing includes family status data, number of dependents of parents, student ranking and parents' income. And based on the data mining process, it was found that the K-Nearest Neighbor algorithm is more accurate than the C4.5 algorithm in classifying scholarship acceptance. And based on the results of the scholarship test with the rapid miner application, the accuracy value was obtained that the KNN algorithm got a higher accuracy value of 88.24% while the C4.5 algorithm got an accuracy value of 82.35%. In this case, it can be seen that the best method to use is the K-Nearest Neighbor algorithm, then followed by the C4.5 algorithm. The suggestion of this research is to use more attributes and use other algorithms for the testing process. The update of the research made is to use 2 algorithms in the data mining process carried out at SMK Indra Bangsa school.

Copyright ©2025 JSCR. All rights reserved.

INFO ARTIKEL

Koresponden

Muhammad Fauzi
Firdaus
dosen03039@umpam.ac.id

Kata kunci:
Beasiswa, Data Mining,
Algoritma C4.5, K-
Nearest Neighbor

Website:
<https://idm.or.id/JSCR/index.php/JSCR>

Hal: 885 - 896

ABSTRAK

Beasiswa merupakan salah satu solusi untuk mengatasi masalah biaya bagi mereka yang kurang mampu. Pemberian beasiswa di SMK Indra Bangsa terkadang hanya berfokus kepada siswa yatim saja dan bagi siswa yang memiliki nilai bagus jadi tidak mendapat beasiswa, maka pada penelitian ini bertujuan untuk menentukan klasifikasi penerimaan beasiswa di SMK Indra Bangsa, untuk data yang dijadikan unit penelitian adalah siswa SMK dari kelas 10-12 yang berjumlah 56 siswa yang dimana nantinya dari jumlah tersebut diambil 30% yang akan dijadikan sebagai data testing. Dengan algoritma data mining tersebut, adapun data atribut yang digunakan untuk melakukan pemrosesan diantaranya data status keluarga, jumlah tanggungan orang tua, peringkat siswa dan penghasilan orang tua. Dan berdasarkan proses data mining didapatkan bahwa algoritma K-Nearest Neighbor lebih akurat dibanding algoritma C4.5 dalam pengklasifikasian penerimaan beasiswa. Dan berdasarkan hasil pengujian beasiswa dengan aplikasi rapid miner didapatkan nilai akurasi bahwa algoritma KNN mendapat nilai akurasi yang lebih tinggi yaitu 88.24% sedangkan algoritma C4.5 mendapatkan nilai akurasi 82.35%. Dalam hal ini bisa diketahui untuk pemakaian metode yang terbaik adalah algoritma K-Nearest Neighbor, kemudian diikuti oleh algoritma C4.5. Adapun saran penelitian ini adalah menggunakan atribut yang lebih banyak dan menggunakan algoritma lain untuk proses pengujian. Adapun pembaruan dari penelitian yang dibuat adalah menggunakan 2 algoritma pada proses mining data yang dilakukan di sekolah SMK Indra Bangsa.

Copyright ©2025 JSCR. All rights reserved.

PENDAHULUAN

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh. Beasiswa dapat diberikan oleh lembaga pemerintah, perusahaan ataupun yayasan. Bantuan ini biasanya berbentuk dana untuk menunjang biaya atau ongkos yang harus dikeluarkan oleh siswa selama menempuh masa pendidikan. Dengan adanya bantuan ini, diharapkan siswa dapat menyelesaikan pendidikannya tanpa ada gangguan terutama yang berhubungan dengan keuangan siswa hingga tuntas atau lulus di jenjang pendidikan. Beasiswa di sekolah pada umumnya berupa beasiswa prestasi dan beasiswa kurang mampu (Safitri, 2020).

Beasiswa Bantuan, merupakan jenis beasiswa yang paling banyak dibutuhkan bagi para pelajar, beasiswa bantuan sendiri biasanya lebih spesifik untuk para pelajar yang tidak mampu membayar uang sekolah. Jenis beasiswa ini biasanya bersifat lokal pada lingkup sekolah atau suatu kota saja. Beasiswa bantuan akan diberikan oleh sekolah kepada pelajar yang datang dari keluarga tidak mampu. Bentuk beasiswanya dapat berupa uang sekolah bulanan atau semester, bantuan buku dan atribut seragam dan lain-lainnya. Beasiswa dapat penuh ditanggung 100% ataupun sebagian saja, bergantung pada program beasiswanya dan bagaimana kondisi penerima beasiswa. Beasiswa ini harapannya dapat menyasar langsung secara tepat sasaran untuk siswa

yang benar-benar membutuhkan, terutama untuk pelajar cerdas yang terkendala biaya (Apdian et al., 2024).

