

Penerapan Data Mining Klasifikasi C4.5 Pada Penerima Beasiswa di SMK Swasta Anak Bangsa

Millah Sari, Agus Perdana Windarto, Harly Okprana

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: ¹millahsari99@gmail.com, ^{2,*}agus.perdana@amiktunasbangsa.ac.id, ³harly@amiktunasbangsa.ac.id

INFORMASI ARTIKEL

Article History

Received : Mar 31, 2021

Accepted : Mar 31, 2021

Published : Mar 11, 2021

KORESPONDENSI

Email:

agus.perdana@amiktunasbangsa.ac.id

ABSTRAK

Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Atribut yang digunakan dalam penelitian ini ada 5 (lima) yaitu: Nilai, Kehadiran, Semester, Penghasilan Orang Tua (PO), dan Jumlah Tanggungan Orang Tua (JTO). Berdasarkan pengolahan data menggunakan *software RapidMiner 5.3.0.0* didapat nilai akurasi sebesar 92,70%, artinya bahwa rule yang dihasilkan tingkat kebenarannya mendekati 100%. Dimana hasil dari *class precision* label Layak sebesar 92,05% dan label Tidak Layak sebesar 93,24%. Sesuai dengan ketentuan tersebut maka hasil perhitungan manual dengan pengujian *RapidMiner* menghasilkan sebanyak 9 model aturan atau *rule* Penerima Beasiswa. Artinya hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 dan *Rapidminer* diperoleh hasil yang sama dan sesuai. Sehingga pengujian dengan *RapidMiner* dapat dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus penerima Beasiswa.

Kata Kunci: Beasiswa; Data Mining; Algoritma C4.5

ABSTRACT

Data Mining is a series of processes to explore added value in the form of knowledge that has not been known manually from a data set. There are 5 (five) attributes used in this study, namely: Value, Attendance, Semester, Parents' Income (PO), and Number of Dependent Parents (JTO). Based on data processing using *Rapid Miner 5.3.0.0* software, an accuracy value of 92.70% is obtained, meaning that the resulting rule is close to 100% correct. Where the results of the feasible precision label class are 92.05% and the inappropriate label is 93.24%. In accordance with these provisions, the results of manual calculations by *Rapid Miner* testing produce 9 models of rules or rules for Scholarship Recipients. This means that the results of the process carried out by researchers on the calculation of the C4.5 Algorithm and *Rapidminer* obtained the same and appropriate results. So that testing with *Rapid Miner* can be said to be successful and can find a decision tree in the case of Scholarship recipients.

Keywords: Scholarship; Data Mining; Algoritma C4.5

1. PENDAHULUAN

Mendapatkan pendidikan adalah merupakan hak asasi sekaligus kewajiban bagi setiap warga negara Indonesia sebagaimana yang tercantum dalam UUD 1945 Pasal 31 ayat 1-2[1]. Beasiswa merupakan penghargaan yang diberikan kepada siswa berprestasi dalam studi yang ditempuhnya. Beasiswa diberikan dengan harapan dapat menumbuhkan dan meningkatkan semangat siswa dalam belajar dan mampu membantu agar orang tua tidak perlu memikirkan biaya sekolah. Penelitian ini dilakukan di SMK Swasta Anak Bangsa yang beralamat di Jl. Utama Bandar siantar kec. Gunung Malela Kab.Simalungun. SMK Swasta Anak Bangsa memiliki kurang lebih 137 siswa, dari jumlah siswa yang di rekomendasikan dalam pemberian beasiswa.

Dalam proses penyeleksian siswa yang berhak mendapatkan beasiswa membutuhkan waktu yang lama, hal tersebut dikarenakan banyaknya calon siswa yang mengajukan dan kuota penerimaan yang terbatas. Sedangkan penerimaan beasiswa harus tepat sasaran sesuai dengan kriteria yang berlaku. Dalam kondisi ini peneliti menemukan kesalahan dalam penentuan calon penerimaan beasiswa pada SMK Swasta Anak Bangsa seperti terpilihnya penerima beasiswa yang kurang tepat, maka dari itu diperlukan klasifikasi untuk membantu menentukan keputusan penerimaan beasiswa. dalam penentuan beasiswa terkadang mengakibatkan sulitnya membuat keputusan dalam menentukan calon penerima beasiswa. Berdasarkan penelitian terdahulu yang dilakukan oleh: [2] dengan judul "algoritma C4.5 untuk menentukan mahasiswa penerima beasiswa" pada penelitian sebelumnya hasil

klasifikasinya menggunakan 3 (tiga) kriteria yaitu: IPK, pekerjaan dan masa kerja yang telah ditentukan sebelumnya, sedangkan dalam penelitian yang peneliti lakukan memiliki 5 (lima) kriteria yaitu: Nilai, Kehadiran, Semester, Penghasilan Orang Tua (PO), Jumlah Tanggungan Orang Tua (JTO).

