

Teknik Data Mining menggunakan Algoritma *Decision Tree* (C4.5) untuk Prediksi Seleksi Beasiswa Jalur KIP pada Universitas Muhammadiyah Kotabumi

Khusnul Khotimah

Universitas Muhammadiyah Kotabumi

khusnul.khotimah@umko.ac.id

ABSTRACT

Smart Indonesia Card Scholarship (KIP) is an educational scholarship provided by the government for equivalent SMA/SMK/MA alumni who have good academic potential and wish to continue their studies to a higher level. KIP lectures themselves are distributed by the government through the Ministry of Education and Culture to universities in their implementation. Universitas Muhammadiyah Kotabumi (UMKO) is one of the universities that opens registration for university admission through KIP scholarships. Based on the data, it is known that there are 210 applicants and from the results of the selection of participants who registered for the KIP selection, there were 204 prospective students who took the test process. So far, in determining the scoring data, the results have not referred to and utilized the previous year's data. While the data can be used as a reference as a source of knowledge data. Therefore, a technique is needed to describe the data more concisely and quickly. Data mining is a technique that can be used to describe a series of processes to obtain knowledge or a pattern from a data set. Decision Tree Algorithm (C4.5) is one of the data mining algorithms that can be used for data classification to help solve classification problems. Based on the results of the research on the implementation of data mining using the decision tree algorithm (C4.5), an accuracy value of 100% was obtained. The accuracy test was also carried out using the Naive Bayes algorithm to obtain a comparison of the levels of accuracy. Based on the accuracy test of the two algorithms, data obtained with an accuracy level of 100% on the Decision Tree (C4.5) algorithm and 90.16% on the Naive bayes algorithm. It can be concluded that the accuracy of the Decision Tree (C4.5) algorithm for predicting prospective students receiving KIP scholarships is better than the Naive Bayes algorithm.

Keywords: *Data Mining; Decision Tree Algorithm (C45); KIP College; Classification*

ABSTRAK

Beasiswa Kartu Indonesia Pintar (KIP) Kuliah merupakan beasiswa Pendidikan yang diberikan pemerintah bagi alumni SMA/SMK/MA sederajat yang memiliki potensi akademik yang baik serta berkeinginan melanjutkan studi ke jenjang yang lebih tinggi. KIP kuliah sendiri pada penyalurannya dilakukan pemerintah melalui kemendikbud ke pihak Perguruan Tinggi dalam implementasinya penyelenggaraannya. Universitas Muhammadiyah kotabumi (UMKO) adalah salah satu perguruan tinggi yang membuka pendaftaran masuk Universitas melalui beasiswa jalur KIP. Berdasarkan data diketahui terdapat 210 pendaftar dan dari hasil seleksi peserta yang mendaftar seleksi KIP terdapat 204 calon mahasiswa yang mengikuti proses tes. Selama ini dalam penentuan data penskoran bobot hasil belum mengacu dan memanfaatkan data tahun sebelumnya. Sedangkan data tersebut dapat dijadikan acuan sebagai sumber data pengetahuan. Oleh karena itu, diperlukan suatu teknik untuk menggambarkan data menjadi lebih ringkas dan cepat. Data mining merupakan Teknik yang dapat digunakan untuk menggambarkan Serangkaian proses mendapatkan pengetahuan atau suatu pola dari kumpulan data. Algoritma *Decision Tree* (C4.5) merupakan salah satu algoritma data mining yang dapat digunakan untuk klasifikasi data membantu menyelesaikan permasalahan klasifikasi. Berdasarkan hasil penelitian implementasi data mining menggunakan algoritma *decision tree* (C4.5) diperoleh nilai akurasi sebesar 100%. Uji akurasi juga dilakukan dengan algoritma *Naive Bayes* untuk memperoleh perbandingan tingkat akurasi. Berdasarkan uji akurasi dua algoritma diperoleh data tingkat akurasi sebesar 100% pada algoritma *Decision Tree* (C4.5) dan 90,16 % pada algoritma *Naive bayes*. Hal ini dapat disimpulkan akurasi algoritma *Decision Tree* (C4.5) untuk prediksi calon mahasiswa penerima beasiswa jalur KIP lebih baik dibandingkan algoritma *Naive bayes*.