Dari beberapa jenis beasiswa yang sudah dijelaskan sebelumnya pada umumnya yang digunakan di sekolah di SMK Indra Bangsa sendiri pembagian beasiswa hanya sebatas di peruntukan untuk siswa yang kurang mampu, yatim atau peserta didik yang orang tua nya mengalami kesulitan ekonomi. Pemberian beasiswa itu sendiri di berikan secara simbolis melalui dana BOS, yang secara langsung memudahkan para siswa di SMK tersebut untuk membiayai SPP bulanan mereka. Maka pada jenis pemberian beasiswa ini dapat kita kategorikan bahwa beasiswa yang diberikan adalah jenis beasiswa bantuan.

Meski demikian pemberian beasiswa tersebut masih terkesan tidak relevan ataupun seimbang, dikarenakan penentuan penerimaan siswa yang akan mendapat beasiswa itu dilakukan berdasarkan keputusan pimpinan yayasan yang meliputi ketua yayasan, kepala sekolah dan para staff guru, yang terkadang hal atau keputusan itu hanya berdasarkan ke satu tujuan masalah saja yaitu ditujukan kepada siswa yang tidak memiliki orang tua (yatim) yang dimana hal tersebut dapat membuat beberapa siswa yang sekira juga perlu mendapatkan bantuan jadi tidak bisa mendapatkan bantuan beasiswa tersebut dikarenakan siswa yang memiliki keunggulan di bidang prestasi akademik tidak bisa mendapatkan beasiswa dikarenakan kondisi kedua orang tua mereka masih lengkap walaupun kadang ada saja siswa yang memiliki orang tua tetapi masih perlu mendapatkan bantuan beasiswa dikarenakan kondisi keuangan yang belum mencukupi.

Pada penelitian ini dilakukan menggunakan data siswa SMK Indra Bangsa dari kelas 10 sampai kelas 12 yang berjumlah 56 siswa, pengujian dilakukan menggunakan menggunakan pengujian beberapa algoritma diantaranya algoritma C4.5, dan algoritma K-Nearest Neighbor. Alasan penulis memilih menggunakan algoritma yang telah disebutkan ialah karena algoritma tersebut memiliki kemampuan yang cukup baik dalam proses pengklasifikasian data. Dalam ini juga terdapat beberapa kategori dalam proses pengujian beasiswa yaitu menggunakan beberapa atribut data yang diperlukan untuk proses mining algoritma, yaitu data status murid dengan orang tua, penghasilan orang tua, tanggungan orang tua dan juga peringkat siswa dikelas. Seluruh data yang akan diuji atau diolah nanti nya akan diproses menggunakan aplikasi data mining Rapid Miner dalam proses pengujian algoritma dan proses mining data.

METODE PENELITIAN

Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasi spesies berdasarkan karakteristik fisik. Selanjutnya dia dikenal sebagai bapak klasifikasi (Josua et al., 2022).

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu Kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Novika et al., 2021).

Metode Pemilihan Sampel

Pemilihan sampel dilakukan hanya kepada siswa SMK Indra Bangsa, data yang dipilih adalah data seluruh siswa dari kelas 10 -12, dengan jumlah data 56 siswa untuk dataset

dan diambil 17 data testing untuk dilakukan proses perhitungan algoritma, berikut data siswa yang akan dilakukan proses perhitungan

Metode Pengumpulan Data

Adapun teknik pengumpulan data yang penulis gunakan sebagai berikut:

Wawancara, dengan cara mengajukan pertanyaan langsung kepada kepala sekolah serta guru dan murid untuk mendapatkan informasi yang lebih akurat.

Studi Pustaka, dengan cara mempelajari dan membaca literatur-literatur, catatan-catatan, dan laporan-laporan yang ada hubungannya dengan masalah yang berhubungan dengan permasalahan yang menjadi obyek penelitian.

Observasi, pada tahap ini penulis memperoleh berbagai data secara pengamatan dan peninjauan langsung terhadap objek penelitian untuk mengetahui gambaran yang terjadi pada SMK Indra Bangsa.

Perancangan Penelitian

Adapun pada penelitian ini atribut dan variable yang digunakan untuk melakukan proses mining data algoritma adalah sebagai berikut:

Tabel 1. Data Atribut Siswa

Atribut	Kategori
Status Orang Tua	Yatim
	Punya Orang Tua
Penghasilan Orang Tua	<1.000.000
	1.000.000 - 2.999.999
	3.000.000 - 4.999.999
	>5.000.000
Tanggungan keluarga	1
	2 - 3
	>3
Peringkat	1
	2 - 3
	4 - 10
	>10

Waktu dan Tempat Penelitian

Sekolah SMK Indra Bangsa adalah SMK swasta yang terletak di jalan KH. Mustofa No. 27 Poris Gaga Baru, Kecamatan Batu Ceper, Kota Tangerang, Provinsi Banten.

Visi SMK Indra Bangsa

“Mewujudkan Insan Yang Unggul, Mandiri, Terampil Dalam Ilmu Teknologi Dan Berakhlakul Karimah”.