Data mining adalah teknik atau cara yang digunakan untuk mengobservasi data menyeluruh secara otomatis. *Data mining* berkonotasi sebagai pencarian informasi bisnis berharga dari sumber data yang besar. [3][4] Klasifikasi data adalah sebuah proses yang menemukan property yang sama pada sebuah himpunan objek dalam sebuah basis data dan di klasifikasikan ke dalam kelas yang berbeda menurut model klasifikasi yang ditetapkan. Klasifikasi memiliki tujuan utama yaitu untuk menemukan model yang membedakan atribut kedalam kategori yang sesuai, kemudian model tersebut digunakan untuk mengklasifikasi atribut yang kelasnya belum diketahui. [5][6][7] Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal [8] [9] [10].

Salah satu metoda Data Mining yang umum digunakan adalah decision tree. Decision tree adalah struktur flowchart yang mempunyai tree(pohon), di mana setiap simpul internal menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas [11] [12][13]. Berdasarkan masalah diatas peneliti mengambil metode C4.5 dari cabang bidang keilmuan *data mining*. *Data mining* berkonotasi sebagai pencarian informasi bisnis yang berharga dari sumber data yang besar. Algoritma C4.5 merupakan pengembangan algoritma ID3 yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan cara memprediksi atau mengklarifikasi yang sangat kuat karena pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. *Algoritma* C4.5 akan mengubah data menjadi pohon keputusan dan aturan-aturan keputusan[14].

2. METODOLOGI PENELITIAN

2.1 Pengumpulan Data

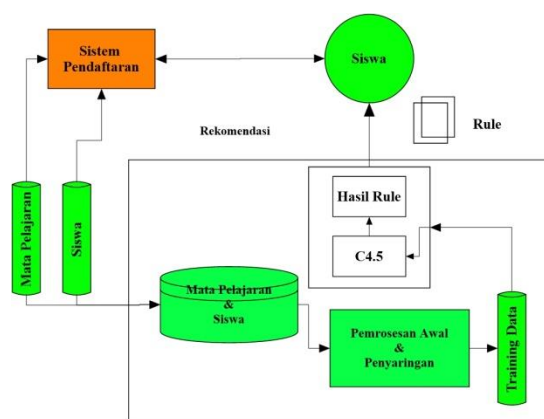
Tujuan dari Skripsi ini adalah untuk membuat model penerima beasiswa menggunakan *algoritma* C4.5 untuk meningkatkan kemauan belajar siswa menjadi lebih baik dengan menyediakan data penerima beasiswa yang dapat digunakan sebagai pedoman analisis dalam pembuatan keputusan. Pada bagian ini kita mulai dengan menggambarkan studi kasus *data mining* pada sistem penerima beasiswa bagaimana mengumpulkan data yang dapat digunakan pada penelitian ini. Data dikumpulkan dari database pendidikan dan mensurvei siswa yang telah menempuh semester 2 sampai dengan tahun 2020 di SMK Swasta Anak Bangsa. Instrumen penelitian yang digunakan harus mempunyai ukuran yang akurat. Secara terperinci, bagaimana mendapatkan *input* yang lebih baik dalam proses *data mining* yang digambarkan pada bagian sebelum pemrosesan data. Penulis memberikan tinjauan singkat dari beberapa *analysis* data yang digunakan pada penelitian ini. Rancangan penelitian ini pertama kali dilakukan dengan melakukan pengamatan (*observasi*) untuk mempelajari klasifikasi data Nilai siswa. Hasil pengamatan kemudian dibuat percobaan yang mendukung, selanjutnya dilakukan eksperimen data dengan menggunakan rapidminer yang merupakan *software open source* untuk membuat model aturan data yang diambil dari kumpulan nilai SMK Swasta Anak Bangsa dan data formulir siswa. Hasil dari eksperimen data ini merupakan pengembangan dari ilmu pengetahuan yang nantinya dapat merupakan masukan bagi pemecahan masalah yang ada di lembaga pendidikan SMK Swasta Anak Bangsa. Dalam melakukan klasifikasi data dari kumpulan nilai SMK Swasta Anak Bangsa penulis melakukan tahapan, antara lain :

