



Artikel

Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4.5

Eko Budiarto¹, Rino², Susanto Hariyanto³, Dera Susilawati⁴

^{1,2,3,4} Universitas Buddhi Dharma, Teknik Informatika, Banten, Indonesia

SUBMISSION TRACK

Received: March 28, 2022

Final Revision: May 03, 2022

Available Online: May 25, 2022

KEYWORD

Beasiswa PIP, *Data Mining*, Klasifikasi, C4.5, Web

KORESPONDENSI

Phone: 083162920690

E-mail: ekobudiarto19@email.com

A B S T R A C T

SD Maria Mediatrix merupakan instansi yang bergerak dalam bidang pendidikan yang mengikuti berbagai program pemerintah dalam menunjang kelancaran belajar, seperti beasiswa Program Indonesia Pintar (PIP). Proses seleksi penerimaan beasiswa pada SD Maria Mediatrix sejauh ini masih dilakukan secara konvensional yang sering kali menimbulkan beberapa permasalahan, antara lain membutuhkan waktu yang lama dan hasilnya kurang tepat sasaran. Oleh karena itu, dilakukan penelitian serta perancangan sebuah aplikasi yang dapat mengklasifikasi penerima beasiswa dengan metode klasifikasi, serta menggunakan C4.5 sebagai algoritmanya. Hasil dari penelitian ini adalah dibuatnya sebuah aplikasi berbasis web yang dapat melakukan klasifikasi penerima beasiswa PIP dari atribut yang digunakan sebagai penentu. Tujuan dari penelitian ini adalah melakukan implementasi klasifikasi dengan algoritma C4.5 ke dalam sebuah aplikasi berbasis web yang diharapkan dapat membantu SD Maria Mediatrix dalam menentukan penerima beasiswa, yaitu dengan mengumpulkan data siswa yang akan digunakan sebagai dataset. Dataset tersebut kemudian dilakukan pengambilan sampel data untuk keperluan perhitungan manual. Proses perhitungan secara manual berfungsi untuk mengetahui alur dari perhitungan agar dapat diimplementasikan dengan baik pada aplikasi yang akan dibuat. Aplikasi berbasis web yang dihasilkan melakukan proses perhitungan C4.5 dan pembentukan aturan klasifikasi secara otomatis dari data siswa yang dimasukkan, pengguna aplikasi cukup menginput nama siswa, penghasilan ayah, penghasilan ibu, pemegang KIP, jumlah saudara, dan jarak rumah hasilnya akan mengetahui siswa tersebut layak mendapat beasiswa atau tidak.

PENGANTAR

Pada era digitalisasi sekarang, perkembangan teknologi informasi begitu melesat hingga menyebabkan tingkat keakuratan suatu data sangat dibutuhkan dalam kehidupan sehari-hari. salah satu cabang teknologi informasi yang dapat menganalisa data adalah *data mining*. SD Maria Mediatrix merupakan instansi yang bergerak dalam bidang pendidikan serta menerapkan sistem beasiswa. Proses seleksi penerimaan beasiswa pada SD Maria Mediatrix sejauh ini masih dilakukan secara konvensional yang sering kali menimbulkan beberapa permasalahan, antara lain membutuhkan waktu yang lama dan ketelitian yang tinggi. Sehingga beasiswa diberikan lama dan hasilnya tidak tepat sasaran, ada siswa yang memang kurang mampu namun tidak masuk kriteria penerima beasiswa oleh pihak sekolah. Oleh karena, itu diperlukan suatu aplikasi berbasis web dengan sebuah metode dan algoritma yang cocok sehingga dapat memenuhi kebutuhan tersebut. Metode klasifikasi dengan algoritma C4.5 digunakan karena dapat menangani atribut diskrit dan kontinyu sehingga cocok untuk diterapkan pada penelitian ini yang memiliki atribut diskrit dan kontinyu. Keunggulan algoritma C4.5 dapat dibuktikan melalui salah satu penelitian yang dibuat oleh Anam[1] dimana berdasarkan hasil implementasi algoritma C4.5 dan Naive Bayes untuk model klasifikasi penerima beasiswa menunjukkan model algoritma C4.5 mempunyai kinerja yang lebih baik dari Naive Bayes. Oleh karena itu, dalam penelitian ini penulis akan menerapkan metode klasifikasi C4.5 untuk menentukan penerima beasiswa berdasarkan attribute jumlah penghasilan orang tua, pemegang KIP, jumlah saudara kandung, jarak rumah ke sekolah, layak PIP.

