

SISTEM PREDIKSI PRESTASI AKADEMIK SISWA GUNA MENDUKUNG KEPUTUSAN PEMBERIAN BEASISWA MENGUNAKAN ALGORITMA C4.5 DI SMK BINA INFORMATIKA BINTARO

Dimas Agung Amrullah¹, Subandi^{2*}

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹dimasagungamrullah@gmail.com, ^{2*}subandi@budiluhur.ac.id

(* : corresponding author)

Abstrak- Permasalahan yang dihadapi oleh SMK Bina Informatika Bintaro adalah bagaimana mengidentifikasi penerima beasiswa secara adil dan efektif. Proses seleksi yang masih dilakukan secara manual dan subjektif seringkali memakan waktu dan tidak selalu menghasilkan hasil yang adil. Dengan adanya masalah tersebut, pihak sekolah membutuhkan sebuah sistem yang akan membantu dalam pengecekan prestasi akademik siswa dengan hasil data yang akurat. Metode untuk penyelesaian masalah ini adalah dengan penggunaan algoritma C4.5 yang digunakan untuk pembuatan sistem dengan tujuan untuk memberikan prediksi pada prestasi akademik berbasis web. Algoritma ini membangun model pohon keputusan yang memiliki kemampuan untuk memprediksi prestasi akademik siswa berdasarkan data historis. Hasil penelitian menunjukkan bahwa sistem prediksi prestasi akademik yang direalisasikan menggunakan algoritma C4.5 memberikan hasil nilai yang akurat yaitu sebesar 88%, recall sebesar 91,67%, dan spesifisitas dengan hasil 0% penghitungan pada data yang dikelompokkan menjadi dua kelompok yaitu persentase sebesar 80% sebagai data latih dan 20% sebagai data uji. Kemudian hasil pada penghitungan ini menunjukkan bahwa sistem ini memiliki kemampuan untuk memberikan prediksi yang cukup akurat dan dapat diandalkan, meskipun masih ada ruang untuk peningkatan. Dengan adanya sistem ini, proses pengambilan keputusan terkait pemberian beasiswa dapat dilakukan dengan lebih efisien dan objektif sehingga menghasilkan proses seleksi yang lebih adil dan transparan.

Kata Kunci: Prediksi Prestasi Akademik, Algoritma C4.5, Pemberian Beasiswa, SMK Bina Informatika Bintaro

STUDENT ACADEMIC ACHIEVEMENT PREDICTION SYSTEM TO SUPPORT SCHOLARSHIP DECISIONS USING C4.5 ALGORITHM AT BINA INFORMATICA VOCATIONAL HIGH SCHOOL

Abstract- *Bintaro Bina Informatika Vocational High School faces problems in determining scholarship recipients objectively and efficiently. The selection process, which is still carried out manually and subjectively, often takes time and does not always produce fair results. Given this problem, the school needs a system that will help in checking student academic achievement with accurate data results. The method for solving this problem is to use the C4.5 algorithm which is used to create a system with the aim of providing web-based predictions of student academic achievement. This algorithm builds a decision tree model that has the ability to predict student academic achievement based on historical data. The results of the research show that the academic achievement prediction system realized using the C4.5 algorithm provides accurate results of 88%, recall of 91.67%, and specificity with 0% results for calculations on data grouped into two groups, namely a percentage of 80 % as training data and 20% as test data. Then the results of this calculation show that this system has the ability to provide fairly accurate and reliable predictions, although there is still room for improvement. With this system, the decision-making process regarding scholarship awards can be carried out more efficiently and objectively, resulting in a fairer and more transparent selection process.*

Keywords: *Prediction Of Academic Achievement, C4.5 Algorithm, Scholarship Award, Bintaro Bina Informatika Vocational High School*

1. PENDAHULUAN

Pendidikan adalah sebuah hal yang menjadi pandangan untuk menilai kualitas diri seseorang atau sumber daya manusia. Untuk merealisasikan sebuah harapan untuk menciptakan generasi muda yang bermanfaat dibutuhkan sekolah yang mendukung hal tersebut. Beasiswa tidak hanya merupakan penghargaan bagi siswa berprestasi namun juga merupakan bentuk dukungan kepada siswa yang membutuhkan dukungan finansial.

SMK Bina Informatika Bintaro adalah lembaga pendidikan yang berkomitmen untuk menghasilkan siswa yang siap bersaing di dunia kerja. Mereka memahami pentingnya memberikan beasiswa kepada siswanya untuk meningkatkan prestasi akademik mereka. Namun, seringkali ada masalah dalam proses menentukan siapa yang menerima beasiswa secara efisien dan efektif. Proses seleksi yang tidak terstruktur dan subjektif dapat menyebabkan ketidakadilan dan menurunkan motivasi siswa.

Dengan berkembangnya teknologi informasi, algoritma C4.5 dianggap sebagai sebuah algoritma yang dapat digunakan sebagai pohon keputusan paling kuat untuk melakukan klasifikasi dengan menggunakan data historis sekolah untuk memprediksi kinerja akademik siswa. Keputusan pemberian beasiswa dapat diambil secara objektif dan berdasarkan data.

