

PENERAPAN KLASIFIKASI ALGORITMA DATA MINING C4.5 UNTUK MEMREDIKSI TINGKAT KELULUSAN SISWA DI LEMBAGA PELATIHAN KERJA SHINJU SEMARANG

Fifi Maqfiroh¹, Sri Mujiyono²

^{1,2}Universitas Ngudi Waluyo, Prodi Teknik informatika, Kabupaten Semarang
Email : ¹indahfafa6@gmail.com, ²mujiyn80@gmail.com

Abstrak

LPK. SHINJU telah menyimpan data-datanya dalam database berupa hardcopy dan softfile. Data-data tersebut jika dimaksimalkan pemanfaatannya dapat memberikan informasi yang berguna, salah satunya adalah prediksi kelulusan siswa. Penelitian bertujuan menerapkan klasifikasi algoritma C4.5 dalam prediksi kelulusan siswa. Data yang digunakan yaitu data alumni siswa yang telah lulus tahun 2020. Atribut yang dipakai adalah Jenis Kelamin, Tempat Tinggal, Asal kelulusan, Status Bekerja, Ekonomi dan Nilai Akhir. Atribut Labelnya yaitu tepat waktu dan terlambat. Implementasi menggunakan aplikasi RapidMiner 5. Metode yang digunakan adalah data mining dengan algoritma C4.5. Metode pengumpulan data yang dipakai dalam penelitian yaitu observasi dan studi literatur. Berdasarkan uji coba diperoleh kesimpulan bahwa bidang ilmu data mining dengan menggunakan algoritma C4.5 dapat diimplementasikan untuk melakukan prediksi kelulusan siswa Lembaga Pelatihan Kerja, setelah melakukan rangkaian uji data set dengan rapid miner, diperoleh hasil accuracy sebagai nilai ketentuan seberapa besar keakuratan menggunakan algoritma C4.5 dalam memprediksi kelulusan siswa Lembaga Pelatihan Kerja SHINJU.

Kata kunci : Algoritma C4.5, Prediksi Kelulusan, Data mining, Rapidminer

Abstrack

LPK. SHINJU has stored is data in the database in the form of hardcopy and softfile. These data, if maximized, can provide useful information, one of which is the prediction of student graduation. This study aims to apply the classification algorithm C4.5 in predicting student graduation. The data used is the alumni data of students who have graduated in 2020. The attributes used are Gender, Place of Residence, Origin of Graduation, Work Status, Economy and Final Value. The Label attribute is on time and late. Implementation using the RapidMiner 5 application. The method used is data mining with the C4.5 algorithm. Data collection methods used in this research are observation and literature study. Based on the trial, it was concluded that the field of data mining science using the C4.5 algorithm can be implemented to predict the graduation of Job Training Institute students. After conducting a series of data set tests with rapid miners, accuracy results are obtained as the value of the provision of how much accuracy is using the C4 algorithm. 5 in predicting the graduation of SHINJU Job Training Institute students.

Keywords: C4.5 Algorithm, Graduation Prediction, Data mining, Rapidminer

PENDAHULUAN

Seiring dengan perkembangan teknologi khususnya dibidang informasi menimbulkan implikasi yang sangat luas di segala aspek hidup manusia. Perkembangan teknologi di bidang komputer telah membawa pengaruh dan dampak yang besar di kehidupan manusia. Bahkan saat ini komputer dapat menggantikan posisi manusia sebagai alat dalam melaksanakan sebuah pekerjaan. Pelatihan kerja adalah keseluruhan kegiatan untuk memberi, memperoleh, meningkatkan, serta mengembangkan kompetensi kerja, produktivitas, disiplin, sikap, dan etos kerja pada tingkat keterampilan dan keahlian tertentu sesuai dengan jenjang dan kualifikasi jabatan atau pekerjaan. ^[1] Lembaga Pelatihan Kerja (LPK) SHINJU adalah salah satu LPK yang berlokasi di Kota Semarang, Jawa Tengah dan bergerak di bidang pelatihan bahasa Jepang dan pengiriman tenaga magang ke Jepang. Siswa yang menempuh pelatihan di LPK. SHINJU minimal lulusan SMA/SMK/Sederajat. Mereka dilatih dan dipersiapkan untuk bekerja magang di Jepang, meliputi bahasa Jepang, Budaya dan Etika. Pada umumnya untuk mengikuti program magang Jepang di LPK. SHINJU diwajibkan menyelesaikan pelatihan dasar minimal 3 bulan. Namun ada beberapa hal yang tidak seimbang antara masuk dan keluarnya siswa yang telah menyelesaikan masa pelatihannya. Terdapat beberapa faktor yang mempengaruhi terlambatnya siswa dalam menyelesaikan pelatihannya seperti latar belakang pendidikan sebelumnya, faktor ekonomi keluarga, faktor perizinan masuk negara Jepang, dan nilai penguasaan bahasa yang masih rendah. Hal ini tentunya akan merugikan lembaga maupun siswa itu sendiri.

