ANALISIS ALGORITMA C 4.5, NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK MENENTUKAN PENERIMAAN BEASISWA

Muhammad Fauzi Firdaus[1], Tukiyat[2]

[1]Teknik Informatika, Universitas Pamulang, Jl. Raya Puspitek, Indonesia, 15310
e-mail: [1]ffauzi889@gmail.com

Abstract

Scholarships are one of the solutions to overcome the problem of costs for those who are less fortunate. Granting scholarships at Indra Bangsa Vocational School sometimes only focuses on orphaned students and scholarships that have good grades so they don't get scholarships, so this study aims to determine the classification of scholarship recipients at Indra Bangsa Vocational School, for the data used as research units are SMK students from class 10-12 with a total of 56 students, of which 30% will be taken as data testing. With this data mining algorithm, the data attributes used for processing include data on family status, number of dependents of parents, student rankings and parents' income. And based on the data mining process it was found that the Naïve Bayes and K-Nearest Neighbor algorithms are more accurate in eliminating the C4.5 algorithm in classifying scholarship recipients. And based on the results of scholarship testing with the rapid miner application, the accuracy value is obtained that the KNN algorithm gets a higher accuracy value of 88.24% while the C4.5 and Naïve Bayes algorithms get an accuracy value of 82.35%. In this case, it can be seen that the best method is the K-Nearest Neighbor algorithm, then the C4.5 and Naïve Bayes algorithms. The suggestion for this research is to use more attributes and use other algorithms for the testing process. The update from the research made is to use 3 algorithms in the data mining process carried out at the Indra Bangsa Vocational School..

***Keywords** : Scholarships, Data Mining, C4.5 Algorithm, Naïve Bayes, K-Nearest Neighbor.*

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 beasiswa 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 nanti nya 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 Naïve Bayes dan K-Nearest Neighbor lebih akurat disbanding 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 dan naïve bayes mendapatkan nilai akurasi 82.35%. Dalam hal ini bisa diketahui untuk pemakaian metode yang terbaik adalah algoritma K-Nearest Neighbor, kemudian algoritma C4.5 dan Naïve Bayes. 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 3 algoritma pada proses mining data yang dilakukan di sekolah SMK Indra Bangsa.
Kata kunci: Beasiswa, Data Mining, Algoritma C4.5, Naïve Bayes, K-Nearest Neighbor.

1. 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.[1]

Beasiswa Bantuan, merupakan jenis beasiswa yang paling banyak dbutuhkan pagi 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.[2]

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 algortima C4.5, algoritma naïve bayes 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 yatu 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.

2. METODE

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. [3]

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.[4]

2.1 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

2.2 Metode Pengumpulan Data

Adapun teknik pengumpulan data yang penulis gunakan sebagai berikut :

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

2.3 Perancangan Penelitian

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

Tabel I. 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

2.4 Waktu dan Tempat Penelitian

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

1. Visi SMK Indra Bangsa
   "Mewujudkan Insan Yang Unggul, Mandiri, Terampil Dalam Ilmu Teknologi Dan Berakhlakul Karimah".
2. Misi SMK Indra Bangsa
   a. Menjalankan Nilai-nilai Agama, Pancasila, Sosial Budaya, Karakter Bangsa Yang Berakhkul Karimah.
   b. Mencetak Siswa Yang Mandiri, Beriman dan Bertaqwa Sebagai Bekal Untuk Masa Depan.
   c. Mengembangkan Ilmu Pengetahuan dan Teknolgi Yang Unggul Sebagai Dasar Untuk Melanjutkan Ke Perguruan Tinggi, Dunia Kerja dan Dunia Usaha.
   d. Menguasai Bahasa Asing (Bahasa Inggris, Bahasa Arab) dan Unggul Dalam Berkompetisi di Era Globalisasi.
   e. Mengembangkan Mutu Dunia Pendidikan Dengan Kerjasama Yang Baik Antara Orang Tua, Stakeholder dan Instansi Terkait Serta Mewujudkan Sekolah Pilihan Masyarakat.

3. HASIL

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 II. 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	Yudamaulana arifin	Punya orang tua	1.000.000 - 2.999.999	2 - 3	1 - 3	Tidak

3.1 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.[5]

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.[6] 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.