Data merupakan bahan mentah yang harus dikelola untuk menghasilkan informasi baik dengan cara kualitatif maupun kuantitatif, atau informasi yang menunjukkan fakta yang akurat sehingga dapat menjadi manfaat untuk peneliti atau memberikan wawasan terhadap suatu kondisi maupun situasi[1].

Data *mining* merupakan sebuah usaha untuk mengumpulkan data yang kemudian akan digunakan sebagai pola untuk mengetahui hubungan antara Kumpulan data secara historis yang diharapkan dapat memberikan hasil untuk digunakan sebagai jawaban untuk memutuskan sebuah jawaban dari permasalahan yang ada dan yang akan datang[2]. Data *Mining* adalah proses berulang dimaksudkan untuk analisis *database* yang bertujuan untuk mencari keakuratan sebuah data dan berpotensi untuk membantu para peneliti untuk pengambilan data hingga pemecahan masalah. Unggul dan berguna dalam definisi paling sederhana, data *mining* bersifat otomatis mengenali pola terkait dalam *database*[3]. Data *Mining* merupakan sekumpulan teknik untuk menemukan pola secara otomatis yang efisien yang bersifat baru, valid, dan dan berguna untuk digunakan dalam database besar. Pola yang dimaksud harus memiliki keberlanjutan yang kemudian memudahkan untuk proses pengambilan keputusan[4]. Data *mining* memiliki beberapa fungsi sebagai berikut, *clustering, classification, association, regression, forecasting, squence analysis, deviation analysis*[5].

Knowledge Discovery in Database atau yang biasa disebut dengan KDD adalah sebuah implementasi proses metode ilmiah dalam kegiatan data *mining*[6]. Klasifikasi adalah proses pengajaran dan pelatihan fungsi tujuan dengan menugaskan setiap kumpulan karakteristik (atribut) ke nomer label kelas yang tersedia[7].

Decision Tree terdiri dari suatu struktur berbentuk diagram alur, dimulai dari simpul pertama yang disebut simpul akar (*root node*). Kemudian setiap *node* internal mewakili pengujian untuk suatu atribut. Cabang (*branch*) kemudian mencatat hasil tes yang dilakukan. Terakhir, daun (*leaves*) yang dicetak pada akhirnya mewakili label *class*[8]. Sebuah proses klasifikasi yang bertujuan untuk membentuk pohon keputusan yang diharapkan yang biasa disebut sebagai algoritma C4.5. Algoritma ini dipilih karena memiliki rangkaian proses klasifikasinya yang cepat dan sederhana[9].

Entropy adalah ukuran tidak pasti yaitu perbedaan hasil akhir nilai dari atribut tertentu. Jika nilai *entropy* semakin tinggi, maka semakin besar perbedaan keputusan yang ada[10]. *Gain* adalah proses pemilihan atribut yang digunakan sebagai pengujian atribut lain pada setiap pohon keputusan[10]. Jika ada atribut dengan banyak nilai atribut maka *split information* harus dilakukan setelah perhitungan *entropy* dan *gain*. Setelah nilai *gain* dan *split info* telah diperoleh, langkah berikutnya adalah menemukan nilai *gain ratio*.

Pada penelitian sebelumnya telah dilakukan penelitian menggunakan algoritma K-Nearest Neighbor (K-NN) untuk memprediksi hasil ujian nasional berdasarkan nilai try out siswa di SMK Bina Informatika Bintaro. Penelitian tersebut mencapai tingkat akurasi 83.21% dengan menggunakan *dataset* selama 3 tahun terakhir berjumlah 475 data[11].

Perbedaan utama penelitian ini dengan penelitian sebelumnya adalah fokus pada pengembangan sistem prediksi akademik siswa menggunakan algoritma C4.5. Meskipun penelitian sebelumnya telah berhasil mengklasifikasi *dataset* siswa dengan baik menggunakan algoritma K-NN, namun penggunaan algoritma C4.5 dapat memberikan yang berbeda dalam pemrosesan data dan interpretasi hasil prediksi prestasi akademik siswa.

Penelitian ini juga memprediksi prestasi siswa guna mendukung keputusan pemberian beasiswa berdasarkan nilai data siswa semester 5 tahun 2023 dan 2024, dengan 250 data yang digunakan. Penelitian ini mencapai tingkat akurasi 88% dengan algoritma C4.5.