- Mengambil data dari kumpulan nilai, seperti pada Gambar 1

LEGER NILAI KELAS X RPL 2																				
T.P 2019/2020																				
Nama Sekolah		: SMK SWASTA ANAK BANGSA																		
Kelas		: XI RPL 2																		
No	Nama Siswa	Apk	Mata Pelajaran Muatan Nasional															Nilai	Jumlah	Rata-rata
			Pendidikan agama Islam	PPKN	B. Indonesia	Matematika	B. Inggris	PAK	Pendidikan Prangkat Lunak	Basis Data	Penggunaan Berbasis Objek	Penggunaan Web Prangkat Bekerak	Produk Kreatif dan Kewirausahaan							
kkkn	70	70	70	70	70	70	75	75	75	75	75									
	A	H	A	H	A	H	A	H	A	H	A	H	A	H						
1	Andika Agustian	KI3	86 B	84 B	83 B	81 B	81 B	85 B	81 C	85 B	83 B	80 C	83 B	912	1819	60.6333				
	KI4	85 B	83 B	80 C	82 B	80 C	85 B	81 C	85 B	83 B	81 C	82 C	907							
2	Akbar Maulana	KI3	85 B	86 B	84 B	81 B	82 B	86 B	82 C	82 B	85 B	85 B	84 B	922	1841	61.3667				
	KI4	86 B	85 B	85 B	82 B	81 B	83 B	81 C	83 B	85 B	85 B	83 B	919							
3	Arjun Priadi	KI3	85 B	84 B	82 B	83 B	82 B	86 B	80 C	76 C	83 B	81 C	83 B	905	1801	60.0333				
	KI4	85 B	81 B	81 B	84 B	80 C	85 B	80 C	76 C	81 C	80 C	83 B	896							
4	Beni Zulfahri	KI3	82 B	85 B	83 B	84 B	81 B	85 B	81 C	81 B	84 B	81 C	84 B	911	1815	60.5				
	KI4	81 B	81 B	82 B	84 B	81 B	85 B	80 C	85 B	81 C	81 C	83 B	904							
5	Dedek Warasto	KI3	82 B	83 B	84 B	77 C	81 B	84 B	80 C	82 B	82 C	78 C	83 B	896	1800	60				
	KI4	83 B	83 B	85 B	80 C	81 B	83 B	80 C	83 B	83 B	80 C	83 B	904							
6	Deni Andrian	KI3	83 B	84 B	84 B	81 B	81 B	83 B	80 C	81 B	83 B	80 C	84 B	904	1805	60.1667				
	KI4	83 B	83 B	85 B	80 C	81 B	84 B	79 C	80 C	83 B	79 C	84 B	901							

Gambar 1. Record dari Kumpulan Nilai

- Membuat beberapa kumpulan nilai data yang berbentuk baris menjadi kolom.
- Melakukan *Data Transformation Service* (DTS) dari beberapa excel diubah ke 1 bentuk excel (.xls).
- Untuk mendapatkan aturan dari data rekomendasi beasiswa yang diambil dari sekumpulan nilai penulis menggunakan program rapidminer dengan rancangan penelitiannya dapat dilihat pada Gambar 2.2



Gambar 2. Rancangan Penelitian Aturan Rekomendasi Beasiswa

Pada Gambar 2 dapat dijelaskan urutan pertama siswa melakukan sistem pendaftaran ke Sekolah Menengah Kejuruan, kemudian siswa mengikuti ujian berdasarkan matapelajaran yang diambil setiap semester. Hasil ujian disimpan kedalam database, data nilai hasil ujian diklasifikasikan dengan melakukan penyaringan data kemudian data yang sudah konsisten dilakukan training data, apabila sampel data sudah *valid* maka dilakukan pembuatan model aturan dengan algoritma C 4.5. Hasil dari model aturan ini menjadi sebuah rekomendasi bagi pihak manajemen Sekolah Menengah Kejuruan Swasta (SMKS).

2.2 Validitas dan Reliabilitas (Keakuratan) Data

Keakuratan instrumen yang digunakan pada penelitian ini adalah penting. Keakuratan mengacu apakah instrumen yang digunakan mengukur secara konsisten setiap waktu dan populasi (Gall et al., 1996).

Survei dalam studi ini diuji dalam jangka waktu dan ukuran internal yang terpercaya yang memiliki keterkaitan antara bagian bagian tes (Brown and Alexander, 1991). Hal ini menjamin apakah pengukuran instrumen secara akurat dimaksudkan untuk mengukur.

2.3 Preprocessing Data

Untuk mendapatkan *input* yang lebih baik dari teknik *data mining*, penulis melakukan beberapa *preprocessing* untuk data yang akan dikumpulkan. Data disusun dalam table yang berbeda yang dihubungkan dalam tabel tunggal. Tabel yang digabungkan adalah :

- tabel Siswa
- tabel Nilai
- tabel Kehadiran

Dari ke 4 tabel di atas, dipilih data yang digunakan untuk data penelitian yang terdiri dari data Nilai, RataMN, RataMK, RataC1, RataC2, RataC3, Kelas.