Adapun rumus untuk menghitung nilai Entropy sebagai berikut :

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

Langkah selanjut nya adalah mencari nilai Gain tertinggi dari setiap nilai entropy dari tiap atribut yang sudah dihitung.

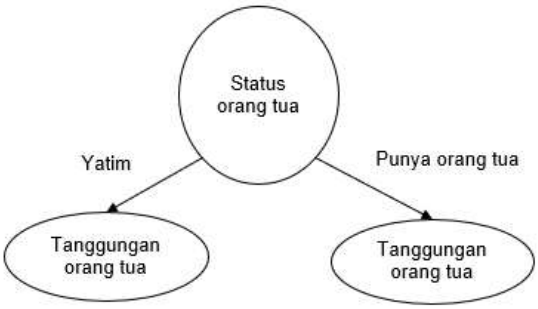
$Gain(S, A) = Entropy(S) - \sum = 1 \ |Si| \ |S| * Entropy(Si)$

Tabel III. 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.



Gbr 1. Pohon keputusan iterasi pertama

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

3.2 Algoritma Naïve Bayes

Pada tahapan berikut ini dilakukan proses data mining dengan menggunakan algoritma naïve bayes. Bayes merupakan Teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema bayes dengan asumsi independensi yang kuat [7]. Probabilitas berasal dari kata Probability dalam Bahasa Inggris yang berarti kemungkinan atau peluang sebuah kejadian akan terjadi.[8]

Data contoh akan dijadikan sebagai data perhitungan untuk proses klafisikasi dalam penentuan penerimaan beasiswa menggunakan algoritma naïve bayes. Proses perhitungan algoritma naïve bayes dimulai dengan menghitung nilai probabilitas dari setiap atribut yang ada. 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”.

Tabel IV. Hasil probabilitas algoritma naïve bayes

Atribut	Kategori	jumlah ya	jumlah tidak	Nilai ya	Nilai Tidak
Status ortu	Punya ortu	1	12	0,25	0,92
	Yatim	3	1	0,75	0,08
Penghasilan	< 1.000.000	3	3	0,75	0,23
	1.000.000 - 2.999.999	1	5	0,25	0,38
	3.000.000 - 5.000.000	0	3	0	0,23
	> 5.000.000	0	2	0	0,15
Tanggungan ortu	1	0	2	0	0,15
	2-3	2	9	0,5	0,69
	>3	2	2	0,5	0,15
Peringkat kelas	1-3	0	3	0	0,23
	4-10	1	5	0,25	0,38
	>10	3	5	0,75	0,38

3.3 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.[9]

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 V. 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 table yang sudah digunakan untuk proses perhitungan algoritma sebelumnya dengan menentukan K=3.

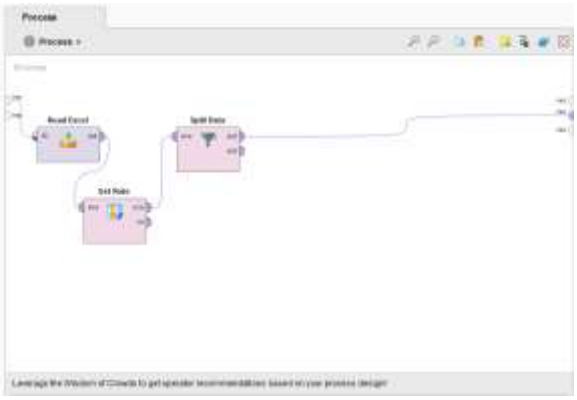
Tabel VI. Hasil perhitungan jarak kNN

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

3.4 Hasil Analisis Rapid Miner

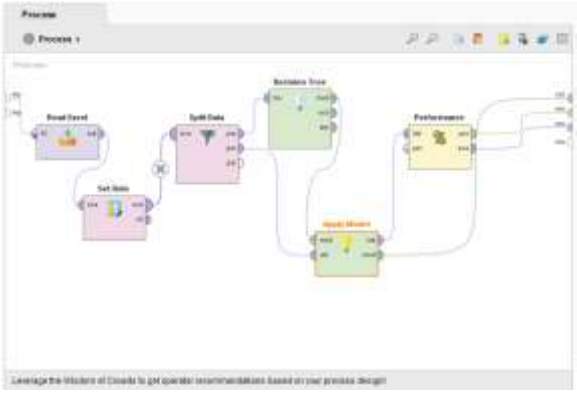
Gbr 2. Halaman awal aplikasi rapid miner

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. Platfrom dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran.[10]



Gbr 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 training 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 semua nya diambil dari seluruh angkatan yaitu kelas 10 hingga kelas 12. Dan pembagian split data menjadi 70% berbanding 30% yaitu membagi data training sebesar 42 siswa dan data training sebesar 17 siswa.