Kata Kunci: *Data Mining; Algoritma Decision Tree (C45); KIP Kuliah; Klasifikasi*

1. PENDAHULUAN

Universitas Muhammadiyah Kotabumi (UMKO) adalah Universitas yang berada di wilayah Lampung Utara. Proses PMB dilakukan melalui beberapa jalur seleksi. Salah satu jalur seleksi yang dilakukan yaitu seleksi beasiswa jalur Kartu Indonesia Pintar (KIP) Kuliah. KIP Kuliah merupakan beasiswa dari pemerintah melalui kemendikbud yang diberikan pada perguruan tinggi yang ditujukan bagi alumni SMA/SMK/MA sederajat yang ingin melanjutkan ke jenjang perguruan tinggi namun kurang mampu secara ekonomi. Alasan UMKO berperan serta membantu pemerintah dalam melakukan seleksi PMB melalui jalur beasiswa KIP tidak lain agar alumni SMA/SMK/MA dengan latar belakang kurang mampu secara ekonomi mendapat kesempatan melanjutkan pendidikannya dengan mencoba mengikuti seleksi ini. Beberapa tahapan seleksi yang harus dilalui diantaranya proses seleksi berkas, seleksi tes potensi akademik (TPA) berbasis CBT (*Computer Based Test*), dan seleksi tes wawancara/*interview*. Ketiga tes tersebut wajib di ikuti seluruh peserta KIP Kuliah.

Adapun metode yang dilakukan oleh panitia penyelenggara proses PMB melalui beasiswa jalur KIP Kuliah dilakukan dengan metode pemberian bobot skor pada atribut/kriteria yang diujikan. Setiap atribut/kriteria yang menjadi syarat masing-masing akan di konversikan dalam bentuk skor nilai yang telah ditentukan sebelumnya oleh Tim panitia penyelenggara dan hasil akhir akan di hitung, di analisis dan di plenokan. Selama ini pada proses penentuan bobot nilai calon mahasiswa penerima beasiswa jalur KIP Kuliah belum memanfaatkan data dari tahun sebelumnya sebagai acuan penskoran nilai. Penggunaan penskoran yang memanfaatkan data di tahun sebelumnya sangat penting karena dapat menjadi acuan sebagai alternatif penentuan dan proses seleksi calon mahasiswa melalui beasiswa jalur KIP Kuliah di tahun berikutnya.

Metode klasifikasi dapat digunakan dalam proses seleksi calon mahasiswa penerima beasiswa jalur KIP Kuliah, Klasifikasi merupakan metode implementasi data mining yang dapat dimanfaatkan untuk mengelompokkan data dalam suatu kelas-kelas yang telah ditentukan sebelumnya. Dalam metode klasifikasi dibagi dalam 2 tahap, yaitu tahap pembelajaran dan tahap klasifikasi. (Han, J., Kamber, M., & Pei, J., 2012). Pada tahap pembelajaran data peserta/pendaftar di tahun sebelumnya dapat digunakan untuk membangun suatu model klasifikasi. Kemudian pada tahap klasifikasi data yang digunakan merupakan data peserta/pendaftar di tahun yang sedang berjalan yang dikelompokkan pada kelas layak atau tidak layak. Untuk membangun model klasifikasi terdapat beberapa algoritma data mining yang dapat digunakan, diantaranya adalah algoritma *Decission Tree (C4.5)*, *K-Nearest Neighbour (k-NN)*, *Support Vektor Machine (SVM)*, serta *Naïve Bayes (NB)*(Prasetyo 2014). Pada penelitian ini yang akan digunakan adalah algoritma *Decission Tree (C4.5)* dengan asumsi algoritma ini memiliki tingkat akurasi lebih baik dari beberapa algoritma yang lain.

Pada kasus data calon mahasiswa peserta seleksi beasiswa jalur KIP Kuliah, jumlah pendaftar yang dinyatakan layak lebih sedikit daripada jumlah pendaftar yang belum layak. Hal ini bergantung pada jumlah target peserta yang di tetapkan. Tingkat pendaftar sangat tinggi, sedangkan berdasarkan data hasil tes jumlah pendaftar yang layak dan belum layak tidakimbang. Untuk mengatasi kasus *class imbalance problem* ini dapat digunakan teknik sampling yang sesuai sehingga model hasil training dapat mendeteksi kelas sebuah data dengan akurat. Pada penelitian ini untuk mengatasi *class imbalance problem* digunakan teknik sampling untuk membangkitkan data training.