Berdasarkan faktor-faktor tersebut maka diperlukan penelitian untuk menggali data siswa sehingga diperoleh informasi yang dapat memprediksi kelulusan siswa. Data yang akan dimanfaatkan disini adalah data siswa yang sudah lulus (yang akan digunakan sebagai *data training* dan *data testing*) maupun yang belum lulus (sedang menempuh studi) yang akan digunakan untuk memprediksi masa studi masing-masing siswa. Penelitian ini dinilai perlu karena jika studi siswa dapat diketahui lebih dini, maka pihak lembaga dapat melakukan tindakan-tindakan yang dirasa perlu supaya siswa dapat lulus tepat waktu sekaligus dapat meningkatkan kualitas lembaga itu sendiri.

Permasalahan prediksi kelulusan dapat diselesaikan menggunakan teknik klasifikasi data mining. Data mining adalah proses yang memanfaatkan teknik-teknik statistik, matematika, dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan knowledge selanjutnya (atau pola-pola) yang berasal dari sekumpulan data yang sangat besar. ^[2] Tahapan di dalamnya berguna untuk mencari pola-pola tertentu dari data yang ada pada database. Biasanya, metode ini banyak ditemukan pada bidang machine learning dan statistika. Klasifikasi data mining adalah sebuah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas (class). Klasifikasi data mining menjadi salah satu metode yang paling umum untuk digunakan. Metode ini dilakukan bertujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui.

Tujuan dari penelitian ini yaitu mengetahui kinerja dari algoritma C4.5 untuk prediksi kelulusan pada Lembaga Pelatihan Kerja SHINJU dengan metode

Algoritma C4.5. pembuatan data mining dalam Algoritma yang akan membantu Lembaga magang ke jepang untuk mengetahui tepat atau terlambatnya siswa.

METODE PENELITIAN

Metode yang akan digunakan dalam penulisan jurnal ini yaitu study literatur dan observasi karena pada dasarnya metode ini merupakan proses keilmuan untuk memperoleh data secara sistematis dan juga berdasarkan bukti, berikut metode untuk jurnal penelitian dalam jurnal ini:

A. *Study Literatur*

Metode ini adalah sebuah pengumpulan data melalui jurnal ilmiah, penelitian sejenis, buku, artikel dan sumber bacaan lainnya yang berkaitan dengan penelitian. Sehingga data atau informasi yang telah didapatkan dari Lembaga tersebut dapat dijadikan sebagai rujukan untuk memperkuat landasan dan argumentasi-argumentasi yang telah ada.

Studi literatur adalah sebuah rangkaian kegiatan yang berkaitan dengan metode yang melakukan pengumpulan data membaca, pustaka, mencatat, dan juga mengolah bahan-bahan penelitian. Tujuan dari metode ini adalah untuk menunjukkan berbagai teori-teori yang sering ada dengan permasalahan yang sering terjadi dan sedang dihadapi dan diteliti sebagai bahan dalam pembahasan untuk hasil penelitian.

Secara umum studi literatur adalah sebuah cara yang digunakan untuk menyelesaikan masalah atau

persoalan dengan cara menelusuri sumber-sumber tulisan yang telah dibuat sebelumnya. Studi literatur ini sangat sering ditemukan dengan mencari di studi pustaka. Dalam penelitian yang akan dijalankan, tentunya seorang peneliti harus mempunyai wawasan yang banyak dan luas dalam hal objek yang akan diteliti. Jika peneliti tidak mempunyai wawasan yang luas terkait dengan objek yang dibahas, maka dapat dipastikan peneliti tersebut akan mempunyai persentase yang memungkinkan peneliti tersebut akan mengalami kegagalan saat akan melakukan penelitian.