Gbr 3. 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 nanti nya 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.

Gbr 4. Confussion matrix algoritma C4.5

Dari hasil perhitungan confussion 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 terkahir ada nilai FN = 1 yaitu merupakan hasil prediksi tidak pada data bernilai ya.

Gbr 5. Confussion matrix algoritma Naïve Bayes

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 = 2 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 = 2 yaitu merupakan hasil prediksi tidak pada data bernilai ya.

Gbr 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.

4. 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 dan naïve bayes 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 VII. Hasil perbandingan nilai akurasi

Algoritma	Akurasi
C4.5	82.35%
Naïve Bayes	82.35%
KNN	88.24%

Berdasarkan table 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% dan nilai akurasi ini sama dengan nilai akurasi algoritma naïve bayes yang dimana nilai nya adalah sebagai berikut :

Nilai akurasi algoritma naïve bayes = (TP + TN) / (TP + TN + FP + FN) = (2+12) / (2+12+1+2) = 14 / 17 = 82.35%

Tabel VIII. Hasil perbandingan nilai precision

Algoritma	Precision
C4.5	60.00%
Naïve Bayes	66.67%
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% dan nilai precision ini lebih rendah dibandingkan dengan nilai precision algoritma naïve bayes yang dimana nilai nya adalah sebagai berikut :

Nilai precision algoritma naïve bayes = (TP) / (TP + FP) = (2) / (2+1) = 2 / 3 = 66.67%.

Tabel IX. Hasil perbandingan nilai recall

Algoritma	Recall
C4.5	75.00%
Naïve Bayes	50.00%
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% dan nilai recall ini lebih tinggi jika dibandingkan dengan nilai recall algoritma naïve bayes yang dimana nilai nya adalah sebagai berikut :

Nilai recall algoritma naïve bayes = (TP) / (TP + FN) = (2) / (2+2) = 2 / 4 = 50.00%

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

Tabel X. Perbandingan akurasi, precision dan recall

Algoritma	Akurasi	Precision	Recall
C4.5	82.35%	60.00%	75.00%
Naïve Bayes	82.35%	66.67%	50.00%
KNN	88.24%	75.00%	75.00%

Berdasarkan table berikut maka bisa dilihat bahwa algoritma k-nearest neighbor memiliki nilai akurasi tertinggi dibangingkan dengan kedua algoritma yang lain yaitu sebesar 88.24% sedangkan algoritma yaitu algoritma C4.5 dan algoritma naïve bayes kedua sama-sama mendapatkan nilai akurasi sebesar 82.35%, begitupun dengan nilai precision yang mana dari table diatas nilai precision pada algoritma k-nearest neighbor lebih dibangingkan algoritma lainnya dengan nilai precision sebesar 75% kemudian dibawahnya ada algoritma naïve bayes dengan nilai precision 66.67% dan yang paling rendah untuk nilai precision ada 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 tinggi ketimbang algoritma yang lainnya, yaitu algoritma C4.5 dan algoritma k-nearest neighbor dengan keduanya memiliki nilai recall sebesar 75.00% sedangkan dibawahnya atau yang lebih rendah adalah algoritma naïve bayes dengan nilai recall sebesar. 50%.

Bagian pembahasan berisi alasan yang menjelaskan hasil penelitian. Pembahasan adalah perbandingan hasil yang diperoleh dengan konsep/teori yang ada dalam tinjauan pustaka. Tidak diperbolehkan menggunakan kalimat yang sama dengan yang tercantum di bagian hasil dan tidak diperbolehkan membaca ulang tabel dan grafik hasil analisis. Namun, hasil dapat dikelompokkan untuk diinterpretasikan dan dibahas berdasarkan teori dan hasil pengabdian kepada masyarakat terdahulu. Penulisan menggunakan TNR 11 point (tegak) dengan spasi 1 atau single.