2. KERANGKA TEORI

2.1 Data Mining

Data mining merupakan teknik proses pembelajaran komputer dalam menganalisis dan melakukan ekstraksi pengetahuan yang secara otomatis. Serangkaian proses untuk menggali nilai tambah dari kumpulan berbagai data menjadi suatu pengetahuan yang selama ini belum. Salah satu algoritma yang terdapat dalam klasifikasi data mining adalah algoritma *C4.5*, yang khusus diterapkan pada Teknik pohon keputusan (*decision tree*) (Azwanti, 2018). Menurut Hermawati (2013) Operasi *data mining* menurut sifatnya dibedakan menjadi dua, yaitu:

1. Prediksi (*prediction driven*). Operasi prediksi digunakan untuk validasi hipotesis, *querying* dan pelaporan (misalnya: spreadsheet dan pivot tabel), analisis multidimensi (*dimensional summary*), OLAP (*Online Analytic Processing*) serta analisis statistik.
2. Penemuan (*discovery driven*) bersifat transparan. Operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi *database*, analisis keterkaitan (*link analysis*) dan deteksi deviasi. (Purwati & Nurlistiani, 2020).

2.2. Klasifikasi data

Menurut Witten et al (2011) klasifikasi adalah salah satu metode yang sering digunakan di data mining. Klasifikasi termasuk kedalam *supervised learning* karena dalam proses klasifikasi terdapat proses pembelajaran dengan data yang telah lampau. Klasifikasi merupakan teknik mining yang digunakan dalam mengelompokkan sesuatu yang didasarkan pada suatu kategori. Klasifikasi di sini berarti suatu data yang sudah digolongkan, misal buku referensi, buku chapter, bahan ajar. (Khotimah, 2020)

2.3 Algoritma Decision Tree (C4.5)

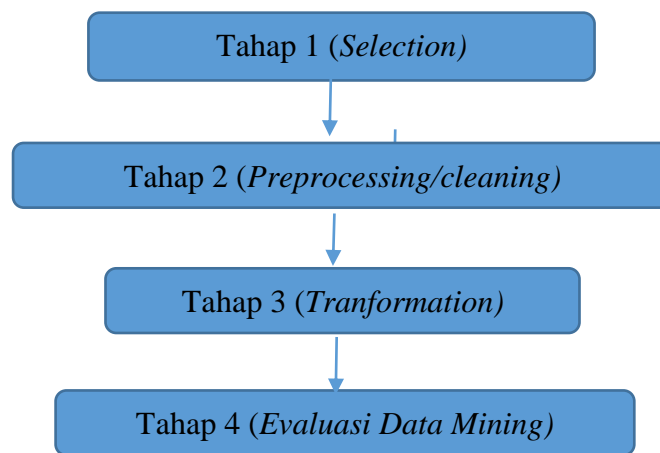
Algoritma C.4.5 merupakan metode probabilistik pengklasifikasian *Decision Tree* berguna untuk mengeksploitasi data dan akan menemukan hubungan yang tersembunyi antara jumlah variabel input dengan variabel target (Kusrini & Luthfi 2009)

2.4 KIP Kartu Indonesia Pintar

KIP Kuliah ini adalah program bantuan pendidikan kepada alumni SMA/SMK/MA sederajat yang memiliki potensi akademik yang baik namun dari keluarga kurang mampu/keterbatasan ekonomi agar dapat melanjutkan pendidikan ke jenjang perguruan tinggi/universitas. (Adit, 2020)

3. METODOLOGI

Metodologi pada penelitian yaitu *Knowledge discovery in Database (KDD)*. Adapun proses tahapannya meliputi: Seleksi data (*Selection*), Memilih kelengkapan data dan membersihkan data (*Preprocessing/cleaning*), Transformasi data (*Transformation*), Evaluasi data mining (*Data Mining Evaluation*):



Gambar 1. Bagan Alir Metode Penelitian

- Peneliti mengajukan permohonan data calon peserta PMB jalur KIP pada bagian penyelenggaraan PMB UMKO. Berdasarkan data yang diperoleh, selanjutnya dilakukan proses pemilihan data yang akan digunakan.
- Peneliti melakukan tahapan *preprocessing* data yaitu proses melihat dan cek *missing value* pada dataset
- Melakukan transformasi terhadap data.
- Melakukan pengujian dan menganalisis hasil pengujian pohon keputusan yang di buat menggunakan algoritma Decision Tree (C4.5) dan membandingkan hasil akurasi dengan algoritma *Naïve Bayes*.