I. METODE

1.1 Data Mining

Data mining adalah langkah analisis terhadap proses penemuan pengetahuan di dalam basis data atau knowledge discovery in databases (KDD). Pengetahuan bisa berupa pola data

atau relasi antar data yang valid (yang tidak diketahui sebelumnya) [2].

Data mining didefinisikan sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar [3]. Data mining juga dapat diartikan sebagai pengestrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga knowledge discovery in databases (KDD).

Berdasarkan beberapa pengertian diatas, data mining dapat diartikan sebagai suatu teknik yang berfungsi untuk menggali informasi berharga pada suatu data yang sehingga dapat ditemukan suatu pola menarik yang sebelumnya tidak diketahui.

1.2 Klasifikasi

Klasifikasi merupakan salah satu metode dalam data mining yang digunakan dalam menganalisis sekumpulan data penting. Metode-metode dalam klasifikasi dapat secara otomatis memprediksi kelas dari data lain yang belum diklasifikasikan [4].

Menurut Islamnuddin [5] Klasifikasi atau taksonomi adalah proses menempatkan suatu objek atau konsep kedalam satu set kategori berdasarkan objek atau konsep yang bersangkutan. Metode klasifikasi berfungsi untuk membantu memahami pengelompokan dari suatu data.

Dari pengertian di atas, dapat disimpulkan bahwa klasifikasi merupakan pengelompokan beberapa parameter kedalam salah satu kategori yang sudah ditetapkan sebelumnya. Algoritma klasifikasi dapat digunakan untuk memprediksi suatu label pada kelas tertentu.

1.3 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal [5]. Algoritma C4.5 dan pohon keputusan merupakan dua hal yang tidak dapat dipisahkan [6], karena pada proses pembangunan sebuah pohon keputusan, dibutuhkan algoritma C4.5.

Cara algoritma C4.5 untuk membangun pohon keputusan yaitu[7]:

1. Pilih atribut yang akan digunakan sebagai akar.
2. Buatlah sebuah cabang untuk setiap nilai.
3. Bagilah kasus dalam sebuah cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Dalam buku [8] ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5.

Menyiapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah dikelompokkan kedalam kelas tertentu.

Setelah itu tentukan akar dari pohon. Pilih akar dari atribut, cara adalah dengan menghitung nilai gain dari semua atribut, yang menjadi akar pertama adalah nilai gain yang paling tinggi. Sebelum menentukan nilai gain, terlebih dahulu hitung nilai entropy. Untuk menentukan nilai entropy gunakan rumus :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan:

S = himpunan kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

Setelah itu tentukan nilai gain menggunakan rumus :

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

Keterangan :

S = Himpunan kasus

A = fitur N = jumlah partisi atribut A

|Si| = proporsi Si terhadap S

|S| = jumlah kasus dalam S

Setelah itu ulangilah langkah ke-2 sampai semua record terpartisi secara sempurna.

Proses partisi pohon keputusan akan berhenti saat:

1. Semua record dalam simpul N mendapat kelas yang sama.
2. Tidak ada atribut di dalam record yang dipartisi lagi.
3. Tidak ada record di dalam cabang yang kosong.

2.4 CRISP-DM

CRISP-DM merupakan sebuah standarisasi data mining yang digunakan sebagai strategi pemecahan masalah yang secara umum merupakan dari segi bisnis atau unit penelitian [9].