5. KESIMPULAN

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, Naïve Bayes dan KNN pada pengklasifikasian penerimaan beasiswa dengan pengujian menggunakan confussion

matrix yang dimana hasil dari tabel confussion matrix tersebut didapat dari proses analisis perhitungan pohon keputusan di algoritma C4.5, lalu analisis proses perhitungan nilai probabilitas untuk algoritma Naïve Bayes dan proses penentuan jarak terdekat untuk algoritma k-Nearest Neighbor, sehingga dari proses perhitungan tersebut lah yang nantinya menjadi bentuk nilai confussion 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% dan algoritma Naïve Bayes mendapatkan nilai akurasi 82.35%. Meskipun sama tetapi algoritma C4.5 dan Naïve Bayes mendapatkan persentasi nilai precision dan recall yang berbeda yaitu algoritma C4.5 mendapatkan nilai precision sebesar 60.00% dan recall sebesar 75.00%, sedangkan algoritma Naïve Bayes mendapatkan nilai precision dan recall sebagai berikut yaitu, precision 66.67% dan recall 50%. 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 dan Naïve Bayes.

6. UCAPAN TERIMAKASIH

Dalam penyusunan ini penulis tidak lepas dari pihak-pihak tertentu yang telah banyak memberikan bantuan bimbingan serta pengarahan, sehingga pada kesempatan ini dengan sebesar-besarnya penulis menyampaikan ucapan terima kasih kepada :

1. Allah SWT, karena oleh berkat kasih dan rahmat karuniaNya penulis bisa ada sebagaimana penulis dapat menyelesaikan tugas akhir tesis.
2. Kedua Orang Tua yang sangat saya cintai, yang telah membantu memberikan semangat, doa, harapan dan dorongan moral kepada Penulis dalam menyelesaikan tugas akhir tesis.
3. Dr. Ir. H. Sarwani, MT., MM, selaku Direktur Pasca sarjana Universitas Pamulang
4. Dr. Ir. Agung Budi Susanto, MM sebagai Ketua Program Studi Magister Teknik Informatika Universitas Pamulang
5. Dr. Ir. Agung Budi Susanto, MM. dan Dr. Tukiyat, M.Si., selaku dosen pembimbing I dan Pembimbing II yang telah menyediakan waktu, tenaga dan pikiran untuk mengarahkan penulisan dan penyusunan tesis ini.
6. Rekan-rekan mahasiswa Program Studi Magister Teknik Informatika Universitas Pamulang yang telah banyak mendukung penulis dalam menyelesaikan tesis ini.
7. Semua Pihak yang terlibat dan tidak penulis sebutkan satu persatu.

7. DAFTAR PUSTAKA

[1] N. A. Safitri, “Tinjauan Pustaka Tinjauan Pustaka,” *Conv. Cent. Di Kota Tegal*, no. 938, pp. 6–37, 2020.

[2] Syarifatul Hilwa, “Pengaruh Pemanfaatan Beasiswa Terhadap Hasil Belajar Siswa di Smk Negeri 4 Jakarta,” *Tarbiya J. Educ. Muslim Soc.*, pp. 1–65, 2016.

[3] Josua Josen A. Limbong, Irwan Sembiring, and Kristoko Dwi Hartomo, “Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes dan K-Nearest Neighbor,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.

[4] Yuli Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.

[5] Ryanwar, “Penerapan Metode Algoritma C4.5 Untuk Memprediksi Loyalitas Karyawan Pada Pt.Xyz Berbasis Web,” *Fak. Sains Dan Teknol. Univ. Buddhi Dharma*, p. 103, 2020.

[6] S. A. Zega, “Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah,” *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, pp. 7–13, 2014.

[7] “68757-ID-sistem-klasifikasi-dan-pencarian-jurnal.pdf.”

[8] Richard, “Modul Probabilitas,” no. 2013150001, pp. 3–18, 2021.

[9] “Tampilan PERBANDINGAN METODE PERHITUNGAN JARAK EUCLIDEAN, HAVERSINE, DAN MANHATTAN DALAM PENENTUAN POSISI KARYAWAN.pdf.”

[10] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, “Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.