Misi SMK Indra Bangsa

- Menjalankan Nilai-nilai Agama, Pancasila, Sosial Budaya, Karakter Bangsa Yang Berakhlakul Karimah.
- Mencetak Siswa Yang Mandiri, Beriman dan Bertaqwa Sebagai Bekal Untuk Masa Depan.
- Mengembangkan Ilmu Pengetahuan dan Teknolgi Yang Unggul Sebagai Dasar Untuk Melanjutkan Ke Perguruan Tinggi, Dunia Kerja dan Dunia Usaha.
- Menguasai Bahasa Asing (Bahasa Inggris, Bahasa Arab) dan Unggul Dalam Berkompetisi di Era Globalisasi.
- Mengembangkan Mutu Dunia Pendidikan Dengan Kerjasama Yang Baik Antara Orang Tua, Stakeholder dan Instansi Terkait Serta Mewujudkan Sekolah Pilihan Masyarakat.

HASIL DAN PEMBAHASAN

Dalam menentukan hasil penelitian ini menggunakan dataset berjumlah 56 data yang akan digunakan untuk pemrosesan data dan 17 data testing untuk melakukan proses perhitungan algoritma.

Adapun data contoh yang akan digunakan untuk melakukan proses perhitungan data mining adalah sebagai berikut:

Tabel 2. Data Testing Siswa

No	Nama	Status dengan orang tua	Penghasilan orang tua	Tanggungan orang tua	Peringkat kelas	Beasiswa
1	Agam Abdul Rohman	Punya orang tua	1.000.000 - 2.999.999	> 3	> 10	Ya
2	Aldila	Punya orang tua	3.000.000 - 5.000.000	1	4 - 10	Tidak
3	Ananda Risma	Yatim	< 1.000.000	1	> 10	Tidak
4	Anindya giani ayu	Punya orang tua	> 5.000.000	> 3	1 - 3	Tidak
5	Auliya Nur Sabrina	Punya orang tua	3.000.000 - 5.000.000	2 - 3	1 - 3	Tidak
6	Aziz Faqih	Punya orang tua	1.000.000 - 2.999.999	2 - 3	> 10	Tidak
7	Durmasarju Egdinan Jason	Punya orang tua	< 1.000.000	2 - 3	> 10	Tidak
8	Fahma Yanti	Yatim	< 1.000.000	2 - 3	> 10	Ya
9	Fajar	Yatim	< 1.000.000	2 - 3	> 10	Ya
10	Farhan Ramadhan	Punya orang tua	1.000.000 - 2.999.999	2 - 3	4 - 10	Tidak
11	Fitri Lubis	Punya orang tua	1.000.000 - 2.999.999	2 - 3	> 10	Tidak
12	Fitriyani Saputri	Punya orang tua	> 5.000.000	> 3	4 - 10	Tidak
13	Muhammad Rizki Ramadhan	Punya orang tua	1.000.000 - 2.999.999	2 - 3	> 10	Tidak
14	Mutiara	Punya orang tua	3.000.000 - 5.000.000	2 - 3	4 - 10	Tidak
15	Nabila Khairunnisah	Yatim	< 1.000.000	> 3	4 - 10	Ya
16	Nayla	Punya orang tua	< 1.000.000	2 - 3	4 - 10	Tidak
17	Yuda maulana arifin	Punya orang tua	1.000.000 - 2.999.999	2 - 3	1 - 3	Tidak

Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross qiunlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai entropy atribut tersebut sebagai poros atribut klasifikasi (Ryanwar, 2020).

Pada tahapan ini dilakukan proses data mining dengan menggunakan algoritma C4.5. Data tersebut akan dijadikan sebagai data contoh perhitungan untuk proses pembentukan pohon keputusan menggunakan algoritma C4.5.

Pohon keputusan atau decision tree merupakan teknik data mining yang digunakan untuk mengeksplorasi data dengan membagi kumpulan data yang besar menjadi

himpunan record yang lebih kecil dan memperhatikan variabel tujuannya (Moerdyanto, et al., 2023) Proses pembentukan pohon keputusan dengan menggunakan algoritma C4.5 ditentukan dari perhitungan nilai Entropy dan nilai Gain. Adapun atribut-atribut yang digunakan sebanyak 4 yaitu status siswa, penghasilan orang tua, tanggungan orang tua dan peringkat kelas serta menggunakan 2 kelas label yaitu “Ya” dan “Tidak”.

Langkah pertama adalah menentukan dan menghitung jumlah siswa yang mendapatkan beasiswa atau tidak, setelah itu maka akan dilakukan proses perhitungan nilai entropy dari total kasus yang ada (Sari, et al., 2024).