B. *Observasi*

Metode ini digunakan untuk mendapatkan data yang diperlukan dengan cara melakukan pengamatan secara langsung. Data penelitian yang akan digunakan dalam penelitian ini diambil dari data alumni siswa pelatihan tahun 2020 yang didapatkan dari bagian administrasi LPK. SHINJU. Data ini akan digunakan sebagai *dataset* dalam proses prediksi.

Dataset adalah data yang digunakan untuk membangun model klasifikator yang sudah diketahui label kelasnya. Pada penelitian ini dataset yang digunakan yaitu data siswa pelatihan yang sudah lulus (alumni) yaitu data siswa pelatihan bahasa Jepang angkatan 2020.

C. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain : ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Adapun kelebihan dari Algoritma C4.5 dibandingkan dengan algoritma ID3 adalah mampu mengatasi nilai yang hilang (missing value), mampu mengatasi data yang bertipe kontinu, dan melakukan pemangkasan pohon (prunning tree). Tahapan-tahapan pembuatan pohon keputusan dengan menggunakan algoritma C4.5 antara lain sebagai berikut:

- a) Menyiapkan data training
Data berasal dari data yang sudah dikumpulkan sebelumnya dan dikelompokkan kedalam kelas tertentu.
- b) Menentukan akar dari pohon
Akar di pilih dari atribut dengan cara menghitung nilai gain dari semua atribut, dan nilai gain yang paling besar

akan menjadi akar pertama. Sebelum menghitung nilai gain, sebaiknya hitung nilai entropy terlebih dahulu. Berikut merupakan rumus yang digunakan dalam menghitung nilai entropy:

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

Keterangan :
S = himpunan kasus
n = jumlah partisi S
p_i = proporsi S_i terhadap S

- c) Hitung nilai gain dengan menggunakan rumus sebagai berikut :

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

Keterangan :
S = Himpunan Kasus
A = Atribut
n = Jumlah Partisi Atribut A
| S_i | = Jumlah Kasus pada partisi ke-i
| S | = Jumlah Kasus dalam S

- d) Hitung Splitinfo dan GainRatio pada atribut menggunakan rumus sebagai berikut:
 - Splitinfo

$$Splitinfo(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} * \log_2 \frac{|D_j|}{|D|}$$

- GainRatio

$$\text{Gain Ratio} = \frac{\text{Gain (S.A)}}{\text{Splitinfo}} \cdot A(D)$$

- e) Kemudian ulangi langkah kedua sampai semua record terpartisi secara sempurna.
- f) Proses partisi pohon keputusan akan berhenti jika :
 - a. Semua record atau tupel pada node N mendapatkan kelas yang sama.
 - b. Tidak ada atribut di dalam record atau tupel yang dipartisi lagi.
 - c. Tidak ada record atau tupel didalam cabang yang kosong.

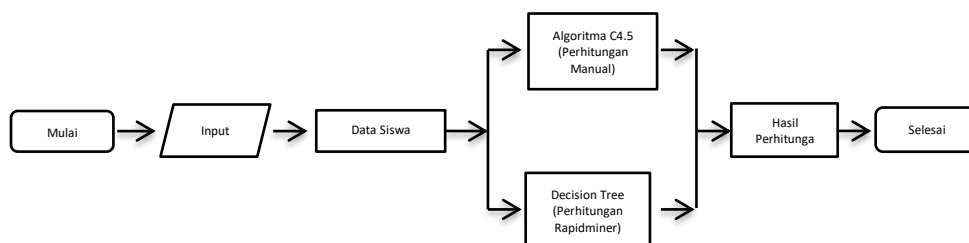
D. Perancangan

Perancangan Algoritma C4.5 menggunakan 2 metode perhitungan, yang pertama perhitungan secara manual dan yang kedua yaitu perhitungan di sebuah aplikasi bernama RapidMiner. RapidMiner juga sangat efektif dalam perhitungan di

berbagai metode salah satunya Decision Tree (C4.5). Proses tahapan metode dalam penelitian ini sebagai berikut:

- 1) Mulai
- 2) Input data atau analisis data yaitu data di pilih, kemudian ditentukan mana labelnya dan atributnya.
- 3) Data semua siswa yang berdasarkan keterkaitan penelitian.
- 4) Proses perhitungan, yang pertama menggunakan perhitungan manual. Dalam hitungan manual ini digunakan rumus yang sudah dijelaskan sebelumnya. Sedangkan untuk perhitungan otomatisnya digunakan aplikasi RapidMiner dan nanti aplikasi tersebut akan secara otomatis menghitung data yang dimasukkan.
- 5) Setelah semua perhitungan selesai maka akan didapat hasil dari kedua perhitungan tersebut.
- 6) Selesai

Proses algoritma penelitian yang diterapkan dapat dijabarkan dalam bentuk flowchart dibawah ini:



Gambar 1 Flowchart Proses Perhitungan

Pembuatan perancangan ini untuk memudahkan dalam melakukan pembuatan jurnal agar susunan urutannya menyesuaikan dengan tepat.

E. Pengolahan Data

Data yang akan digunakan adalah data alumni siswa di Lembaga Pelatihan Kerja Shinju. Data perhitungan ini menggunakan tabel Atribut sesuai dari data alumni siswa LPK SHINJU, yaitu:

- Tempat Tinggal (Jawa Tengah, Jawa Barat, Jawa Timur, Luar Jawa)
- Jenis Kelamin (Laki-laki, Perempuan)
- Asal Kelulusan (SMA, SMK)
- Status Bekerja (Pernah, Belum)
- Ekonomi (Sedang, Rendah, Tinggi)
- Nilai Akhir (80.35, 80)
- Status Kelulusan (Tepat, Terlambat)

ATRIBUT	KATEGORI	KET
TEMPAT TINGGAL	JAWA TENGAH	52
	JAWA TIMUR	4
	JAWA BARAT	35
	LUAR JAWA	6
JENIS KELAMIN	L	92
	P	4
ASAL KELULUSAN	SMA	1
	SMK	96
STATUS BEKERJA	PERNAH	53
	BELUM	44
EKONOMI	RENDAH	52
	SEDANG	45
	TINGGI	0
NILAI AKHIR	≤ 80.35	61
	> 80.35	36
	≤ 80	61
	> 80	36
STATUS KELULUSAN	TEPAT	86
	TERLAMBAT	11

Gambar 2 Informasi Data

HASIL DAN PEMBAHASAN

A. Pengolahan Data

Data dikumpulkan kemudian dikelola, pada tahap ini dilakukan perhitungan atribut-atribut seperti tempat tinggal dan data atribut yang lain. Tahap sebelumnya sudah dilakukan perancangan data yang didapat sehingga dapat lanjut ke tahap berikutnya. Tahap selanjutnya perhitungan entropy beserta gain untuk melakukannya dilihat dari jumlah data pada siswa lulus tepat waktu dan siswa lulus terlambat. Setelah mendapatkan hasil dari Gain, Splitinfo, dan Gainratio langsung ke pohon keputusan pada setiap tabel.

Tabel di atas juga sudah dijelaskan analisis perancangan data nya, yang pertama

perhitungan pada tabel sebagai berikut :

Tabel 1 Tabel Perhitungan Node 1

NODE 1	JUMLAH (S)	TEPAT (Si)	TERLAMBAT (Si)	ENTROPHY	GAIN	SPLITINFO	GAINRATIO
TOTAL	97	89	8	0.182962218			
TEMPAT TINGGAL					0.011145193	1.450865977	0.007681757
JAWA TENGAH	52	49	3	0.156647458			
JAWA BARAT	35	32	3	0.185597286			
JAWA TIMUR	4	3	1	0.188721876			
LUAR JAWA	6	5	1	0.211631745			
JENIS KELAMIN					0.007046754	0.262102401	0.026885498
LAKI	92	85	7	0.177270789			
PEREMPUAN	4	3	1	0.188721876			
ASAL KELULUSAN					0.001178894	0.082836553	0.014231572
SMA	1	1	0	0			
SMK	96	88	8	0.1836769			
STATUS BEKERJA					0.087568252	0.993781138	0.08811623
PERNAH	53	53	0	0			
BELUM PERNAH	44	36	8	0.210300335			
EKONOMI					0.0000252	-1.314737201	-0.000019167
RENDAH	52	47	5	0.193030142			
SEDANG	45	42	3	0.167559411			
NILAI AKHIR					-0.07882343	1.903077692	-0.041418924
<=80,35	61	53	8	0.208141048			
>80,35	36	36	0	0			
<=80	61	53	8	0.208141048			
>80	36	36	0	0			