2. METODE PENELITIAN

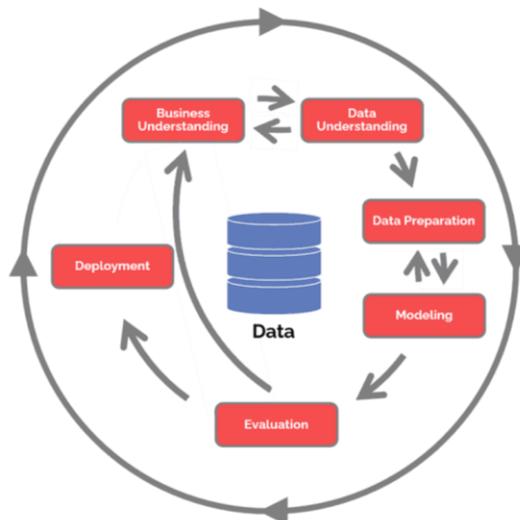
2.1 Metode Pembanding

Studi sebelumnya menggunakan algoritma C4.5 untuk memprediksi kelulusan tepat waktu mahasiswa di Uin Syarif Hidayatullah Jakarta. Penelitian ini menggunakan metode CFS untuk memilih fitur untuk meningkatkan akurasi prediksi kelulusan tepat waktu mahasiswa. Dengan menggunakan algoritma C4.5, penelitian ini mencapai tingkat akurasi sebesar 75,52% [12].

Sedangkan penelitian ini juga memprediksi prestasi siswa guna mendukung keputusan pemberian beasiswa berdasarkan nilai data siswa semester 5 tahun 2023 dan 2024, dengan 250 data yang digunakan. Penelitian ini mencapai tingkat akurasi 88% dengan algoritma C4.5.

2.2 CRISP-DM

Penggunaan metodologi data *mining* CRISP-DM sebagai pemecah masalah umum untuk bisnis dan penelitian. Metodologi ini terdiri dari enam fase, yaitu wawasan bisnis, wawasan data, persiapan data, pemodelan, evaluasi, dan implementasi. Proses metodologi ini terdiri dari 6 fase yang dapat dijelaskan pada Gambar 1 sebagai berikut[13].



Gambar 1. CRISP-DM

Berikut ini adalah penjelasan untuk setiap tahapan seperti yang ditunjukkan pada Gambar 1:

- Business Understanding*, Fase pertama dalam *crisp-dm* adalah pemahaman bisnis atau *business understanding*. Pada tahap ini pendefinisian masalah bisnis dapat dilakukan dengan mudah tetapi tahap ini harus dilakukan dengan tepat karena meskipun pada tahap ini terlihat sederhana, namun sebenarnya tahap ini memerlukan penguasaan nyata terhadap proses bisnis dan cara pengolahannya untuk mencapai tujuan. Tahapan ini merupakan tahap yang sangat penting untuk dapat melakukan proses pada tahap selanjutnya.
- Data Understanding*, Dalam tahap kedua adalah data *understanding* atau pemahaman yang digunakan data sebagai pemasang data yang dimiliki dengan data yang dibutuhkan. Pada proses tersebut jenis data digunakan sebagai penentu algoritma yang digunakan, serta tujuan suatu *project* analisis yang mana digunakan untuk mendapatkan *knowledge* dari kumpulan data *mining* tersebut. Mengumpulkan, menjelaskan dan juga meningkatkan kualitas data merupakan tahapan dalam fase ini.
- Data Prepration*, Pada tahap data *prepration* proses akhir pengembangan *dataset* dibuat dalam bentuk data mentah. Proses yang dilakukan pada fase ini melewati proses pembersihan data, pemilihan data, hingga transformasi data yang digunakan untuk tahap input permodelan.
- Modelling*, *Modelling* adalah tahap di mana pembelajaran mesin secara langsung mengidentifikasi alat yang digunakan untuk data *mining*, teknik data *mining* dan algoritma yang diperlukan. Klasifikasi menggunakan model pohon keputusan dari algoritma pilihan *classification and regression tree* dengan kriteria *information gain* untuk membuat pohon keputusan.
- Evaluation*, Pengujian juga dikenal sebagai *evaluation* adalah suatu proses yang dilakukan dengan mempelajari tingkat pola kinerja yang dihasilkan oleh algoritma yang digunakan. Proses ini melibatkan penggunaan matrik konfusi sebagai parameter yang memiliki nilai akurasi, presisi dan *recall* sebagai acuannya.
- Deployment*, Penyebaran atau *deployment* merupakan tahap penarikan kesimpulan dari hasil berupa laporan atau jurnal dengan menggunakan metode yang diperoleh.

2.3 Confusion Matrix

Confusion Matrix digunakan dalam masalah klasifikasi, di mana N adalah jumlah kelas yang akan diprediksi, dan *matrix* ini berukuran $N \times N$ [14]. Berikut ini adalah tabel 1 *confusion matrix*.

- a. *Data Selection*, Data yang diperoleh kemudian dipilih dengan menghilangkan atribut nama, nis, nilai produk kreatif dan kewirausaha, nilai bimbingan uji kompetensi, nilai bimbingan karir sehingga jumlah atribut yang digunakan sebanyak 9 yaitu nilai matematika, nilai bahasa Indonesia, nilai agama, nilai bahasa inggris, nilai *project work*, jumlah izin, jumlah sakit, jumlah alpha, nilai ekskul.
- b. *Data Transformation*. Pada tahap ini, atribut baru ditambahkan yaitu atribut prestasi. Label kelas dengan kategori “Tinggi” dan “Rendah” dibuat atau ditambahkan untuk memenuhi *Attribute/Feature Construction* yang diperlukan. Proses transformasi atau perubahan data digunakan untuk mengubah data asli menjadi data dengan menggunakan berbagai kategori yang ditetapkan berdasarkan persyaratan pengolahan data *mining*. Ada 9 atribut yang ditransformasikan dan Tabel 2 merupakan penjabaran dari masing-masing kategori dari masing-masing atribut.