Adapun rumus untuk menghitung nilai Entropy sebagai berikut:

$$Entropy(S) = \sum - p_j * \log_2 p_j$$

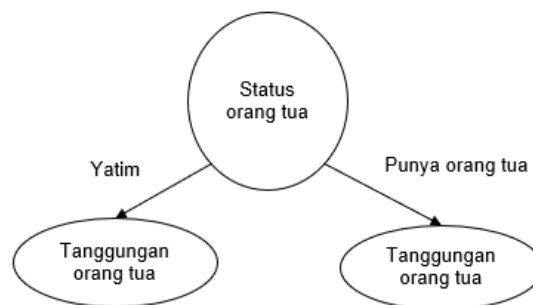
Langkah selanjut nya adalah mencari nilai Gain tertinggi dari setiap nilai entropy dari tiap atribut yang sudah dihitung (Halimah, et al., 2022).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{|S|} \frac{|S_i|}{|S|} * Entropy(S_i)$$

Tabel 3. Hasil Perhitungan Algoritma C4.5 Iterasi 1

Atribut		Jumlah kasus	Ya	Tidak	Entropy	Gain
total kasus	kategori	17	4	13	0,787	
Status Ortu	Yatim	4	3	1	0,811	0,297
	Punya Ortu	13	1	12	0,391	
Penghasilan Ortu	<1 jt	6	3	3	1,000	0,205
	1 - 2.9 jt	6	1	5	0,650	
	3 - 5 jt	3	0	3	0,000	
	> 5 jt	2	0	2	0,000	
Tanggungan Ortu	1	2	0	2	0,000	0,109
	2-3	11	2	9	0,684	
	>3	4	2	2	1,000	
Peringkat	1-3	3	0	3	0,000	0,109
	4-10	6	1	5	0,650	
	>10	8	3	5	0,954	

Pada data ditabel tertulis bahwa nilai gain terbesar berada pada atribut status orang tua siswa, maka atribut status orang tua siswa akan menjadi node akar pertama untuk proses perhitungan selanjutnya.



Gambar 1. Pohon Keputusan Iterasi Pertama

Langkah selanjutnya adalah mengulangi proses perhitungan entropy dan gain sampai terbentuk pohon keputusan akhir.

Algoritma K-Nearest Neighbor

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya Euclidean distance dan manhattan distance (city block distance), yang paling sering digunakan adalah euclidean distance.

Sebelum masuk ke proses perhitungan langkah pertama, pada algoritma KNN jika data nya bukan numerik, maka perlu dibuat pembobotan terlebih dahulu, berikut data pembobotan dari setiap atribut yang digunakan.

Tabel 4. Pembobotan nilai kNN

Atribut	kategori	bobot
Status ortu	Punya ortu	0,5
	Yatim	1
Penghasilan	< 1.000.000	1
	1.000.000 - 2.999.999	0,75
	3.000.000 - 5.000.000	0,5
	> 5.000.000	0,25
Tanggungan ortu	1	0,5
	2-3	0,75
	>3	1
Peringkat kelas	1-3	1
	4-10	0,75
	>10	0,5

Langkah pertama yang harus dilakukan adalah menghitung jarak pada tiap atribut dengan data yang ada pada kasus baru yang akan diuji. Jadikan data yang ada pada kasus baru, menjadi nilai bobot terlebih dahulu dan data yang digunakan untuk melakukan proses perhitungan adalah data pada tabel yang sudah digunakan untuk proses perhitungan algoritma sebelumnya dengan menentukan K=3.

Perhitungan jarak data uji dengan data ke 1

$$= \sqrt{(0,5 - 0,5)^2 + (0,75 - 1)^2 + (1 - 1)^2 + (0,5 - 0,75)^2} = 0,353$$

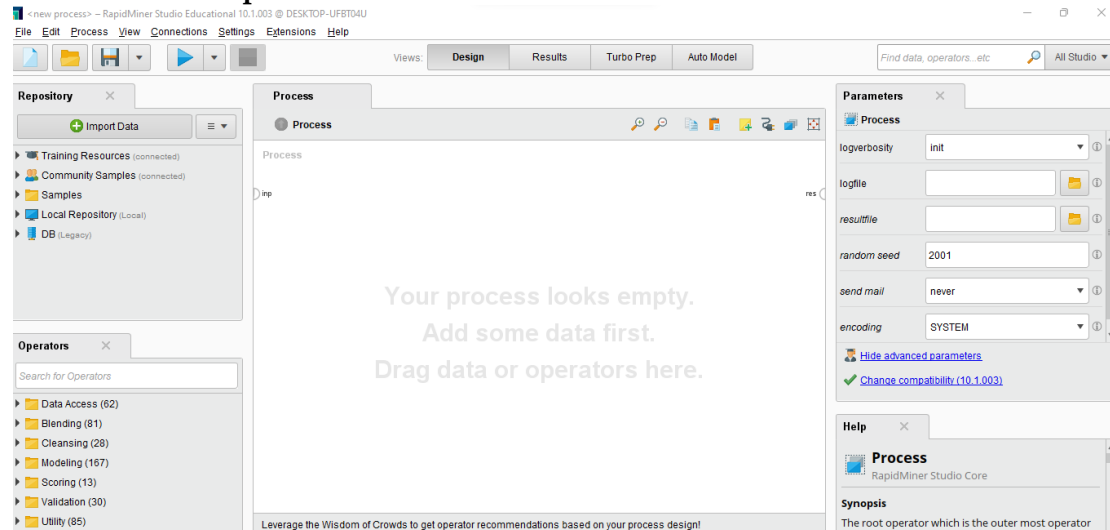
Ulangi proses perhitungan hingga data ke-17.