Berikut ini merupakan tahap-tahap dalam CRISP-DM:

a. Business understanding

Pada tahapan ini, dilakukan pemahaman tujuan dan kebutuhan dari sudut pandang bisnis yang nantinya diterjemahkan dalam pendefinisian masalah pada data mining. Setelah itu dilakukan penentuan rencana dan strategi untuk mencapai tujuan dan kebutuhan tersebut.

b. data understanding

Pada tahap ini, dilakukan pengumpulan data untuk kemudian diolah sesuai dengan tujuan penelitian ini. Data awal yang diolah merupakan data sekunder berbentuk excel yang berisi 372 data siswa SD Maria Mediatrix Tahun Pelajaran 2020/2021. Pengolahan data yang dilakukan bertujuan untuk klasifikasi beasiswa dilihat dari beberapa atribut sehingga menjadi atribut penentu.

Tabel 1. Data awal

No	Nama	NIPD	JK	NISN	Tempat Lahir	Jarak Rumah
1	Abigail Patricia Thong	2973	P	009774287	Bekasi	1 km
2	Abner Moses Susilo	3180	L	012983154	Tangerang	2 km
...
372	Zefanya Salomi	3380	P	011263244	Tangerang	9 km

Berikut ini merupakan atribut-atribut yang terdapat pada data awal:

Tabel 2. Atribut data awal

Atribut	Keterangan	Tipe Data
No	Nomor urutan	Integer
Nama	Nama Peserta Didik	Varchar
NIPD	Nomor Induk Peserta Didik	Varchar
JK	Jenis Kelamin	Varchar
NISN	Nomen Induk Siswa Nasional	Varchar
Tempat Lahir	Tempat Lahir Peserta Didik	Varchar
Tanggal Lahir	Tanggal Lahir Peserta Didik	Date
NIK	Nomor Induk Kependudukan	Varchar
Agama	Agama Peserta Didik	Varchar
Alamat	Alamat Rumah Peserta Didik	Text
RT	RT Peserta Didik	Varchar
RW	RW Peserta Didik	Varchar
Kelurahan	Kelurahan Peserta Didik	Varchar
Kecamatan	Kecamatan Peserta Didik	Varchar
Jenis Tinggal	Tinggal Bersama Tua/Wali/Lainnya	Varchar
Alat Int	Transportasi untuk ke sekolah	Varchar
HP	Nomor HP Peserta Didik	Varchar
Nama	Nama Ayah	Varchar
Pekerjaan	Pekerjaan Ayah	Varchar
Penghasilan	Penghasilan Ayah	Integer
Nama	Nama Ibu	Varchar
Pekerjaan	Pekerjaan Ibu	Varchar
Penghasilan	Penghasilan Ibu	Integer
Rombel Saat Ini	Rombongan belajar	Varchar
Penerima KIP	Informasi penerima KIP	Varchar
Nomor KIP	Informasi nomor KIP	Varchar
Nama di KIP	Nama pada KIP	Varchar
Nomor KKS	Nomor KKS	Varchar
Layak PIP (usulan dari sekolah)	Informasi dapat bantuan PIP	Varchar
Kebutuhan Khusus	Informasi keadaan fisik Peserta Didik	Varchar
Anak ke-berapa	Informasi Anak ke-	Varchar
Lintang	Titik Lintang Alamat Peserta Didik	Varchar
Bujur	Titik Bujur Alamat Peserta Didik	Varchar
Jumlah Saudara Kandung	Jumlah Saudara Kandung	Integer
Jarak Rumah (KM)	Jarak tempat tinggal dengan sekolah	Integer

c. data preparation

Pada tahap ini, dilakukan proses pembangunan dataset yang meliputi pemilihan tabel dan atribut-atribut data, serta proses *data cleaning*. *Data preparation* dalam penelitian ini dimulai dengan *data cleaning* dengan menghilangkan atribut-atribut yang dianggap tidak berkontribusi terhadap hasil akhir.

Atribut-atribut yang tidak berkontribusi pada hasil akhir tersebut meliputi No, Nama, NIPD, JK, NISN, Tempat Lahir, Tanggal Lahir, NIK, Agama, Alamat, RT, RW, Kelurahan, Kecamatan, Jenis Tinggal, Alat Transportasi, HP, Nama Ayah, Pekerjaan Ayah, Nama Ibu, Pekerjaan Ibu, Rombel Saat Ini, Nomor KIP, Nama di KIP, Nomor KKS, Kebutuhan Khusus, Anak ke-berapa, Lintang dan Bujur sehingga hanya menyisakan enam atribut yaitu :

- Penghasilan Ayah
- Penghasilan Ibu
- Penerima KIP
- Layak PIP(usulan dari sekolah).
- Jumlah Saudara Kandung
- Jarak Rumah ke Sekolah (KM).