Tabel 2. *Data Transformation*

No	Atribut	Nilai
1	Nilai	100-91 (A)
	Agama	90-79 (B)
		78-60 (C)
2	Nilai	100-91 (A)
	Bindo	90-79 (B)
		78-60 (C)
3	Nilai Mtk	100-91 (A)
		90-79 (B)
		78-60 (C)
4	Nilai Binggris	100-91 (A)
		90-79 (B)
		78-60 (C)
5	Nilai Pw	100-91 (A)
		90-79 (B)
		78-60 (C)
6	Jumlah Izin	>5 (Tinggi)
		2-5 (Sedang)
		0-1 (Rendah)
7	Jumlah Sakit	>5 (Tinggi)
		2-5 (Sedang)
		0-1 (Rendah)
8	Jumlah Alpha	>5 (Tinggi)
		2-5 (Sedang)
		0-1 (Rendah)
9	Nilai Ekskul	B+-A (Sangat Baik)
		B(-)-B+ (Baik)
		C-B(-) (Kurang Baik)

3.4 Modelling

Pendekatan C4.5 dipilih sebagai metode setelah melakukan tahap *preprocessing* dan membuat model untuk memprediksi prestasi akademik siswa. Berikut Gambar 3 adalah beberapa contoh aturan yang ditemukan:

```
Atribut terpilih = bindo, dengan nilai gain = 0.041
=====
Opsi 1 :
jumlah A/C = 74
jumlah B = 127
Split = 0.949
Rasio = 0.043
Opsi 2 :
jumlah C/B = 136
jumlah A = 65
Split = 0.908
Rasio = 0.045
Opsi 3 :
jumlah B/A = 192
jumlah C = 9
Split = 0.264
Rasio = 0.155
=====
cabang 1
(bindo='C')
Jumlah data = 9
Jumlah Tinggi = 5
Jumlah Rendah = 4
Entropy All = 0.991
```

Gambar 3. Aturan / Rule

Berikut ini Gambar 4 merupakan contoh perhitungan dari pembentukan pohon keputusan:

Nilai Atribut	Jumlah data	Jumlah Tinggi	Jumlah Rendah	Entropy	Gain
agama='B'	69	62	7	0.474	
agama='C'	59	48	11	0.694	
agama='A'	73	64	9	0.539	0.007
bindo='B'	127	107	20	0.628	
bindo='A'	65	62	3	0.27	
bindo='C'	9	5	4	0.991	0.041
mtk='B'	177	152	25	0.587	
mtk='A'	22	21	1	0.267	
mtk='C'	2	1	1	1	0.013
binggris='B'	110	97	13	0.524	
binggris='A'	74	65	9	0.534	
binggris='C'	17	12	5	0.874	0.012
pw='C'	42	31	11	0.83	
pw='B'	136	123	13	0.455	
pw='A'	23	20	3	0.559	0.024
izin='Sedang'	51	44	7	0.577	
izin='Rendah'	126	110	16	0.549	
izin='Tinggi'	24	20	4	0.65	0.001
sakit='Rendah'	163	141	22	0.571	
sakit='Sedang'	31	27	4	0.555	
sakit='Tinggi'	7	6	1	0.592	-0
alpha='Rendah'	192	169	23	0.529	
alpha='Sedang'	7	3	4	0.985	
alpha='Tinggi'	2	2	0	0	0.029
ekskul='Kurang Baik'	41	34	7	0.659	
ekskul='Baik'	120	104	16	0.567	
ekskul='Sangat Baik'	40	36	4	0.469	0.003

Gambar 4. Perhitungan Pembentukan Pohon Keputusan

Karena *gain* tertinggi adalah atribut bindo, atribut tersebut berfungsi sebagai titik acuan dan referensi untuk perhitungan entropi dan *gain* berikutnya. Setelah itu, ulangi hingga setiap atribut memiliki kelasnya sendiri.

3.5 Confusion Matrix

Pengujian ini mencari nilai akurasi, *recall* / sensitivitas dan spesivitas dengan menggunakan metode *confusion matrix*. Di bawah ini adalah Tabel 3 dan Tabel 4 penjelasan untuk menghitung *confusion matrix*.