Lakukan proses perhitungan dengan rumus *Euclidean* jarak, sehingga berdasarkan hasil perhitungan yang telah dilakukan dengan rumus tersebut, maka bisa dibuat data perhitungan ke dalam table berikut:

Tabel 5. Hasil Perhitungan Jarak kNN

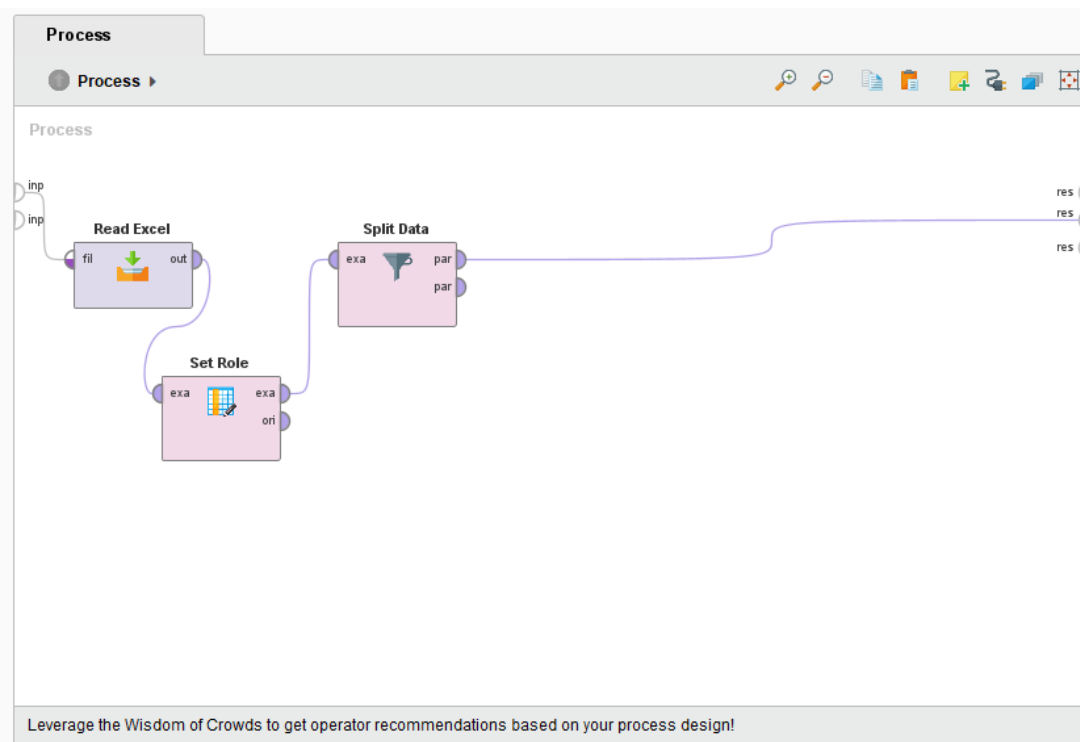
Date ke	Jarak Terdekat	Urutan	Keputusan
1	0,353553391	2	Ya
2	0,707106781	14	Tidak
3	0,75	15	Tidak
4	0,790569415	17	Tidak
5	0,612372436	11	Tidak
6	0,433012702	5	Tidak
7	0,353553391	2	Tidak
8	0,612372436	11	Ya
9	0,612372436	11	Ya
10	0,353553391	2	Tidak
11	0,433012702	5	Tidak
12	0,75	15	Tidak
13	0,433012702	5	Tidak
14	0,559016994	10	Tidak
15	0,5	9	Ya
16	0,25	1	Tidak
17	0,433012702	5	Tidak