Rumus yang digunakan pada pencarian nilai Entropy pada atribut yang dipakai adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i = \left(- \binom{3}{4} * \log_2 \left(\frac{3}{4} \right) \right) + \left(- \binom{1}{4} \log_2 \left(\frac{1}{4} \right) \right)$$

- Entropy (Jawa Tengah)

$$= \left(- \binom{49}{52} * \log_2 \left(\frac{49}{52} \right) \right) + \left(- \binom{3}{52} \log_2 \left(\frac{3}{52} \right) \right) = 0.156647458$$
- Entropy (Jawa Barat)

$$= \left(- \binom{32}{35} * \log_2 \left(\frac{32}{35} \right) \right) + \left(- \binom{3}{35} \log_2 \left(\frac{3}{35} \right) \right) = 0.85597286$$
- Entropy (Jawa Timur)

- Entropy (Luar Jawa)

$$= \left(- \binom{5}{6} * \log_2 \left(\frac{5}{6} \right) \right) + \left(- \binom{1}{6} \log_2 \left(\frac{1}{6} \right) \right) = 0.211631745$$
- Entropy (Laki - laki)

$= \left(- \binom{85}{92} * \text{Log}_2 \binom{85}{92} \right)$ $+ \left(- \binom{7}{92} \text{Log}_2 \binom{7}{92} \right)$ $= 0,177270789$ <ul style="list-style-type: none"> ○ Entropy (Perempuan) $= \left(- \binom{3}{4} * \text{Log}_2 \binom{3}{4} \right)$ $+ \left(- \binom{1}{4} \text{Log}_2 \binom{1}{4} \right)$ $= 0,188721876$ <ul style="list-style-type: none"> ○ Entropy (SMA) $= \left(- \binom{1}{1} * \text{Log}_2 \binom{1}{1} \right)$ $+ \left(- \binom{0}{1} \text{Log}_2 \binom{0}{1} \right)$ $= 0$ <ul style="list-style-type: none"> ○ Entropy (SMK) $= \left(- \binom{88}{96} * \text{Log}_2 \binom{88}{96} \right)$ $+ \left(- \binom{8}{96} \text{Log}_2 \binom{8}{96} \right)$ $= 0,1836769$ <ul style="list-style-type: none"> ○ Entropy (Pernah) $= \left(- \binom{53}{53} * \text{Log}_2 \binom{53}{53} \right)$ $+ \left(- \binom{0}{53} \text{Log}_2 \binom{0}{53} \right)$ $= 0$ <ul style="list-style-type: none"> ○ Entropy (Belum Pernah) $= \left(- \binom{36}{44} * \text{Log}_2 \binom{36}{44} \right)$ $+ \left(- \binom{8}{44} \text{Log}_2 \binom{8}{44} \right)$ $= 0,210300335$ <ul style="list-style-type: none"> ○ Entropy (Sedang) 	$= \left(- \binom{42}{45} * \text{Log}_2 \binom{42}{45} \right)$ $+ \left(- \binom{3}{45} \text{Log}_2 \binom{3}{45} \right)$ $= 0,167559411$ <ul style="list-style-type: none"> ○ Entropy (Rendah) $= \left(- \binom{47}{52} * \text{Log}_2 \binom{47}{52} \right)$ $+ \left(- \binom{5}{52} \text{Log}_2 \binom{5}{52} \right)$ $= 0,193030142$ <ul style="list-style-type: none"> ○ Entropy (<= 80.35) $= \left(- \binom{53}{61} * \text{Log}_2 \binom{53}{61} \right)$ $+ \left(- \binom{8}{61} \text{Log}_2 \binom{8}{61} \right)$ $= 0,208141048$ <ul style="list-style-type: none"> ○ Entropy (> 80.35) $= \left(- \binom{36}{36} * \text{Log}_2 \binom{36}{36} \right)$ $+ \left(- \binom{0}{36} \text{Log}_2 \binom{0}{36} \right)$ $= 0$ <ul style="list-style-type: none"> ○ Entropy (<=80) $= \left(- \binom{53}{61} * \text{Log}_2 \binom{53}{61} \right)$ $+ \left(- \binom{8}{61} \text{Log}_2 \binom{8}{61} \right)$ $= 0,208141048$ <ul style="list-style-type: none"> ○ Entropy (>80) $= \left(- \binom{36}{36} * \text{Log}_2 \binom{36}{36} \right)$ $+ \left(- \binom{0}{36} \text{Log}_2 \binom{0}{36} \right)$ $= 0$
--	--