2.4 Alat Analisis Data

RapidMiner adalah aplikasi perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (machine learning), pembelajaran mendalam (deep learning), penambangan teks (text mining), dan analisis prediktif (predictive analytics). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta digunakan untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan [15]. Rapidminer 5.3 digunakan pada studi ini untuk menampilkan kelompok siswa pada kumpulan data dan memperlihatkan matriks presentasi yang tersebar dari kelompok data.

2.5 Instrument Penelitian

Dalam penelitian ini, penulis melakukan beberapa dataset file dalam bentuk *spreadsheet* file excel 2007 (xls). Transformasi data ini diperlukan sebagai masukan untuk perangkat lunak yang akan digunakan dalam penelitian ini. Dalam pengolahan data transformasi, penulis menggunakan perangkat lunak rapidminer versi 5.3 Instrumen penelitian.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian yang sesuai dengan pertanyaan-pertanyaan yang diajukan pada permulaan. Penelitian ini dilaksanakan pada satu data set yaitu penulis menggunakan dataset Nilai yang diambil dari database DAPODIK SMK Swasta Anak Bangsa. Data set bersifat nominal yang terdiri dari kehadiran, semester, Penghasilan

Orang Tua (PO) dan Jumlah Tanggungan Orang Tua (JTO). Kemudian data ditransformasikan ke format data excel 2003.

Kumpulan data yang diperoleh digunakan sebagai contoh sumber input untuk membuat model aturan menggunakan algoritma C4.5 dengan *software Rapidminer*. Dalam percobaan ini, penulis menampilkan gambar pemetaan model aturan keterhubungan data nilai menggunakan algoritma C4.5 untuk menganalisis Penerima Beasiswa. Pohon Keputusan digunakan untuk membuat model aturan yang akan dipilih dalam mengambil keputusan pada bagian manajemen sekolah.

3.1 Pengolahan Data

Untuk menentukan penerima beasiswa menggunakan metode algoritma C4.5 maka data yang digunakan dan dianalisa yaitu data siswa di SMK Swasta Anak Bangsa, dimana diperoleh dari data DAPODIK yaitu: Nilai Raport, Kehadiran, Semester, Penghasilan Orang Tua (PO), Jumlah Tanggungan Orang Tua (JTO). Kelima Atribut predicator untuk menghasilkan atribut target, dimana atribut target tersebut menjadi class output untuk penerima beasiswa yaitu dibedakan menjadi 2 class (layak dan tidak layak).

3.1.1 Proses Perhitungan Algoritma C4.5

Perhitungan Algoritma C4.5 untuk memperoleh model aturan pohon keputusan dapat diuraikan sebagai berikut :

Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk keputusan Layak, jumlah kasus untuk keputusan Tidak Layak.

Langkah 2: Menghitung *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan kelas atribut dengan persamaan (1). Selanjutnya dilakukan penghitungan *Gain* untuk masing-masing atribut dengan persamaan (2). Berikut ini adalah perhitungan nilai *entropy* dan *gain*.

Menghitung *entropy* total :

$$\text{Entropy [Total]} = \left(-\frac{63}{137} \times \log_2 \left(\frac{63}{137} \right) \right) + \left(-\frac{74}{137} \times \log_2 \left(\frac{74}{137} \right) \right)$$

$$= 0,995344607$$

Menghitung *entropy* dan *gain* Nilai:

$$\text{Entropy [Nilai – Tinggi]} = 0$$

$$\text{Entropy [Nilai – Sedang]} = \left(-\frac{45}{80} \times \log_2 \left(\frac{45}{80} \right) \right) + \left(-\frac{35}{80} \times \log_2 \left(\frac{35}{80} \right) \right)$$

$$= 0,988699408$$

$$\text{Entropy [Nilai – Rendah]} = \left(-\frac{5}{44} \times \log_2 \left(\frac{5}{44} \right) \right) + \left(-\frac{39}{44} \times \log_2 \left(\frac{39}{44} \right) \right)$$

$$= 0,5107878230,995344607 - \left(\left(\frac{13}{137} \times 0 \right) + \left(\frac{80}{137} \times 0,988699408 \right) + \left(\frac{44}{137} \times 0,510787823 \right) \right) =$$

$$0,253953243$$

Gain [Minat siswa]

Menghitung *entropy* dan *gain* cara Kehadiran:

$$\text{Entropy [Kehadiran – Hadir]} = \left(-\frac{63}{137} \times \log_2 \left(\frac{63}{137} \right) \right) + \left(-\frac{74}{137} \times \log_2 \left(\frac{74}{137} \right) \right)$$

$$= 0,995344607$$

$$\text{Entropy [Kehadiran – Tidak Hadir]} = 0$$

$$\text{Gain [Kehadiran]} = 0,995344607 - \left(\left(\frac{137}{137} \times 0,995344607 \right) \right) = 0$$