Hasil Analisis Rapid Miner



Gambar 2. Halaman Awal Aplikasi Rapid Miner

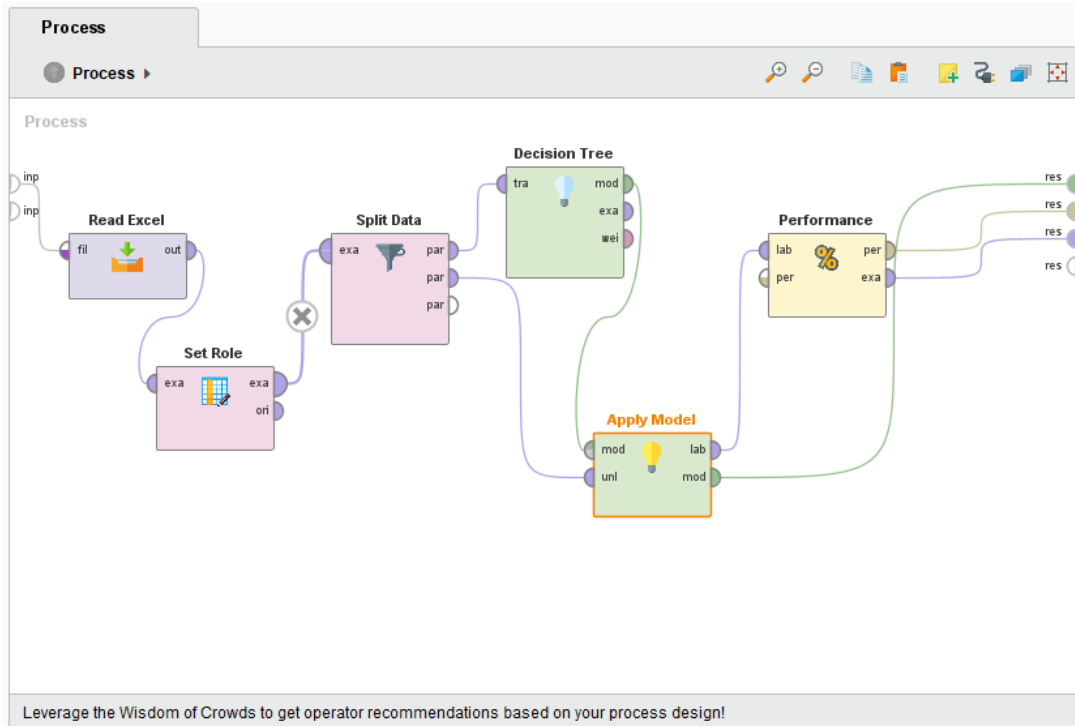
Tampilan halaman awal aplikasi rapid miner yang digunakan untuk melakukan proses analisis algoritma data mining. RapidMiner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Platform dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran (Prasetyo et al., 2021)



Gambar 3. Halaman Menu Split Data Rapid Miner

Untuk melakukan pengujian, dataset yang di awal dibagi menjadi 2 bagian yaitu data training dan data testing dimana data training sebesar 70% dan data testing sebesar 30% dan berikut tampilan ratio split level di rapid miner. Dimana pembagian data menjadi 2 bagian tersebut diambil dari jumlah total siswa SMK yang akan diuji dan diproses menggunakan ketiga algoritma yang digunakan, data siswa yang dipakai untuk pengujian mining berjumlah 56 siswa yang semuanya diambil dari seluruh angkatan

yaitu kelas 10 hingga kelas 12. Dan pembagian split data menjadi 70% berbanding 30% yaitu membagi data training sebesar 39 siswa dan data testing sebesar 17 siswa.



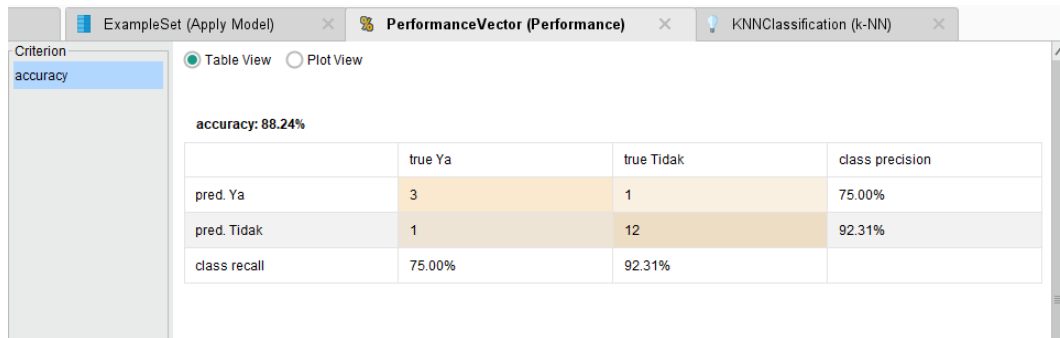
Gambar 4. Proses Klasifikasi Algoritma Di Rapid Miner

Langkah selanjutnya dalam proses data mining algoritma menggunakan aplikasi Rapid Miner adalah dengan menambahkan proses atribut data performance, yang dimana atribut tersebut berguna untuk menentukan dan menghasilkan nilai akurasi yang nantinya akan menjadi hasil utama pada penelitian yang dilakukan. Untuk melakukan proses dalam mendapatkan nilai akurasi, precision dan recall nantinya, maka terlebih dahulu harus dilakukan koneksi atau menghubungkan atribut model algoritma dengan atribut apply model, lalu setelah itu dari atribut apply model ini yang akan dihubungkan ke atribut performance klasifikasi yang berguna untuk mengklasifikasi data penerimaan beasiswa.

	true Ya	true Tidak	class precision
pred. Ya	3	2	60.00%
pred. Tidak	1	11	91.67%
class recall	75.00%	84.62%	

Gambar 5. Confussion Matrix Algoritma C4.5

Dari hasil perhitungan confusion matrix untuk algoritma C4.5 yang telah dilakukan maka hasilnya dari 17 data siswa yang merupakan data testing bisa dijelaskan sebagai berikut yaitu nilai TP = 3 yang didapat dari hasil prediksi benar ya pada siswa benar yang mendapatkan beasiswa, nilai TN = 11 didapat dari hasil prediksi tidak pada data yang bernilai tidak, dilanjutkan nilai FP = 2 yaitu didapat dari hasil prediksi ya pada data bernilai tidak. Kemudian yang terakhir ada nilai FN = 1 yaitu merupakan hasil prediksi tidak pada data bernilai ya.