Tabel 3. Tabel Perbandingan

No	Label Sebenarnya	Label Prediksi	Kategori
1	Tinggi	Tinggi	TP
2	Tinggi	Rendah	FN
3	Tinggi	Tinggi	TP
4	Rendah	Tinggi	FP
5	Rendah	Tinggi	FP
6	Tinggi	Rendah	FN
7	Tinggi	Tinggi	TP
8	Tinggi	Rendah	FN
9	Tinggi	Rendah	FN

10	Tinggi	Tinggi	TP
...			
50	Tinggi	Tinggi	TP

Tabel 4. Pengujian *Confusion Matrix*

<i>True Positive</i> (TP)	<i>True Negative</i> (TN)	<i>False Positive</i> (FP)	<i>False Negative</i> (FN)
44	0	2	4

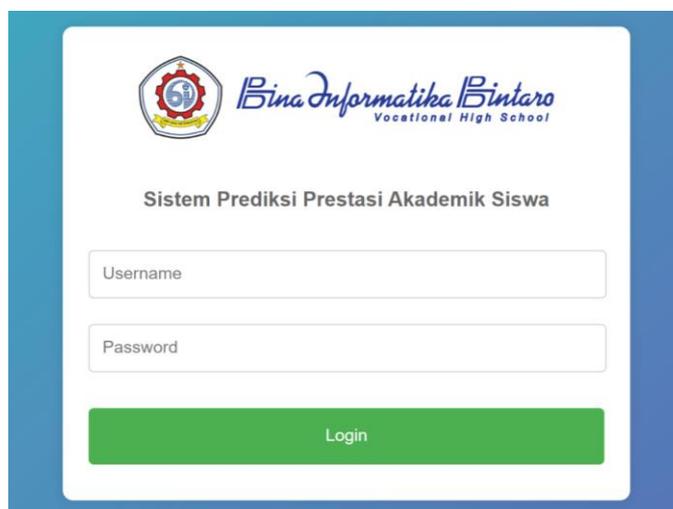
- a. Berdasarkan hasil pengujian dengan data uji sebanyak 50 baris data
 1. TP (*True Positive*) = 44 (prediksi benar siswa berprestasi tinggi).
 2. TN (*True Negative*) = 0 (prediksi benar siswa berprestasi rendah).
 3. FP (*False Positive*) = 2 (prediksi salah siswa berprestasi tinggi padahal prestasi rendah).
 4. FN (*False Negative*) = 4 (prediksi salah siswa berprestasi rendah padahal berprestasi tinggi).
- b. Setelah mengetahui berapa jumlah TP, TN, FP dan FN. Kita dapat menghitung *confusion matrix* sebagai berikut:
 1. Akurasi: $(44+0)/(44+0+2+4)=44/50=0.88$ atau 88%
 2. Sensivitas: $44/(44+4)=44/48=0.9167$ atau 91.67%
 3. Spesivitas: $0/(0+2)=0$

Hasil *recall* / sensitivitas sebesar 91,67% dan spesivitas 0% dan akurasi hingga 88% hal tersebut dapat memastikan bahwa model ini dapat memprediksi data dengan baik.

3.6 Tampilan Layar

3.6.1. Tampilan Layar Menu Login

Menu login pertama muncul seperti yang ditunjukkan pada gambar 5. Pada menu *login*, pengguna harus memasukkan username dan password yang benar. Jika benar maka pengguna akan berhasil *login* dan masuk ke halaman beranda. Jika tidak benar maka pengguna harus memasukkan kembali *username* dan *password*.



Gambar 5. Tampilan Layar Menu Login

3.6.2. Tampilan Layar Halaman Beranda

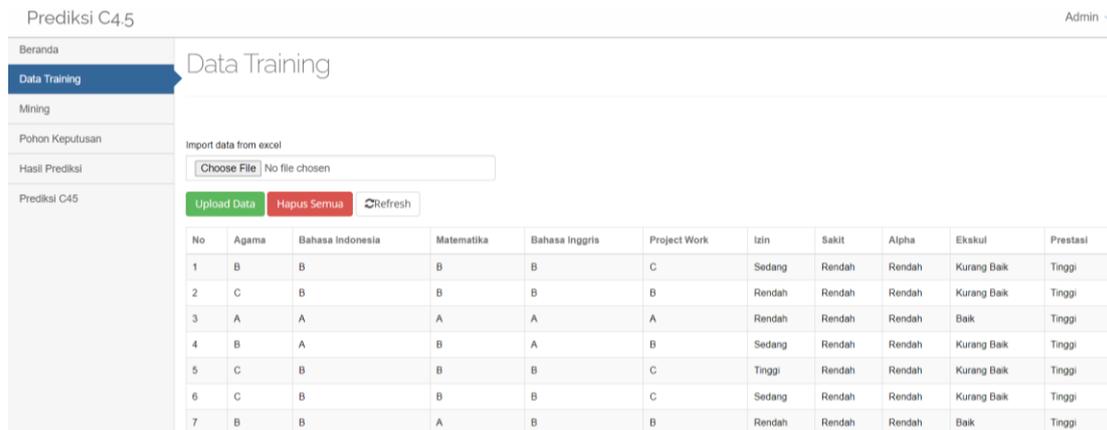
Gambar 6 menunjukkan layar halaman beranda yang muncul setelah pengguna berhasil *login*. Pada menu halaman beranda pengguna dapat memilih halaman data *training* untuk menginput data latih, halaman *mining* untuk melihat proses *mining* C4.5, halaman pohon keputusan untuk melihat hasil *mining* yang sudah terbentuk menjadi *rule* dan pengguna bisa melakukan uji data untuk menghitung akurasi, halaman prediksi C4.5 untuk menginput data apakah sudah *valid* sesuai hasil pohon keputusan dan hasil halaman prediksi C4.5 untuk melihat hasil data yang pengguna input dari halaman prediksi C4.5.