Menghitung *entropy* dan *gain* Semester :

$$\text{Entropy [Semester - Dua]} = \left(-\frac{7}{46} \times \log_2 \left(\frac{7}{46} \right) \right) + \left(-\frac{39}{46} \times \log_2 \left(\frac{39}{46} \right) \right)$$

$$= 0,615253891$$

$$\text{Entropy [Semester – Empat]} = \left(-\frac{38}{59} \times \log_2 \left(\frac{38}{59} \right) \right) + \left(-\frac{21}{59} \times \log_2 \left(\frac{21}{59} \right) \right)$$

$$= 0,939254721$$

$$\text{Entropy [Semester – Enam]} = \left(-\frac{18}{32} \times \log_2 \left(\frac{18}{32} \right) \right) + \left(-\frac{14}{32} \times \log_2 \left(\frac{14}{32} \right) \right)$$

$$0,995344607 - \left(\left(\frac{46}{137} \times 0,615253891 \right) + \left(\frac{59}{137} \times 0,939254721 \right) + \left(\frac{32}{137} \times 0,988699408 \right) \right) =$$

$$0,153329362 = 0,988699408$$

$$\text{Gain [Semester]} =$$

Menghitung *entropy* dan *gain* Penghasilan Orang Tua (PO) :

$$\text{Entropy [PO – Tinggi]} = \left(-\frac{1}{7} \times \log_2 \left(\frac{1}{7} \right) \right) + \left(-\frac{6}{7} \times \log_2 \left(\frac{6}{7} \right) \right)$$

$$= 0,591672779$$

$$\text{Entropy [PO – Sedang]} = \left(-\frac{8}{41} \times \log_2 \left(\frac{8}{41} \right) \right) + \left(-\frac{33}{41} \times \log_2 \left(\frac{33}{41} \right) \right)$$

= 0,712064055

$$\begin{aligned} \text{Entropy [PO - Rendah]} &= \left(-\frac{54}{89} \times \log_2 \left(\frac{54}{89} \right) \right) + \left(-\frac{35}{89} \times \log_2 \left(\frac{35}{89} \right) \right) \\ 0,995344607 - \left(\left(\frac{46}{137} \times 0,591572779 \right) + \left(\frac{59}{137} \times 0,939254721 \right) + \left(\frac{32}{137} \times 0,988699408 \right) \right) &= 0,123900955 \\ 0,966870165 \\ \text{Gain [PO]} &= \end{aligned}$$

Menghitung *entropy* dan *gain* Jumlah Tanggungan Orang Tua (JTO) :

$$\begin{aligned} \text{Entropy [JTO - Banyak]} &= 0 \\ \text{Entropy [JTO - Sedang]} &= \left(-\frac{29}{40} \times \log_2 \left(\frac{29}{40} \right) \right) + \left(-\frac{11}{40} \times \log_2 \left(\frac{11}{40} \right) \right) = 0,848548178 \\ \text{Entropy [JTO - Sedikit]} &= \left(-\frac{34}{97} \times \log_2 \left(\frac{34}{97} \right) \right) + \left(-\frac{63}{97} \times \log_2 \left(\frac{63}{97} \right) \right) = 0,934527566 \\ 0,995344607 - \left(\left(\frac{0}{137} \times 0 \right) + \left(\frac{40}{137} \times 0,848548178 \right) + \left(\frac{97}{137} \times 0,934527566 \right) \right) &= 0,085920512 \text{Gain [JTO]} \end{aligned}$$