Gambar 6. Confussion matrix algoritma kNN

Dari hasil perhitungan confussion matrix untuk algoritma naïve bayes yang telah dilakukan maka hasilnya dari 17 data siswa yang merupakan data testing bisa dijelaskan sebagai berikut yaitu nilai TP = 3 yang didapat dari hasil prediksi benar ya pada siswa benar yang mendapatkan beasiswa, nilai TN = 12 didapat dari hasil prediksi tidak pada data yang bernilai tidak, dilanjutkan nilai FP = 1 yaitu didapat dari hasil prediksi ya pada data bernilai tidak. Kemudian yang terkahir ada nilai FN = 1 yaitu merupakan hasil prediksi tidak pada data bernilai ya.

Pembahasan

Percobaan dilakukan untuk mengetahui tingkat akurasi algoritma mana yang paling tinggi nilai akurasi nya. Berdasarkan hasil percobaan yang sudah dilakukan diatas, hasil menunjukkan bahwa algoritma C4.5 mendapatkan nilai pengujian akurasi sebesar 82.35% sedangkan untuk algoritma k-Nearest Neighbor mendapatkan nilai 88.24% dari ketiga pengujian algoritma tadi semua nya mendapatkan nilai keakuratan prediksi yang sudah cukup baik.

Tabel 6. Hasil Perbandingan Nilai Akurasi

Algoritma	Akurasi
C4.5	82.35%
KNN	88.24%

Berdasarkan tabel perbandingan nilai akurasi tersebut dapat diketahui bahwa algoritma yang memiliki nilai akurasi paling tinggi pada proses mining data menggunakan aplikasi data mining rapid miner adalah algoritma k-nearest neighbor yaitu sebesar 88.24% yang hasil di dapat dari perhitungan nilai akurasi berikut:

Yaitu nilai akurasi merupakan $(TP + TN) / (TP + TN + FP + FN) = (3+12) / (3+12+1+1) = 15 / 17 = 88.24%$ dimana nilai ini lebih tinggi dibandingkan nilai akurasi dari algoritma C4.5 yang nilai akurasi nya sebagai berikut:

Yaitu nilai akurasi algoritma C4.5 = $(TP + TN) / (TP + TN + FP + FN) = (3+11) / (3+11+2+1) = 14 / 17 = 82.35%$.

Tabel 8. Hasil Perbandingan Nilai Precision

Algoritma	Precision
C4.5	60.00%
KNN	75.00%

Berdasarkan tabel perbandingan nilai precision tersebut dapat diketahui bahwa algoritma yang memiliki nilai precision yang paling tinggi pada proses mining data menggunakan aplikasi data mining rapid miner adalah algoritma k-nearest neighbor yaitu sebesar 75.00% yang hasil di dapat dari perhitungan nilai precision berikut:

Yaitu nilai precision merupakan nilai $(TP) / (TP + FP) = (3) / (3+1) = 3 / 4 = 75.00\%$ dimana nilai ini lebih tinggi dibandingkan nilai precision dari algoritma C4.5 yang nilai precision nya sebagai berikut:

Yaitu nilai precision algoritma C4.5 = $(TP) / (TP + FP) = (3) / (3+2) = 3 / 5 = 60.00\%$.

Tabel 9. Hasil perbandingan nilai recall

Algoritma	Recall
C4.5	75.00%
KNN	75.00%

Berdasarkan tabel perbandingan nilai recall tersebut dapat diketahui bahwa algoritma yang memiliki nilai recall yang paling tinggi pada proses mining data menggunakan aplikasi data mining rapid miner adalah algoritma k-nearest neighbor dan algoritma C4.5 yaitu sebesar 75.00% yang hasil ini di dapat dari perhitungan nilai recall berikut:

Yaitu nilai recall algoritma k-nearest neighbor merupakan nilai $(TP) / (TP + FN) = (3) / (3+1) = 3 / 4 = 75.00\%$ dimana nilai ini sama jika dibandingkan nilai recall dari algoritma C4.5 yang nilai recall nya sebagai berikut:

Yaitu nilai recall algoritma C4.5 = $(TP) / (TP + FN) = (3) / (3+1) = 3 / 4 = 75.00\%$.