Pada *data preparation* penelitian ini, tidak terdapat proses *data transformation* dan *data reduction*.

d. Modeling

Pada tahap ini, dilakukan perancangan konsep serta pemilihan dan penerapan teknik pemodelan yang disesuaikan agar mendapatkan hasil yang optimal. Tahap modeling ini menggunakan teknik data mining klasifikasi dengan algoritma C4.5 yang nantinya akan dilakukan pengolahan data yang bertujuan untuk klasifikasi beasiswa

e. Evaluation

Pada tahap ini, dilakukan evaluasi terhadap hasil dari pemodelan yang dilakukan sebelumnya serta permasalahan bisnis yang belum dipertimbangkan. Pada penelitian ini, penulis menggunakan *confussion matrix* sebagai acuan validitas aturan yang telah dihasilkan. *Confussion matrix* merupakan alat pengukuran yang dapat digunakan untuk

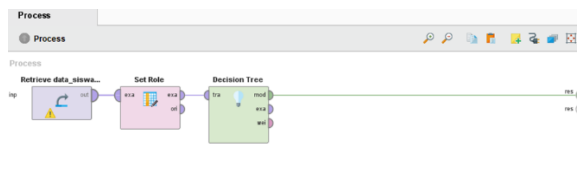
menghitung kinerja atau tingkat kebenaran proses klasifikasi [10].

f. Deployment

Pada tahap ini, informasi dan pengetahuan yang telah diperoleh akan diimplementasikan kedalam sebuah bentuk khusus agar dapat digunakan oleh pengguna. Tahap ini berkaitan dengan pengimplementasian data mining dalam sebuah perusahaan.

II. HASIL

Dataset yang sudah terbentuk akan diolah menggunakan aplikasi RapidMiner Studio 9.9.02. Setelah itu, dilakukan desain proses C4.5 dengan memasukkan data hasil *import*, operator *Set Role*, dan operator *Decision Tree*. ketiga operator tersebut kemudian dihubungkan.



Gambar 1. Desain Proses C4.5 pada RapidMiner
Berikut ini adalah hasil dari proses data mining yang didapat dengan menggunakan aplikasi RapidMiner Studio 9.9.02

Tree

```

penghasilan ayah = Menengah
| jml saudara > 2.500: Ya {Ya=56, Tidak=50}
| jml saudara ≤ 2.500
| | jarak rumah = Dekat: Tidak {Ya=0, Tidak=142}
| | jarak rumah = Jauh
| | | Penghasilan Ibu = Menengah: Tidak {Ya=0, Tidak=2}
| | | Penghasilan Ibu = Tidak Berpenghasilan: Ya {Ya=8, Tidak=0}
penghasilan ayah = Rendah
| jml saudara > 1.500
| | Penghasilan Ibu = Menengah
| | | jml saudara > 2.500: Ya {Ya=3, Tidak=1}
| | | jml saudara ≤ 2.500: Tidak {Ya=0, Tidak=2}
| | | Penghasilan Ibu = Rendah: Ya {Ya=5, Tidak=1}
| | | Penghasilan Ibu = Tidak Berpenghasilan: Ya {Ya=3, Tidak=0}
| | jml saudara ≤ 1.500: Tidak {Ya=2, Tidak=20}
penghasilan ayah = Tidak Berpenghasilan
| Penghasilan Ibu = Menengah
| | jml saudara > 1.500: Ya {Ya=6, Tidak=0}
| | jml saudara ≤ 1.500: Tidak {Ya=0, Tidak=3}
| | Penghasilan Ibu = Tidak Berpenghasilan: Ya {Ya=13, Tidak=0}
penghasilan ayah = Tinggi: Tidak {Ya=0, Tidak=55}

```

Gambar 2. Aturan Kalsifikasi

Hasil dari proses data mining diatas terdapat rules yang dapat di jadikan sebagai rekomendasi dalam menentukan Penerima Beasiswa yaitu sebagai berikut:

1. Jika Penghasilan Ayah = Menengah, Jml saudara > 2 maka hasilnya YA

2. Jika Penghasilan Ayah = Menengah, Jml saudara ≤ 2 , Jarak rumah = Dekat maka hasilnya Tidak
3. Jika Penghasilan Ayah = Menengah, Jml saudara ≤ 2 , Jarak rumah = Jauh, Penghasilan Ibu = Menengah maka hasilnya Tidak
4. Jika Penghasilan Ayah = Menengah, Jml saudara ≤ 2 , Jarak rumah = Jauh, Penghasilan Ibu = Tidak Berpenghasilan maka hasilnya Ya
5. Jika Penghasilan Ayah = Rendah, Jml saudara > 2 maka hasilnya YA
6. Jika Penghasilan Ayah = Rendah, Jml saudara ≤ 2 maka hasilnya Tidak
7. Jika Penghasilan Ayah = Rendah, Penghasilan Ibu = Rendah maka hasilnya YA
8. Jika Penghasilan Ayah = Rendah, Penghasilan Ibu = Tidak Berpenghasilan maka hasilnya YA
9. Jika Penghasilan Ayah = Rendah, Jml saudara ≤ 2 maka hasilnya Tidak
10. Jika Penghasilan Ayah = Tidak Berpenghasilan, Penghasilan Ibu = Menengah, Jml Saudara > 1 maka hasilnya Ya
11. Jika Penghasilan Ayah = Tidak Berpenghasilan, Penghasilan Ibu = Menengah, Jml Saudara ≤ 1 maka hasilnya Tidak
12. Jika Penghasilan Ayah = Tidak Berpenghasilan, Penghasilan Ibu = Tidak Berpenghasilan maka hasilnya Ya
13. Jika Penghasilan Ayah = Tinggi maka hasilnya Tidak

III. PEMBAHASAN

3.1 Perhitungan C4.5

Perhitungan ini digunakan untuk menunjukkan bagaimana proses menghitung entropy dan gain sehingga terbentuk suatu pohon keputusan. Kemudian menentukan akar pertama dari pohon. Akar pertama akan diambil dari atribut yang terpilih dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung

dahulu nilai Entropy dari masing-masing atribut. Perhitungan manual Algoritma C4.5 menggunakan sampel 20% dari dataset yang ada yaitu 75 data. Dari 75 data yang

digunakan jumlah kasus layak mendapatkan PIP sebanyak 22 data dan jumlah tidak layak mendapatkan PIP sebanyak 53 data. Berikut data sampel yang digunakan untuk perhitungan manual :

Tabel 3. Sampel Dataset

Penghasilan Ayah	Penghasilan Ibu	Pemegang KIP	Jml Saudara	Jarak Rumah	Layak PIP
Menengah	Tidak	Tidak	3	Dekat	Ya
Menengah	Berpenghasilan	Tidak	3	Dekat	Ya
Menengah	Tidak	Tidak	2	Jauh	Ya
Menengah	Berpenghasilan	Tidak	2	Dekat	Tidak
Menengah	Berpenghasilan	Tidak	1	Dekat	Tidak
Menengah	Menengah	Tidak	1	Dekat	Tidak
Menengah	Tidak	Tidak	1	Dekat	Tidak
Menengah	Berpenghasilan	Tidak	1	Dekat	Tidak
Menengah	Menengah	Tidak	2	Dekat	Tidak
Menengah	Menengah	Tidak	2	Jauh	Tidak
Tidak Berpenghasilan	Menengah	Tidak	2	Dekat	Ya
Menengah	Tidak	Tidak	3	Dekat	Ya
Menengah	Berpenghasilan	Tidak	3	Dekat	Tidak
Menengah	Menengah	Tidak	2	Dekat	Tidak
.....
Tinggi	Tidak	Tidak	2	Dekat	Tidak
	Berpenghasilan				

Pada tabel di atas kemudian menghitung nilai entropy dan gain. Berikut perhitungan nilai entropy dan nilai gain node pertama