Gambar 6. Tampilan Layar Halaman Beranda

3.6.3. Tampilan Layar Halaman Data Training

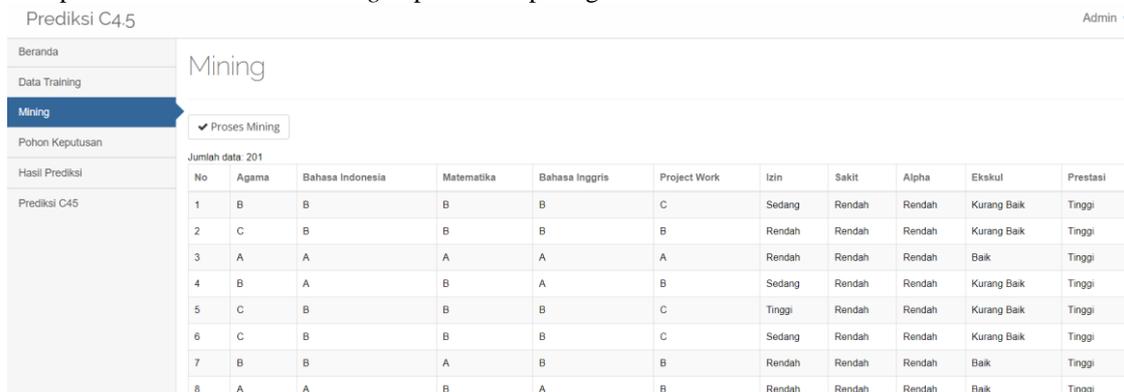
Gambar 7 adalah halaman data *training* untuk import *file excel*. Ada tombol *upload* data untuk menampilkan data ke layar dalam bentuk tabel dan ada tombol hapus data untuk menghapus semua data.



Gambar 7. Tampilan Layar Halaman Data Training

3.6.4. Tampilan Layar Mining dan Proses Mining

Gambar 8 menunjukkan halaman *mining* dimana datanya siap diproses dengan algoritma C4.5. Proses *mining* ini kemudian dilakukan dengan algoritma C4.5 untuk mencari *entropy* dan *gain*. Setelah *gain* ditentukan maka pohon keputusan dibuat. Proses *mining* dapat dilihat pada gambar 9.



Gambar 8. Tampilan Layar Halaman Mining

Prediksi C4.5 Admin

Nilai Atribut	Jumlah data	Jumlah Tinggi	Jumlah Rendah	Entropy	Gain
agama='B'	69	62	7	0.474	
agama='C'	59	48	11	0.684	
agama='A'	73	64	9	0.539	0.007
bindo='B'	127	107	20	0.628	
bindo='A'	65	62	3	0.27	
bindo='C'	9	5	4	0.991	0.041
mtk='B'	177	152	25	0.587	
mtk='A'	22	21	1	0.267	
mtk='C'	2	1	1	1	0.013
binggris='B'	110	97	13	0.524	
binggris='A'	74	65	9	0.534	
binggris='C'	17	12	5	0.874	0.012

Gambar 9. Tampilan Layar Halaman Proses *Mining*

3.6.5. Tampilan Layar Pohon Keputusan

Gambar 10 menunjukkan halaman pohon keputusan dimana *rule* atau aturan-aturan yang terbentuk merupakan hasil dari proses *mining*.

Prediksi C4.5 Admin

Pohon Keputusan

Jumlah rule: 65

Id	Aturan
1	IF (bindo='C') AND (zini='Sedang') THEN Label = Tinggi
2	IF (bindo='C') AND (zini='Tinggi' OR zini='Rendah') AND (binggris='A') THEN Label = Tinggi
3	IF (bindo='C') AND (zini='Tinggi' OR zini='Rendah') AND (binggris='B' OR binggris='C') AND (ekskul='Sangat Baik') THEN Label = Rendah
4	IF (bindo='C') AND (zini='Tinggi' OR zini='Rendah') AND (binggris='B' OR binggris='C') AND (ekskul='Baik' OR ekskul='Kurang Baik') AND (pwr='C') AND (ekskul='Baik') THEN Label = Rendah
5	IF (bindo='C') AND (zini='Tinggi' OR zini='Rendah') AND (binggris='B' OR binggris='C') AND (ekskul='Baik' OR ekskul='Kurang Baik') AND (pwr='C') AND (ekskul='Kurang Baik') THEN Label = Tinggi
6	IF (bindo='C') AND (zini='Tinggi' OR zini='Rendah') AND (binggris='B' OR binggris='C') AND (ekskul='Baik' OR ekskul='Kurang Baik') AND (pwr='B') THEN Label = Tinggi
7	IF (bindo='B' OR bindo='A') AND (alpha='Tinggi') THEN Label = Tinggi
8	IF (bindo='B' OR bindo='A') AND (alpha='Rendah' OR alpha='Sedang') AND (pwr='A') AND (sakit='Tinggi') THEN Label = Rendah
9	IF (bindo='B' OR bindo='A') AND (alpha='Rendah' OR alpha='Sedang') AND (pwr='A') AND (sakit='Rendah' OR sakit='Sedang') AND (ekskul='Kurang Baik') AND (zini='Sedang') THEN Label = Tinggi
10	IF (bindo='B' OR bindo='A') AND (alpha='Rendah' OR alpha='Sedang') AND (pwr='A') AND (sakit='Rendah' OR sakit='Sedang') AND (ekskul='Kurang Baik') AND (zini='Rendah' OR zini='Tinggi') AND (mtk='A') THEN Label = Rendah