3.1.2 Aturan Rules

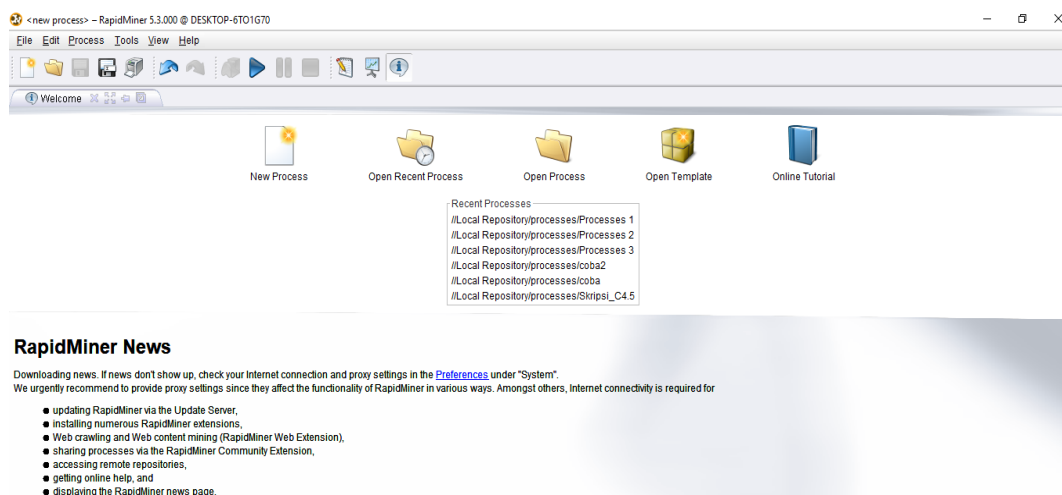
Dari perhitungan diatas terdapat 9 (Sembilan) *rules* yang dapat di jadikan sebagai referensi dalam menentukan Penerima Beasiswa. Adapun aturan atau *rule* yang terbentuk berdasarkan pohon keputusan yaitu 5 (empat) *rules* keputusan layak dan 4 (enam) *rules* keputusan tidak layak dijelaskan melalui teks narasi sebagai berikut:

- 1) Jika Nilai = Tinggi, Maka hasilnya Layak {Layak = 13, Tidak Layak = 0}
- 2) Jika Nilai = Sedang dan Penghasilan Orang Tua (PO) =Tinggi, maka hasilnya Tidak Layak {Layak = 0, Tidak Layak = 4}
- 3) Jika Nilai = Sedang dan Penghasilan Orang Tua (PO) =Rendah, maka hasilnya Layak dan Tidak Layak {Layak = 40, Tidak Layak = 10}
- 4) Jika Nilai = Sedang, Penghasilan Orang Tua (PO) = Sedang dan Jumlah Tanggungan Orang tua (JTO) = Sedang, maka hasilnya Layak dan Tidak Layak {Layak = 5, Tidak Layak = 5}
- 5) Jika Nilai = Sedang, Penghasilan Orang Tua (PO) = Sedang dan Jumlah Tanggungan Orang tua (JTO) = Sedikit, maka Tidak Layak {Layak = 0, Tidak Layak = 16}
- 6) Jika Nilai = Rendah, Jumlah Tanggungan Orang Tua (JTO) = Sedikit, maka hasilnya Tidak Layak {Layak = 0, Tidak Layak = 35}
- 7) Jika Nilai = Rendah, Jumlah Tanggungan Orang Tua (JTO) = Sedang, Penghasilan Orang Tua (PO) = Sedang, maka hasilnya Layak dan Tidak Layak { Layak = 1, Tidak Layak = 1}
- 8) Jika Nilai = Rendah, Jumlah Tanggungan Orang Tua (JTO) = Sedang, Penghasilan Orang Tua (PO) = Rendah, Semester = Dua, maka hasilnya Tidak Layak { Layak = 0, Tidak Layak = 3}
- 9) Jika Nilai = Rendah, Jumlah Tanggungan Orang Tua (JTO) = Sedang, Penghasilan Orang Tua (PO) = Rendah, Semester = Empat, maka hasilnya Layak { Layak = 4, Tidak Layak = 0}

3.2 Hasil Percobaan

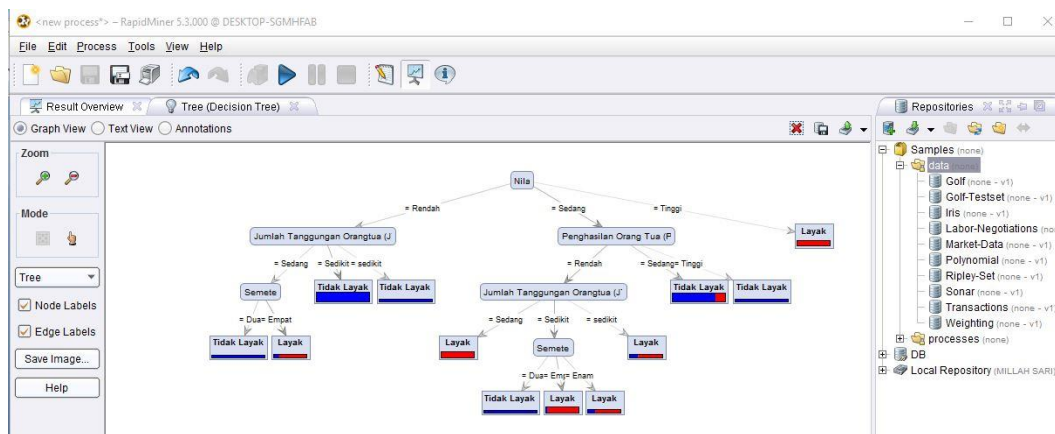
Pengujian terhadap hasil perhitungan manual menggunakan *software RapidMiner* dilakukan melalui beberapa tahapan proses yaitu :