Rumus Entropy:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Perhitungan Entropy Total:

$$Entropy(\text{total}) = (-\frac{22}{75} * \log_2(\frac{22}{75})) + (-\frac{53}{75} * \log_2(\frac{53}{75})) = 0,872988294$$

Entropy Penghasilan Ayah:

$$Entropy(\text{Total, Tidak Berpenghasilan}) = (-\frac{7}{8} * \log_2(\frac{7}{8})) + (-\frac{1}{8} * \log_2(\frac{1}{8})) = 0,543564443$$

Rumus Gain:

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

$$Gain(\text{Total, Penghasilan Ayah}) = 0,872988294 - ((\frac{8}{75} * 0,543564443) + (\frac{9}{75} * 0,918295834) + (\frac{55}{75} * 0,64293835) + (\frac{3}{75} * 0)) = 0,233324464$$

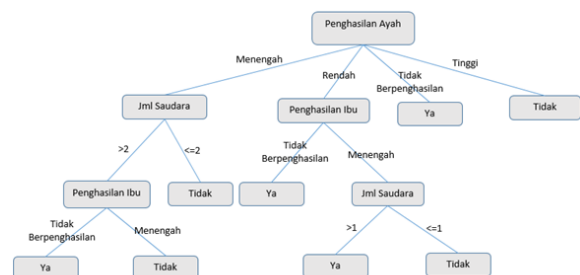
Gain(Total, Penghasilan Ibu) = 0,872988294 - (($\frac{48}{75} * 0,896038233 + (\frac{27}{75} * 0,825626526)$)) = 0,002298276

Dari contoh perhitungan tersebut diuraikan secara rinci dalam tabel berikut:

Tabel 4. Perhitungan Node 1

Node	Atribut	Jum Kasus	Ya	Tidak	Entropy	Gain
1	Total	75	22	53	0,872988294	0,233324464
Penghasilan Ayah	Tidak	8	7	1	0,543564443	0,002298276
	Berpenghasilan Rendah	9	6	3	0,918295834	
	Menengah	55	9	46	0,64293835	
	Tinggi	3	0	3	0	
Penghasilan Ibu	Tidak	48	15	33	0,896038233	0,003712807
	Berpenghasilan Menengah	27	7	20	0,825626526	
Pemegang KIP	Tidak	73	21	52	0,865693994	0,051048277
	Ya	2	1	1	1	
Jumlah Saudara	<=2	63	15	48	0,791858353	0,057282477
	>2	12	7	5	0,979868757	
Jarak Rumah ke Sekolah	Dekat	75	22	53	0,822404226	0,057282477
	Jauh	5	4	1	0,721928095	

Dari hasil tabel diatas diketahui bahwa atribut dengan Gain tertinggi adalah Penghasilan Ayah, yaitu 0,233324464. Setelah didapatkan nilai gain yang tertinggi maka selanjutnya dihitung dengan cara yang sama dengan menggunakan persamaan entropy untuk mendapatkan nilai entropy dan persamaan gain untuk mendapatkan nilai gain. Setelah dilakukan hasil perhitungan nilai entropy dan nilai gain, maka akan terbentuk pohon keputusan seperti gambar berikut:



Gambar 3. Pohon Keputusan Hitung Manual

Model yang telah dibentuk diuji tingkat akurasi dengan menggunakan confusion matrix sebagai acuan validitas aturan yang

telah dihasilkan. Tabel confusion matrix ditunjukkan pada tabel berikut ini:

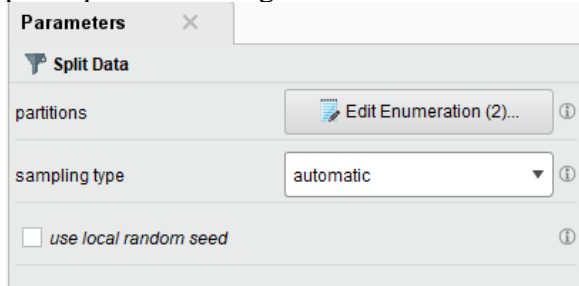
Tabel 5. Tabel Confusion Matrix

Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a (true positive – TP)	b (false negative – TN)
Class = No	c (false positive – FP)	d (true negative – TN)