Rumus yang digunakan pada pencarian nilai Gain pada atribut

yang dipakai adalah sebagai berikut:

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

- ❖ Gain (Tempat Tinggal)

$$= 0.182962218 - \left(\binom{52}{97} * 0.156647458 + \binom{35}{97} * 0.185597286 + \binom{4}{97} * 0.188721876 + \binom{6}{97} * 0.211631745 \right)$$

$$= 0.011145199$$
- ❖ Gain (Jenis Kelamin)

$$= 0.182962218 - \left(\binom{92}{97} * 0.007046754 + \binom{4}{97} * 0.188721876 \right)$$

$$= 0.004671627$$
- ❖ Gain (Asal Kelulusan)

$$= 0.182962218 - \left(\binom{1}{97} * 0 + \binom{96}{97} * 0.1836769 \right)$$

$$= 0.001178894$$

- ❖ Gain (Status Bekerja)

$$= 0.182962218 - \left(\binom{53}{97} * 0 + \binom{44}{97} * 0.210300335 \right)$$

$$= 0.087568252$$

- ❖ Gain (Ekonomi)

$$= 0.182962218 - \left(\binom{52}{97} * 0.193030142 + \binom{45}{97} * 0.167559411 \right)$$

$$= 0.0000252$$
- ❖ Gain (Nilai Akhir)

$$\begin{aligned}
 &= 0.182962218 \\
 &\quad - \left(\binom{61}{97} \right) \\
 &\quad \quad * 0.208141048 \Big) \\
 &\quad + \left(\binom{36}{97} * 0 \right) \\
 &\quad + \left(\binom{61}{97} \right) \\
 &\quad \quad * 0.208141048 \Big) \\
 &\quad + \left(\binom{36}{97} * 0 \right) \\
 &= -0.07882343
 \end{aligned}$$

Hitung Splitinfo dan GainRatio pada atribut menggunakan rumus sebagai berikut:

- Splitinfo

$$\frac{|D|}{|D|} \log_2 \frac{|D|}{|D|}$$
- GainRatio

$$\text{Gain Ratio} = \frac{\text{Gain (S.A)}}{\text{Splitinfo A(D)}}$$

✚ Splitinfo (Tempat Tinggal)

$$\begin{aligned}
 &= \left(- \binom{52}{97} * \log_2 \binom{52}{97} \right) \\
 &\quad + \left(- \binom{35}{97} \right) \\
 &\quad \quad * \log_2 \binom{35}{97} \Big) \\
 &\quad + \left(- \binom{4}{97} \right) \\
 &\quad \quad * \log_2 \binom{4}{97} \Big) \\
 &\quad + \left(- \binom{6}{97} \right) \\
 &\quad \quad * \log_2 \binom{6}{97} \Big) \\
 &= 1.450865977
 \end{aligned}$$

✚ Splitinfo (Jenis Kelamin)

$$\begin{aligned}
 &= \left(- \binom{92}{97} * \log_2 \binom{92}{97} \right) \\
 &\quad + \left(- \binom{4}{97} \right) \\
 &\quad \quad * \log_2 \binom{4}{97} \Big) \\
 &= 0.262102401
 \end{aligned}$$