Gambar 10. Tampilan Layar Halaman Pohon Keputusan

3.6.6. Tampilan Layar Halaman Uji *Rule* dan Hasil Akurasi

Gambar 11 adalah halaman *uji rule* untuk *import file* data uji. Ada tombol *upload* data untuk menampilkan data ke layar dalam bentuk tabel, ada tombol *hapus* data untuk menghapus semua data dan ada tombol *hitung* akurasi untuk menghitung akurasi C4.5. Gambar 12 menunjukkan hasil perhitungan akurasi dengan *confusion matrix*.

Prediksi C4.5 Admin

Uji Rule

Import data from excel

No file chosen

Jumlah data uji: 50

No	Agama	Bahasa Indonesia	Matematika	Bahasa Inggris	Project Work	Izin	Sakit	Alpha	Ekskul	Prestasi
1	A	B	B	A	C	Rendah	Rendah	Rendah	Sangat Baik	Tinggi
2	A	A	C	A	B	Rendah	Rendah	Rendah	Sangat Baik	Tinggi
3	A	A	C	A	B	Rendah	Tinggi	Rendah	Baik	Tinggi
4	B	B	B	B	B	Sedang	Rendah	Rendah	Baik	Tinggi
5	A	B	B	A	B	Rendah	Rendah	Rendah	Kurang	Tinggi
6	A	A	B	B	C	Sedang	Rendah	Rendah	Baik	Tinggi
7	B	A	C	A	A	Rendah	Rendah	Rendah	Baik	Tinggi

Gambar 11. Tampilan Layar Halaman Uji *Rule*

Prediksi C4.5

No	A	B	C	B	B	Tinggi	Sedang	Rendah	Baik	Tinggi	Tinggi	TP	Status
41	B	B	C	B	B	Rendah	Sedang	Rendah	Baik	Tinggi	Tinggi	17	benar
42	B	B	B	B	B	Sedang	Sedang	Rendah	Baik	Tinggi	Tinggi	33	benar
43	A	A	A	A	A	Rendah	Rendah	Rendah	Baik	Tinggi	Tinggi	13	benar
44	A	B	B	B	B	Tinggi	Tinggi	Rendah	Kurang	Tinggi	Tinggi	45	benar
45	B	C	B	A	B	Sedang	Sedang	Rendah	Sangat Baik	Tinggi	Tinggi	1	benar
46	B	A	B	C	C	Rendah	Rendah	Rendah	Sangat Baik	Tinggi	Rendah	19	salah
47	B	B	B	B	C	Rendah	Rendah	Rendah	Sangat Baik	Tinggi	Tinggi	31	benar
48	B	A	C	A	B	Rendah	Rendah	Rendah	Sangat Baik	Tinggi	Tinggi	17	benar
49	C	B	B	C	B	Rendah	Rendah	Rendah	Baik	Tinggi	Rendah	20	salah
50	B	B	B	B	B	Rendah	Rendah	Rendah	Kurang	Tinggi	Tinggi	45	benar

Jumlah prediksi: 50
 Jumlah tepat: 44
 Jumlah tidak tepat: 6

AKURASI = 88 %
LAJU ERROR = 12 %

TP=44 TN=0 FP=2 FN=4
 Sensitivitas = 91.67 %
 Spesifisitas = 0 %

Gambar 12. Hasil Perhitungan *Confusion Matrix*

3.6.7. Tampilan Layar Halaman Prediksi C4.5

Gambar 13 menunjukkan halaman yang memiliki *form input* untuk membuat prediksi. Jika pengguna ingin melakukan prediksi maka pengguna harus menekan tombol *submit*.

Prediksi C4.5

Beranda
Data Training
Mining
Pohon Keputusan
Hasil Prediksi
Prediksi C4.5

Prediksi

Nama: Ulok
 Nilai Agama: B
 Nilai Bahasa Indonesia: B
 Nilai Matematika: C
 Nilai Bahasa Inggris: C
 Nilai Project Work: A
 Jumlah Izin: Sedang
 Jumlah Sakit: Tinggi
 Jumlah Alpha: Rendah
 Nilai Ekskul: Sangat Baik

Gambar 13. Tampilan Layar Halaman Prediksi C4.5

3.6.8. Tampilan Layar Hasil Prediksi

Gambar 14 menunjukkan hasil prediksi yang sebelumnya sudah di *input* di halaman prediksi C4.5 yang kemudian ditampilkan dalam bentuk tabel.