4. HASIL PEMBAHASAN

a. Data selection

Pada tahap selection dilakukan proses memilih data yang sesuai dengan data yang diperoleh sehingga tidak semua data digunakan. Data yang digunakan berjumlah 204 dari 210 data pendaftar, pemilihan atribut-atribut didasarkan pada data yang berpengaruh pada proses seleksi beasiswa jalur KIP Kuliah. Data tersebut meliputi data id peserta, program studi, skor nilai berkas, skor nilai tes CBT, dan skor nilai tes wawancara/interview.

No	ID Peserta	Nama Kelamin	Pendid	Wawancara	200%	CDT	40%	Berkas	80%	Total Skor	Keterangan
1	11-C-2021001	Laki-laki	IKL-K3UM	71.25	14.25	27.50	15.00	75.00	70.00	59.25	KKS
2	11-C-2021002	Pemempuan	PGSI	81.25	16.25	47.50	10.00	100.00	40.00	75.25	KIP dan PK
3	11-C-2021003	Pemempuan	PGSD	78.75	15.75	35.00	14.00	100.00	40.00	69.75	PK21, KIP
4	11-C-2021004	Pemempuan	PGSD	70.75	15.75	30.00	4.00	87.50	85.00	54.97	KIP
5	11-C-2021005	Laki-laki	IKL-K3UM	73.00	14.00	30.00	6.00	12.50	5.00	23.60	Tidak ada K
6	11-C-2021006	Laki-laki	IKL-K3UM	76.25	15.25	35.00	14.00	12.50	8.00	54.23	Tidak ada K
7	11-C-2021007	Pemempuan	PGSD	70.00	15.00	40.00	16.00	87.50	25.00	66.80	KKS
8	11-C-2021008	Pemempuan	IKL-K3UM	73.75	13.75	35.00	15.00	75.00	40.00	60.73	KIP dan PK
9	11-C-2021009	Laki-laki	IKL-K3UM	80.00	16.00	42.50	17.00	100.00	40.00	73.00	KIP dan PK
10	11-C-2021010	Laki-laki	IKL-K3UM	75.25	15.00	32.50	15.00	87.50	35.00	65.05	KIP
11	11-C-2021011	Laki-laki	PG-N2AN	82.50	16.50	52.50	21.00	100.00	40.00	77.50	KIP dan PK
12	11-C-2021012	Laki-laki	IKL-K3UM	78.00	15.00	42.50	17.00	75.00	60.00	62.60	KKS
13	11-C-2021013	Laki-laki	IKL-K3UM	77.25	13.75	22.50	11.00	75.00	30.00	56.85	PK21
14	11-C-2021014	Laki-laki	IKL-K3UM	75.75	15.75	35.00	14.00	100.00	40.00	69.75	KIP, KKS
15	11-C-2021015	Pemempuan	PKSI	79.25	15.25	42.50	17.00	87.50	55.00	67.53	KIP
16	11-C-2021016	Laki-laki	IKL-K3UM	75.50	15.50	40.00	16.00	100.00	40.00	71.18	KIP, KKS
17	11-C-2021017	Pemempuan	PGSD	82.38	16.48	55.00	22.00	100.00	40.00	78.48	KIP, PK21

Gambar 2. *Dataset* Rekapitulasi Peserta KIP

b. *Preprocessing/cleaning*

Pada tahap ini yang dilakukan adalah proses pembersihan data calon mahasiswa seleksi beasiswa jalur KIP Kuliah tahun 2021. Tahapan yang dilakukan dengan melihat dan mengecek kembali *missing value* yang terdapat

pada data calon mahasiswa peserta KIP. Kemudian dilakukan pengecekan kesalahan penulisan dan melakukan reduksi data sesuai kebutuhan. Pada data yang ada tidak ditemukan data yang tidak lengkap maka proses penanganan *missing value* tidak dilakukan. Proses lainnya yang dilakukan memperbaiki penamaan atribut data dan mereduksi data agar sesuai dengan data yang dibutuhkan. *Tools* yang digunakan adalah *rapidminer* versi 9.10 untuk melakukan *preprocessing* seperti tampilan berikut:

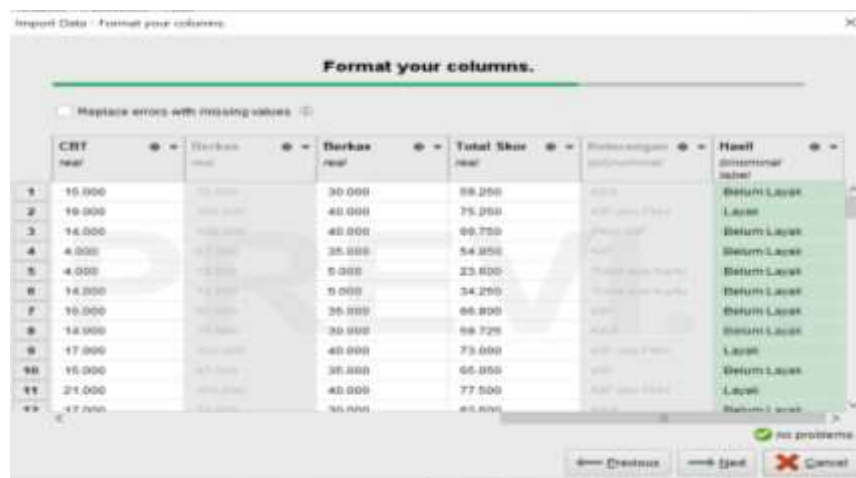


ID Peserta	Jenis Kelamin	Prodi	Interview	CBT
1	Male	Engineering	15.000	30.000
2	Male	Engineering	16.000	40.000
3	Male	Engineering	14.000	40.000
4	Male	Engineering	4.000	35.000
5	Male	Engineering	4.000	5.000
6	Male	Engineering	14.000	5.000
7	Male	Engineering	10.000	35.000
8	Male	Engineering	14.000	30.000
9	Male	Engineering	17.000	40.000
10	Male	Engineering	15.000	30.000
11	Male	Engineering	21.000	40.000

Gambar 3. Tampilan data cek *Missing Value*

c. Data Transformation

Pada tahap tranformasi data dilakukan proses perubahan data sesuai dengan bentuk data yang akan digunakan dalam proses data mining. Beberapa atribut tidak akan digunakan untuk proses data mining karena tidak semua atribut memiliki pengaruh pada proses hasil data mining dengan tujuan agar memudahkan proses klasifikasi yang dilakukan. Beberapa proses tranformasi dilakukan seperti merubah tipe data dari integer menjadi real. Adapun kriteria yang di tranformasi meliputi: nilai skor berkas, nilai skor tes CBT, dan nilai skort tes wawancara/*interview*.



CBT	Interview	Berkas	Total Skor	Hasil
15.000	30.000	45.000	90.000	Belum Lulus
16.000	40.000	56.000	112.000	Lulus
14.000	40.000	54.000	108.000	Belum Lulus
4.000	35.000	39.000	78.000	Belum Lulus
4.000	5.000	9.000	18.000	Belum Lulus
14.000	5.000	19.000	38.000	Belum Lulus
10.000	35.000	45.000	90.000	Belum Lulus
14.000	30.000	44.000	88.000	Belum Lulus
17.000	40.000	57.000	114.000	Lulus
15.000	30.000	45.000	90.000	Belum Lulus
21.000	40.000	59.000	120.000	Lulus

Gambar 4. Tampilan Tranformasi *Dataset*

d. Evaluasi Data Mining

Pada proses data mining dilakukan pemilihan algoritma yang akan digunakan untuk menentukan pola klasifikasi data calon mahasiswa penerima beasiswa jalur KIP UMKO. Algoritma ynag digunakan adalah *Decicion Tree (C4.5)*

sebagai implementasi data mining. Hasil pengolahan algoritma *Decision Tree (C4.5)* selanjutnya akan di uji tingkat akurasi dengan membagi data dalam rasio 8:2. 80% data dijadikan *training* dan 20% data menjadi data *testing*. Berikut ini merupakan hasil pemodelan *Decision Tree (C4.5)*:

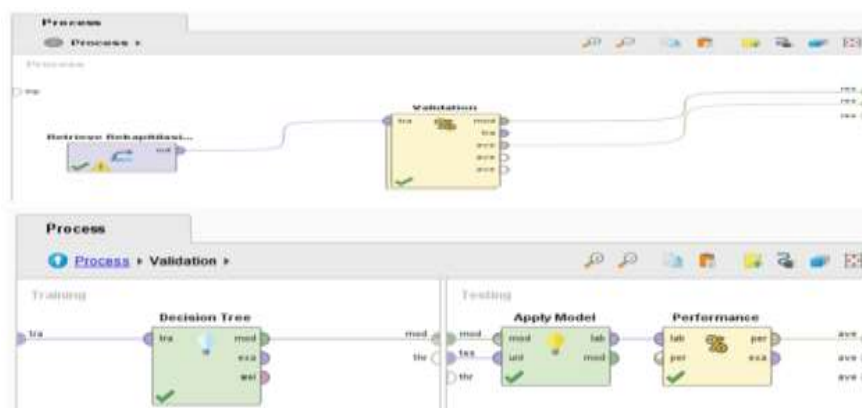


Gambar 5. Tampilan Uji Akurasi Dataset

ID	Nama	Jenis	Status	...
1	Adi	Laki-laki	Belum Lajang	...
2	Budi	Laki-laki	Lajang	...
3	Cici	Laki-laki	Belum Lajang	...
4	Dani	Laki-laki	Lajang	...
5	Evi	Laki-laki	Belum Lajang	...
6	Fani	Laki-laki	Lajang	...
7	Gani	Laki-laki	Belum Lajang	...
8	Hani	Laki-laki	Lajang	...
9	Iani	Laki-laki	Belum Lajang	...
10	Jani	Laki-laki	Lajang	...
11	Kani	Laki-laki	Belum Lajang	...
12	Lani	Laki-laki	Lajang	...
13	Mani	Laki-laki	Belum Lajang	...
14	Nani	Laki-laki	Lajang	...
15	Oani	Laki-laki	Belum Lajang	...
16	Pani	Laki-laki	Lajang	...
17	Qani	Laki-laki	Belum Lajang	...
18	Rani	Laki-laki	Lajang	...
19	Sani	Laki-laki	Belum Lajang	...
20	Tani	Laki-laki	Lajang	...

Gambar 5. Tampilan View Dataset

Selanjutnya dilakukan proses membandingkan tingkat akurasi klasifikasi algoritma *Decision Tree (C4.5)* dengan Algoritma *Naive Bayes*, untuk perhitungan *Decision Tree (C4.5)* adalah sebagai berikut:



Gambar 6. Perhitungan Algoritma *Decision Tree (C4.5)*

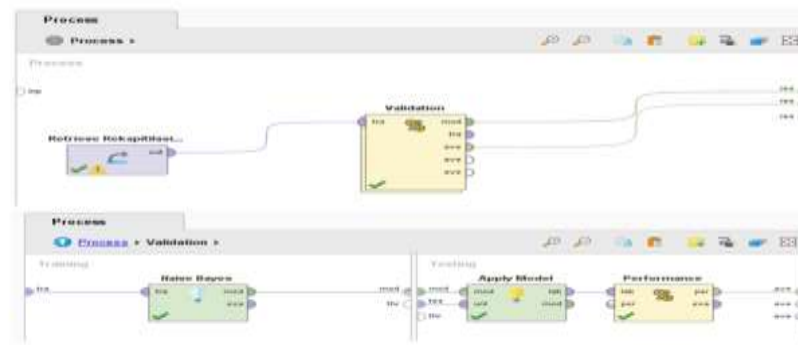
Perhitungan tingkat akurasi menggunakan *confussion matrix* pada *Decision Tree (C4.5)* seperti pada tabel berikut:

Tabel 1. Tampilan Akurasi Algoritma *Decision Tree (C4.5)*

accuracy: 100.00%

	true Belum Lajang	true Lajang	class precision
pred Belum Lajang	32	0	100.00%
pred Lajang	0	9	100.00%
class recall	100.00%	100.00%	