✚ Splitinfo (Asal Kelulusan)

$$\begin{aligned}
 &= \left(- \binom{1}{97} * \log_2 \binom{1}{97} \right) \\
 &\quad + \left(- \binom{96}{97} \right) \\
 &\quad \quad * \log_2 \binom{96}{97} \Big) \\
 &= 0.082836553
 \end{aligned}$$

✚ Splitinfo (Status Bekerja)

$$= \left(-\binom{53}{97} * \log_2 \left(\binom{53}{97} \right) \right. \\ \left. + \left(-\binom{44}{97} \right) \right. \\ \left. * \log_2 \left(\binom{44}{97} \right) \right)$$

$$= 0.993781138$$

✚ Splitinfo (Ekonomi)

$$= \left(-\binom{52}{97} * \log_2 \left(\binom{52}{97} \right) \right) \\ + \left(-\binom{45}{97} \right) \\ * \log_2 \left(\binom{45}{97} \right)$$

$$= -1.314737201$$

✚ Splitinfo (Nilai Akhir)

$$= \left(-\binom{61}{97} * \log_2 \left(\binom{61}{97} \right) \right) \\ + \left(-\binom{36}{97} \right) \\ * \log_2 \left(\binom{36}{97} \right) \\ + \left(-\binom{61}{97} \right) \\ * \log_2 \left(\binom{61}{97} \right) \\ + \left(-\binom{36}{97} \right) \\ * \log_2 \left(\binom{36}{97} \right)$$

$$= 1.903077692$$

✚ Gainratio (Tempat Tinggal)

$$= \frac{0.011145199}{1.450865977} = 0.007681757$$

✚ Gainratio (Jenis Kelamin)

$$= \frac{0.007046754}{0.262102401} = 0.026885498$$

✚ Gainratio (Asal Kelulusan)

$$= \frac{0.001178894}{0.082836553} = 0.014231572$$

✚ Gainratio (Status Bekerja)

$$= \frac{0.087568252}{0.993781138} = 0.0881162 \quad (\checkmark)$$

✚ Gainratio (Ekonomi)

$$= \frac{0.0000252}{-1.314737201} \\ = -0.000019167$$

✚ Gainratio (Nilai Akhir)

$$= \frac{-0.07882343}{1.903077692} \\ = -0.041418924$$

Dari hasil perhitungan nilai entropi, gain pada tabel perhitungan kelulusan maka dapat disimpulkan nilai hasil gain tertinggi yaitu gain status bekerja yang mendapatkan hasil 0.0881162 untuk menentukan nilai akarnya.

B. Lakukan node sampai selesai dan membentuk pohon keputusan

Tabel 2 Tabel Perhitungan Node 13

NODE 13	JUMLAH (S)	TEPAT (Si)	TERLAMBAT (Si)	ENTROPHY	GAIN	SPLITINFO	GAINRATIO
TOTAL	7	4	3	0.062536797			
ASAL KELULUSAN					0	0	0
SMA	0	0	0	0			
SMK	7	4	3	0.062536797			

- o Perhitungan Entropy Asal Kelulusan (SMK)

$$= \left(- \binom{4}{7} * \text{Log}_2 \left(\frac{4}{7} \right) \right) + \left(- \binom{3}{7} \text{Log}_2 \left(\frac{3}{7} \right) \right) = 0.062536797$$

- ❖ Gain (Asal Kelulusan)

$$= 0.062536797 - \left(\binom{0}{7} * 0 \right) + \left(\binom{7}{7} * 0.062536797 \right) = 0$$