Prediksi C4.5

Beranda
Data Training
Mining
Pohon Keputusan
Hasil Prediksi
Hasil Prediksi
Prediksi C4.5

Hasil

Ops: [Hapus Semua Data](#)

Jumlah data: 1

No	Nama	Agama	Bahasa Indonesia	Matematika	Bahasa Inggris	Project Work	Izin	Sakit	Alpha	Ekskul	Prestasi	Id Rule
1	Ulok	B	B	C	C	A	Sedang	Tinggi	Rendah	Sangat Baik	Rendah	8

Gambar 14. Tampilan Layar Halaman Hasil Prediksi C4.5

4. KESIMPULAN

Berdasarkan dengan hasil data dalam penelitian ini, dapat memberikan kesimpulan yaitu sistem prediksi prestasi akademik siswa guna mendukung keputusan pemberian beasiswa menggunakan algoritma C4.5 di SMK Bina Informatika Bintaro berbasis web membuat proses penentuan penerimaan beasiswa menjadi lebih mudah dan objektif. Keberagaman nilai dari akurasi yang diperoleh juga dipengaruhi oleh jumlah data yang digunakan. Ini

terlihat dari proses analisis dan dari seluruh proses pengujian *dataset* yang telah dilakukan. Namun, nilai spesivisitas yang rendah menunjukkan bahwa model ini kurang mampu mengidentifikasi siswa dengan prestasi yang rendah dengan benar. Berikut hasil dari *dataset* yang digunakan adalah, jumlah data 250 baris, pembagian data latih (80%) dan data uji (20%), nilai akurasi 88%, nilai *recall* / sensivisitas 91,67%, nilai spesivisitas 0%. Penelitian ini diharapkan dapat membantu sekolah menentukan penerima beasiswa secara lebih efektif dan akurat.

DAFTAR PUSTAKA

- [1] F. Gunadi and S. Rheno Widiyanto, "Perbandingan Data Warehouse Cloud Computing Menggunakan Konvensional Berbasis Kriptografi," *Seminar Nasional Teknologi Komputer Sains (SAINTEKS)*, 2020, pp. 69-73.
- [2] E. T. Naldy and A. Andri, "Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN," *J. Nas. Ilmu Komput.*, vol. 2, no. 2, pp. 89-101, 2021, doi: 10.47747/jurnalnik.v2i2.525.
- [3] N. Iriadi, L. Setioningtias, and P. Priatno, "Implementasi Data Mining Pada Klasifikasi Ketidakhadiran Pegawai Menggunakan Metode C4.5," *Comput. Sci.*, vol. 1, no. 1, pp. 53-61, 2021, doi: 10.31294/coscience.v1i1.198.
- [4] P. Bhatia, *Data Mining and Data Warehousing*. 2019. doi: 10.1017/9781108635592.
- [5] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, pp. 35-43, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [6] R. Buaton, M. Zarlis, H. Mawengkang, and S. Efendi, "Optimization of forecasting time series with RBT (Rule Best Time Series)," *J. Theor. Appl. Inf. Technol.*, vol. 98, no. 15, pp. 2977-2989, 2020.
- [7] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Informatika Budidarma*, vol. 4, no. 2, pp. 437-444, 2020, doi: 10.30865/mib.v4i2.2080.
- [8] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innov. Res. Informatics*, vol. 1, no. 2, pp. 63-69, 2019, doi: 10.37058/innovatics.v1i2.918.
- [9] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," *J. Krisnadana*, vol. 1, no. 2, 2022, doi: 10.58982/krisnadana.v1i2.108.
- [10] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 21-32, 2020.
- [11] R. Aljabar and D. Kusumaningsih, "Penerapan Algoritma Klasifikasi K-Nearest Dengan Menggunakan Hasil Nilai Try Out Siswa Sekolah Menengah Kejuruan Berbasis Desktop Pada SMK Bina Informatika Bintaro," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 1, no. 1, pp. 136-142, 2018.
- [12] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 6, no. 1, pp. 61-74, 2023, doi: 10.36080/skanika.v6i1.2976.
- [13] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103-108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [14] T. A. Yoga Siswa, G. M. Putra, and A. Prafanto, "Seleksi Fitur Information Gain dan Teknik Pruning Untuk Memperbaiki Akurasi Algoritma C4.5 dalam Kasus Keterlambatan Biaya Kuliah," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 17, no. 2, pp. 102-112, 2022, doi: 10.30872/jim.v17i2.11794.
- [15] A. F. Pulungan, "Analisis Kinerja Bray Curtis Distance Canberra Distance Dan Euclidean Distance Pada Algoritma K-Nearest Neighbor," *Doctoral dissertation, Universitas Sumatera Utara*, 2019.