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{TP} + \text{FN} + \text{FP}) * 100$$

$$\text{Error} = (\text{FP} + \text{FN}) / (\text{FP} + \text{FN} + \text{TN} + \text{TP}) * 100$$

Pengujian *confussion matrix* menggunakan data sampel sejumlah 75 data yang telah dilakukan perhitungan, pengujian ini dilakukan pada aplikasi Rapid Miner Studio dengan pengaturan metode *sampling type* pada *split data* sebagai berikut:



Gambar 4. Split data

Dengan aplikasi rapidminer dihasilkan nilai akurasi dan nilai error yang akan dijelaskan pada gambar dibawah ini:

accuracy: 81.08%

	true Ya	true Tidak	class precision
pred. Ya	7	4	63.64%
pred. Tidak	3	23	88.46%
class recall	70.00%	85.19%	

Gambar 5. Accuracy Algoritma C4.5

Gambar 5 adalah hasil pengukuran accuracy menggunakan rapidminer untuk algoritma c4.5. Selanjutnya akan dilakukan perhitungan manual untuk mencari nilai accuracy dan error dari hasil klasifikasi dari dataset.

$$\text{Accuracy} = (23 + 7) / (23 + 7 + 4 + 3) * 100 = 81,08\%$$

$$\text{Error} = (3 + 4) / (3 + 4 + 23 + 7) * 100 = 18,92\%$$

dari perhitungan diatas, diketahui bahwa accuracy dari hasil klasifikasi dataset penelitian ini adalah sebesar 81,08%, sedangkan tingkat error dari hasil hasil klasifikasi dataset pada penelitian adalah 18,92%. Perhitungan accuracy merupakan hasil dari perbandingan tingkat kedekatan antara nilai prediksi dengan nilai aktual [11]. Semakin tinggi kedekatan nilai prediksi dengan nilai aktual, maka semakin tinggi tingkat akurasi, untuk meningkatkan akurasi dari hasil klasifikasi bisa dengan menggunakan data yang lebih banyak agar menghasilkan aturan yang lebih akurat.

3.2 Tampilan Program

Berikut ini merupakan tampilan program yang telah dihasilkan :



Gambar 6. Tampilan Data Traning

Halaman data traning adalah halaman untuk melihat dan mengubah data siswa yang akan digunakan sebagai dataset untuk dilakukan perhitungan.



Gambar 7. Tampilan Proses Mining

Halaman Proses Mining adalah halaman untuk memastikan data yang akan di hitung sudah benar dan bisa dilanjutkan dengan klik tombol Proses, jika masih ada data yang salah bisa kembali ke halaman data traning untuk memperbaiki data yang salah.



Gambar 8. Hasil Perhitungan

Halaman hasil proses mining berfungsi untuk menampilkan jumlah data, jumlah kelas positif, jumlah kelas negatif, nilai entropy dan nilai gain dari masing-masing atribut serta menampilkan atribut yang terpilih dengan nilai gain tertinggi sebagai akar pertama setelah proses perhitungan



Gambar 9. Hasil Rule

Halaman Hasil Rule merupakan tampilan dari hasil analisa klasifikasi algoritma c4.5 kemudian terbentuk rules yang dapat digunakan untuk rekomendasi penerima beasiswa.



Gambar 10. Halaman Prediksi

Halaman Prediksi merupakan halaman yang digunakan user untuk melakukan proses prediksi penerima beasiswa berdasarkan aturan yang telah ditentukan sebelumnya oleh admin dalam pembentukan rules.



Gambar 11. Cetak Hasil Prediksi

Halaman merupakan halaman display history dari data yang pernah dilakukan proses prediksi

IV. SIMPULAN DAN SARAN

4.1 Simpulan

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan teknik data mining untuk mendapatkan aturan klasifikasi penerima beasiswa PIP SD Maria Mediatrix yang telah dijabarkan pada bab-bab sebelumnya, maka dapat ditarik kesimpulan, yaitu sebagai berikut :

1. Algoritma C4.5 dapat diimplementasikan kedalam sebuah aplikasi berbasis web yang berfungsi untuk mempermudah proses pemilihan beasiswa PIP, setelah menguji data set dengan rapid miner, lalu menghitung kembali data set dengan perhitungan manual maka didapat kan

hasil accuracy sebagai nilai ketentuan aplikasi dalam menentukan secara akurat siswa penerima beasiswa PIP pada SD Maria Mediatrix.