- a) Langkah pertama membuka *software RapidMiner*, dimana tampilan utamanya dapat kita lihat pada Gambar 3



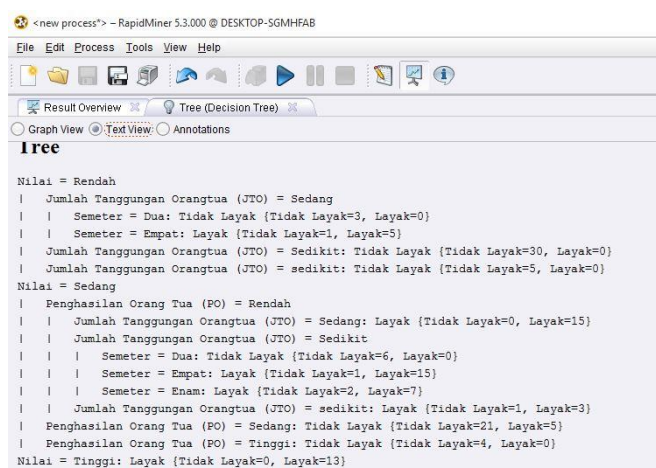
Gambar 3. Tampilan Utama *Rapidminer 5.3*

- b) Langkah kedua pilih *New Process* untuk membuat proses baru dan melakukan pengujian data. Berikut ini adalah tampilan *New Process* dan hasil pengolahan data dengan model pohon keputusan sesuai dengan *software RapidMiner* dapat dilihat pada Gambar 4 sebagai berikut :



Gambar 4. Decision Tree Pada Rapidminer

Gambar 4 diatas merupakan pohon keputusan yang dihasilkan pada *Rapidminer* dengan aturan atau *rule* yang dapat dilihat pada text view pada Gambar 5. berikut :



Gambar 5. Rule Decision Tree Pada Rapidminer

Hasil penerapan Algoritma C4.5 menggunakan *software RapidMiner* dengan operator *Split Validation* diperoleh nilai akurasi yaitu sebesar 92,70%. Hasil akurasi tersebut diperoleh dengan pengaturan pada operator *split validation* dengan nilai *split ratio* = 0,5 dan *sampling type* = *linear sampling*. Berikut ini adalah hasil akurasi yang diperoleh.

Multiclass Classification Performance Annotations			
Table View Plot View			
accuracy: 92.70%			
	true Tidak Layak	true Layak	class precision
pred. Tidak Layak	69	5	93.24%
pred. Layak	5	58	92.06%
class recall	93.24%	92.06%	

Gambar 6. Nilai Akurasi Algoritma C4.5

PerformanceVector			
PerformanceVector:			
accuracy: 92.70%			
ConfusionMatrix:			
True:	Tidak Layak	Layak	
Tidak Layak:	69	5	
Layak:	5	58	

Gambar 7. Performance Vektor Algoritma C4.5

Dari gambar diatas nilai *Accuracy* sebesar 92,70%, artinya aturan atau *rule* yang dihasilkan mendekati 100%, dimana untuk *Class Precision* pada prediksi label Layak sebesar 92,05% dan prediksi label Tidak Paham sebesar 93,24%.

Hasil yang dilakukan peneliti dalam perhitungan Algoritma C4.5 diperoleh 9 model aturan atau *rule* Penerima beasiswa. Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sebesar 92,70%, artinya bahwa *rule* yang dihasilkan tingkat kebenarannya mendekati 100%. Dimana hasil dari *class*

precision label Layak sebesar 92,05% dan label Tidak Layak sebesar 93,24%. *Parameters* yang digunakan pada *decision tree* telah dilakukan penyesuaian terhadap kriteria *decision tree* yang digunakan pada *RapidMiner* yaitu *information gain*, *maksimal depth*= 20, *confidance*= 0,25, *minimal gain*= 0,1, *minimal leaf size*= 1, *minimal size for split*= 2, *number of pruning alt*= 3. Sesuai dengan ketentuan tersebut maka hasil perhitungan manual dengan pengujian *RapidMiner* menghasilkan sebanyak 9 model aturan atau *rule* Penerima Beasiswa. Artinya hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 dan *Rapidminer* diperoleh hasil yang sama dan sesuai. Sehingga pengujian dengan *RapidMiner* dapat dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus penerima Beasiswa.