Sehingga berdasarkan hasil perhitungan nilai akurasi, precision dan recall pada kedua algoritma tersebut maka dapat dibuatlah table perbandingan nilai algoritma sebagai berikut:

Tabel 10. Perbandingan akurasi, precision dan recall

Algoritma	Akurasi	Precision	Recall
C4.5	82.35%	60.00%	75.00%
KNN	88.24%	75.00%	75.00%

Berdasarkan table berikut maka bisa dilihat bahwa algoritma k-nearest neighbor memiliki nilai akurasi tertinggi dibandingkan dengan algoritma yang lain, yaitu sebesar 88.24% sedangkan algoritma yang lain nya, yaitu algoritma C4.5 mendapatkan nilai akurasi sebesar 82.35%, begitupun dengan nilai precision yang mana dari table diatas nilai precision pada algoritma k-nearest neighbor lebih tinggi dibandingkan algoritma lainnya dengan nilai precision sebesar 75% dibandingkan dengan algoritma C4.5 untuk nilai precision algoritma C4.5 dengan nilai 60.00%. dan yang terakhir ialah nilai recall dimana pada nilai recall ini ada 2 algoritma yang memiliki nilai paling sama tinggi, yaitu algoritma C4.5 dan algoritma k-nearest neighbor dengan keduanya memiliki nilai recall sebesar 75.00%.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan penulis terhadap klasifikasi penerimaan beasiswa di SMK IT Indra Bangsa, maka dapat disimpulkan bahwa:

1. Berdasarkan hasil perbandingan nilai algoritma C4.5 dan KNN pada pengklasifikasian penerimaan beasiswa dengan pengujian menggunakan confusion matrix yang dimana hasil dari tabel confusion matrix tersebut didapat dari proses analisis perhitungan pohon keputusan di algoritma C4.5, lalu analisis proses perhitungan nilai proses penentuan jarak terdekat untuk algoritma k-Nearest Neighbor, sehingga dari proses perhitungan tersebut lah yang nantinya menjadi bentuk nilai confusion matrix pada tiap kasus dari algoritma data mining.
2. Berdasarkan perhitungan algoritma data mining yang didapatkan dari nilai akurasi algoritma K-Nearest Neighbor mendapat nilai akurasi 88.24 %, algoritma C4.5 mendapatkan tingkat akurasi atau nilai akurasi 82.35%. Meskipun begitu algoritma C4.5 mendapatkan persentasi nilai precision dan recall yang sebagai berikut, yaitu algoritma C4.5 mendapatkan nilai precision sebesar 60.00% dan recall sebesar

75.00%. Berdasarkan hasil nilai akurasi, precision dan recall ini sehingga dapat ditentukan bahwa algoritma yang lebih baik dalam klasifikasi beasiswa pada penelitian ini adalah algoritma K-Nearest Neighbor, lalu setelahnya adalah algoritma C4.5.

DAFTAR PUSTAKA

- Apdian, D., M. T. B. Hutabarat, R. Jayawiguna, and Y. Suherman, (2024) "Sistem Penunjang Keputusan Beasiswa Pada Smk Ristek Karawang Berbasis Web Menggunakan Metode Smart," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 18, no. 4, pp. 17-24, Jan. 2024, doi: 10.35969/interkom.v18i4.320.
- Halimah, D., M. Ridwan, L. Stikom, T. Bangsa, and W. Saputra (2022). "Algoritma C4.5 Untuk Menentukan Klasifikasi Tingkat Pemahaman Mahasiswa Pada Matakuliah Bahasa Pemrograman," *Jurnal Teknik Mesin, Industri, Elektro Dan Informatika (JTMEI)*, vol. 1, no. 3, 2022.
- Josua Josen A. Limbong, Irwan Sembiring, and Kristoko Dwi Hartomo (2022, "Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes dan K-Nearest Neighbor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, pp. 347-356, 2022, doi: 10.25126/jtiik.202294960.
- Moerdyanto, O. P., I. Kadek, and D. Nuryana, (2023). "Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma Decision Tree," *Journal of Informatics and Computer Science*, vol. 05, 2023
- Novika, T., P. Poningsih, H. Okprana, A. P. Windarto, and H. Siahaan (2021), "Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, p. 9, Jan. 2021, doi: 10.30865/mib.v5i1.2498.
- Prasetyo, V. R., H. Lazuardi, A. A. Mulyono, and C. Lauw (2021), "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, pp. 8-17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- Ryanwar, (2020). "Penerapan Metode Algoritma C4.5 Untuk Memprediksi Loyalitas Karyawan Pada Pt.Xyz Berbasis Web," *Fakultas Sains Dan Teknologi Universitas Buddhi Dharma*, p. 103, 2020.
- Safitri, N. A. (2020). "Tinjauan Pustaka Tinjauan Pustaka," *Convention Center Di Kota Tegal*, no. 938, pp. 6-37, 2020.
- Sari, Widia, P. G. Giwangga, N. Dian Fitriyanti, A. Arlisdia Putra, and H. Irawan (2024), "Implementasi Algoritma C4.5 Untuk Klasifikasi Data Insentif Karyawan PT. Adhi Cakra Utama Mulia," *Jurnal Komputer Antartika*, vol. 2, pp. 95-101, 2024, [Online]. Available: <https://ejournal.mediaantartika.id/index.php/jka>