2. Analisa pemilihan beasiswa dari data siswa SD Maria Mediatrix dapat dihasilkan dengan menggunakan algoritma C4.5 pada metode klasifikasi, yaitu dengan cara menentukan atribut penentu yang digunakan untuk perhitungan entropy dan gain, dari perhitungan tersebut menghasilkan pohon keputusan dan aturan yang digunakan untuk klasifikasi penerima beasiswa PIP.
3. Hasil dari penelitian ini adalah dibuatnya sebuah aplikasi berbasis web yang dapat melakukan proses perhitungan C4.5 dan pembentukan aturan klasifikasi secara otomatis dari data siswa yang dimasukkan. Tersedia fitur prediksi penerima beasiswa PIP dimana pengguna aplikasi cukup menginput nama siswa, penghasilan ayah, penghasilan ibu, pemegang KIP, jumlah saudara, dan jarak rumah. Setelah itu aplikasi akan secara otomatis menganalisa siswa tersebut layak mendapat beasiswa atau tidak.

REFERENCES

- [1] C. Anam, “Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa,” *Magister Tek. Inform. Univ. AMIKOM Yogyakarta*, vol. 8, no. 1, pp. 13–19, 2018.
- [2] Suyanto, *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung: Bandung:Informatika, 2017.
- [3] E. Prasetyo, *DATA MINING – Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: C.V ANDI OFFSET, 2012.
- [4] Y. T. U. Heni Sulistiani, “Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa,” *Snti*, pp. 300–305, 2018.
- [5] N. Islamuddin, “Data Mining Untuk Seleksi Kelayakan Calon Penerima Beasiswa di STMIK Bina Bangsa Kendari,” *Simkom*, vol. 4, no. 2, pp. 9–20, 2019, doi: 10.51717/simkom.v4i2.30.
- [6] M. Tabrani, “Klasifikasi Penerima Beasiswa Kopertis Dengan,” *Pilar Nusa Mandiri J. Comput. Inf. Syst.*, vol. 12, no. 1, pp. 72–80, 2016, [Online]. Available: <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/261>
- [7] a fahrudin, l listiyoko, p surya, and ..., “Prediksi Peringkat Kelulusan Mahasiswa Untuk Menentukan Strategi Pemasaran Kampus Menggunakan Pohon Keputusan,” *Pros. Semin. ...*, no. November, 2017, [Online]. Available: <http://ejournal.lppmunsera.org/index.php/senasset/article/view/459/522>
- [8] P. Rahmadya, T. H; Herlawanti, *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains, 2013.
- [9] D. Feblian and D. U. Daihani, “Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X,” *J. Tek. Ind.*, vol. 6, no. 1, pp. 1–12, 2017, doi: 10.25105/jti.v6i1.1526.
- [10] Karsito and S. Susanti, “Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia,” *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.
- [11] T. B. Sasongko, “Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM,” *J. Tek. Inform. dan Sist. Inf.*, vol. 2, no. 2, pp. 244–253, 2016.

BIOGRAPHY

Eko Budiarto lahir di Tangerang pada tanggal 19 April 1997. Menyelesaikan pendidikan Strata I (S1) pada tahun 2022 pada Program Studi Teknik Informatika di Universitas Buddhi Dharma. Saat ini bekerja sebagai staff Tata Usaha disekolah Maria Mediatrix.

Rino S.Kom., M.Kom, Saat ini bekerja sebagai dosen Tetap dan Kaprodi pada Program Studi Teknik Informatika di Universitas Buddhi Dharma.

Susanto Hariyanto, Saat ini bekerja sebagai dosen Tetap pada Program Studi Teknik Informatika di Universitas Buddhi Dharma

Dera Susilawati, Saat ini bekerja sebagai dosen Tetap pada Program Studi Teknik Informatika di Universitas Buddhi Dharma