4. KESIMPULAN

Implementasi Algoritma C4.5 untuk menentukan Kelayakan penerima beasiswa pada SMK Swasta Anak Bangsa telah berhasil dilakukan. Variable yang memiliki prioritas utama terhadap prediksi penerima beasiswa adalah siswa yang memiliki nilai tinggi maka siswa tersebut dinyatakan dengan status Layak. Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sebesar 92,70%, artinya bahwa *rule* yang dihasilkan tingkat kebenarannya mendekati 100%. Dimana hasil dari *class precision* label Layak sebesar 92,05% dan label Tidak Layak sebesar 93,24%. *Parameters* yang digunakan pada *decision tree* telah dilakukan penyesuaian terhadap kriteria *decision tree* yang digunakan pada *RapidMiner* yaitu *information gain*, *maksimal depth*= 20, *confidance*= 0,25, *minimal gain*= 0,1, *minimal leaf size*= 1, *minimal size for split*= 2, *number of pruning alt*= 3. Sesuai dengan ketentuan tersebut maka hasil perhitungan manual dengan pengujian *RapidMiner* menghasilkan sebanyak 9 model aturan atau *rule* Penerima Beasiswa. Artinya hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 dan *Rapidminer* diperoleh hasil yang sama dan sesuai. Sehingga pengujian dengan *RapidMiner* dapat dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus penerima Beasiswa. Dengan melihat hasil pengujian tersebut dapat diambil sebuah kesimpulan yaitu pengujian Data Mining Algoritma C4.5 dapat digunakan dalam menentukan kelayakan penerima Beasiswa di SMK Swasta Anak Bangsa.

REFERENCES

- [1] N. Hijriana and M. Rasyidan, "Penerapan Metode Decision Tree Algoritma C4 . 5 Untuk Seleksi," *Sains Dan Teknol.*, vol. 3, pp. 9–13, 2017.
- [2] M. A. Rahman, "Algoritma C45 Untuk Menentukan Mahasiswa Penerima Beasiswa (Studi Kasus : Pps Iain Raden Intan Bandar Lampung)," *J. Teknol. Inf. Magister Darmajaya*, vol. 1, no. 02, pp. 118–128, 2015.
- [3] A. Khoerunnisa, B. Irawan, and M. R. Rumani, "Analisis dan implementasi perbandingan algoritma c.45 dengan naïve bayes untuk prediksi penawaran produk," *E-Proceeding Eng.*, vol. 3, no. 3, pp. 5029–5035, 2016.
- [4] Tumini and L. Damayanti, "Aplikasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naive Bayes Studi Kasus SMP Negeri 11 Kotabumu Utara," *Inform. SIMANTIK*, vol. 3, no. 2, pp. 23–30, 2018.
- [5] N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu)," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 11–22, 2018, doi: 10.24176/simet.v9i1.1627.
- [6] H. Okprana, M. R. Lubis, and J. T. Hadinata, "Prediksi Kelulusan TOEFL Menggunakan Metode Resilient Backpropagation," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 275, 2020, doi: 10.26418/jp.v6i2.41224.
- [7] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [8] D. Maurina and A. Z. Fanano, "Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada Sma Muhammadiyah Gubug Menggunakan Algoritma C4.5," *J. UDINUS*, pp. 5–6, 2015.
- [9] J. H. Jaman and N. I. P. Astuti, "Penerapan Algoritma C4.5 untuk Penentuan Penerima Beasiswa (Studi Kasus: SDN Karawang Kulon III)," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, pp. 25–29, 2018, doi: 10.36805/technoxplore.v3i1.796.
- [10] A. S. Suweleh, D. Susilowati, and U. Bumigora, "Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4 . 5 Jurnal BITE: Jurnal Bumigora Information Technology Jurnal BITE: Jurnal Bumigora Information Technology," vol. 2, no. 1, pp. 12–21, 2020, doi: 10.30812/bite.v2i1.798.
- [11] Rismayanti, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan," *Media Infotama*, vol. 12, no. 2, pp. 116–120, 2016.
- [12] G. R. P. A. P. Windarto, E. Irawan, and W. Saputra, "Penerapan Data Mining Menggunakan Algoritma C4 . 5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS," *Pros. Semin. Ris. dan Inf. sains*, vol. 2, pp. 376–385, 2020.
- [13] R. Winanjaya, F. Amir, and R. Doni, "Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 1, 2019, doi: 10.30645/senaris.v1i0.1.
- [14] R. Nofitri and N. Irawati, "ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN PENDAHULUAN Penerapan teknologi informasi saat ini berkembang begitu pesat . Salah satunya penerapan teknologi yang dapat diterapkan di dunia industri yaitu untuk evaluasi terhadap kinerja perusahaan . Evaluasi me," vol. V, no. 2, pp. 199–204, 2019.
- [15] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.