Adapun untuk perhitungan *Naive Bayes* adalah sebagai berikut:



Gambar 7. Perhitungan Algoritma *Naive Bayes*

Untuk perhitungan tingkat akurasi *Naive Bayes* dengan *confussion matrix* pada tabel berikut:

Tabel 2. Tampilan Akurasi Algoritma *Naive Bayes*

accuracy: 90.16%			
	true Belum Layak	true Layak	class precision
pred. Belum Layak	43	1	97.73%
pred. Layak	5	12	70.59%
class recall	89.58%	92.31%	

Pada hasil penelitian diketahui bahwa untuk klasifikasi data seleksi beasiswa jalur KIP calon mahasiswa dengan menggunakan algoritma *Decision Tree (C4.5)* merupakan algoritma terbaik. Algoritma *Decision Tree (C4.5)* tingkat akurasi sebesar 100 %, sedangkan algoritma *Naive Bayes* sebesar 90,16 %.

5. KESIMPULAN

Penelitian ini dapat disimpulkan bahwa analisa dan perhitungan menggunakan algoritma ini sangat membantu proses penentuan keputusan untuk menentukan calon mahasiswa yang layak dan belum layak mendapatkan beasiswa jalur KIP berdasarkan bobot nilai tes yang telah dilalui. Klasifikasi menggunakan Teknik data mining dapat digunakan untuk mempersingkat waktu dalam penyajian data pendukung dalam pengambilan keputusan. Algoritma data mining dapat memberikan gambaran jelas atribut apa saja yang dapat dijadikan prioritas sebagai acuan pengambilan keputusan. Implementasi data mining menggunakan menggunakan algoritma *Decision Tree (C4.5)* dalam menentukan calon mahasiswa yang layak dan belum layak mendapatkan beasiswa melalui jalur KIP Kuliah dapat dijadikan acuan sehingga proses seleksi menjadi tepat sasaran. Berdasarkan uji tingkat akurasi algoritma *Decision Tree (C4.5)* dengan algoritma *Naive Bayes* diperoleh data tingkat akurasi sebesar 100% pada algoritma *Decision Tree (C4.5)* dan 90,16 % pada algoritma *Naive bayes*. Hal ini dapat disimpulkan akurasi algoritma *Decision Tree (C4.5)* untuk prediksi calon mahasiswa penerima beasiswa jalur beasiswa KIP Kuliah lebih baik.

DAFTAR PUSTAKA

- Adit, A. 2020. Apa itu KIP Kuliah? Ini penjelasan dan Fasilitas Pembiayaannya. Website: Retrieved from Kompas.com: <https://edukasi.kompas.com/read/2020/02/19/10543491/apa-itu-kip-kuliah-ini-penjelasan-dan-fasilitas-pembiayaannya?page=all>. Diakses tanggal 15 Oktober 2021
- Azwanti, N., 2018. Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu). Jurnal SIMETRIS(Teknik Industri, Mesin, Elektro dan Ilmu Komputer) 9(1), 11-22. Print ISSN 2252-4983 dan Online ISSN 2549-3108.
- Han, J., Kamber, M., & Pei, J. 2012. Data Mining Concepts and Techniques 3rd Edition (3rd ed.). Morgan Kaufmann
- Hermawati, Fajar Astuti. (2013). Konsep dan Teori Pengolahan Citra Digital, Yogyakarta: Andi Offset
- Khotimah, K., & Husniyati. 2020. Pengelolaan Big Data Perpustakaan dengan Teknik Data Mining. Jurnal SIENNA (Sistem dan Teknologi Informasi) 1(2), 64-70, P-ISSN: 2745-987X dan E-ISSN: 2745-9861.
- Kusrini & Luthfi, E.T., 2009. Algoritma Data Mining Ed. 1. T. A. Prabawati, ed., Yogyakarta: Andi Offset.
- Prasetyo, E. 2014. Data Mining – Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi Offset.
- Purwati N., Nurlistiani R., 2020. *Data Mining Dengan Algoritma Neural Network Dan Visualisasi Data Untuk Prediksi Kelulusan Mahasiswa*. Jurnal Informatika, Vol. 20, No. 2, E-ISSN 2407-1544.