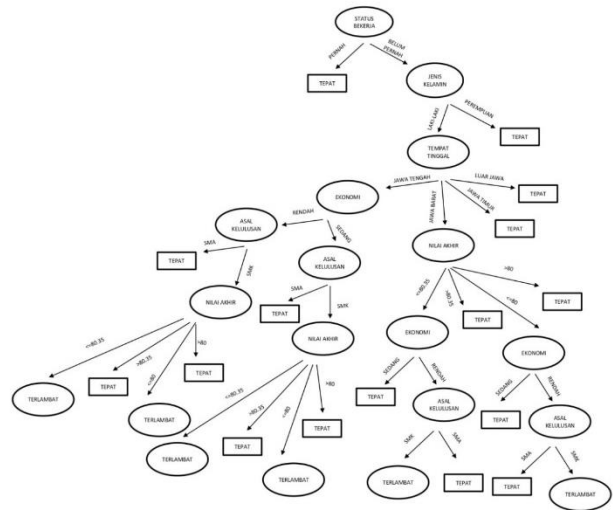
- Splitinfo (Asal Kelulusan)

$$= \left(- \binom{0}{7} * \log_2 \left(\frac{0}{7} \right) \right) + \left(- \binom{7}{7} * \log_2 \left(\frac{7}{7} \right) \right) = 0$$

- Gainratio (Asal Kelulusan)

$$= \frac{0}{0} = 0$$

Setelah semua akar diperoleh maka hasil dari pohon keputusan adalah sebagai berikut :



C. RapidMiner

RapidMiner [20] adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi

AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

Beberapa Fitur dari RapidMiner, antara lain:

- Banyaknya algoritma data mining, seperti decision tree dan self-organization map.
- Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi.
- Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.
- Mengintegrasikan proyek data mining Weka dan statistika R

KESIMPULAN

Berdasarkan uji coba dan evaluasi yang sudah dilakukan maka dapat diambil kesimpulan sebagai berikut: Penerapan Algoritma C4.5 pada prediksi kelulusan siswa Pelatihan Kerja dapat menghasilkan

accuracy yang mencapai 100% dengan menggunakan aplikasi RapidMiner. Berdasarkan penelitian ini, dapat dikatakan bahwa tepat waktu atau terlambatnya kelulusan siswa dapat di lihat dari perolehan Nilai Akademis siswa tersebut selama menempuh pelatihan di LPK. SHINJU.

DAFTAR PUSTAKA

- ADMIN IN KOMPUTER, 2020. *RapidMiner : Mengenal Aplikasi Data Mining Terkemuka di Dunia*. [Online] Available at: <https://www.doavers.com/blog/rapidminer-mengenal-aplikasi-data-mining-terkemuka-di-dunia> [Accessed 03 2022].
- Anggreany, M. S., 2020. *Confusion Matrix*. [Online] Available at: <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> [Accessed 30 12 2021].
- Bajabir, A. Z. A. M., 2018. *Penerapan Metode Naive Bayes Untuk Prediksi Menentukan Karyawan Tetap Pada PT. YSP Industries Indonesia..* [Online] Available at: <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=22968&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain> [Accessed 20 12 2021].
- Cahya, D., 2018. *Naïve Bayes Classifier (NBC)*. [Online] Available at: https://cahyadsn.phpindonesia.id/extra/naive_bayes.php [Accessed 22 12 2021].

- DPPAP2TKILN, 2015. *Proses Pelaksanaan Pelatihan di LPK (bagian 1)*. [Online] Available at: <https://dppap2tkiln.wordpress.com/2015/04/07/proses-pelaksanaan-pelatihan-di-lpk-bagian-1/> [Accessed 25 01 2022].
- Hardiansyah, Z., 2022. *Mengenal Microsoft Excel Beserta Manfaat dan Rumusnya*. [Online] Available at: <https://tekno.kompas.com/read/2022/02/10/14150067/mengenal-microsoft-excel-beserta-manfaat-dan-rumusny?page=all> [Accessed 04 2022].
- Heryana, D., 2019. *Data Mining untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika UIN Raden Intan Lampung Menggunakan Naive Bayes*, s.l.: s.n.
- Hidayatullah, A. F. P. A. D. S. D. P. P. I., 2011. *Analisis Kualitas Data dan Klasifikasi Data Pasien Kanker*. [Online] Available at: https://www.researchgate.net/publication/270823880_Analisis_Kualitas_Data_dan_Klasifikasi_Data_Pasien_Kanker [Accessed 22 01 2022].
- Hulu, S., 2020. *Analisis Kinerja Metode Cross Validation Dan K-Nearest Neighbor Dalam Klasifikasi Data*, Sumatera utara: Teknik Informatika, Universitas Sumatera Utara.
- Kholifah, I. N., 2020. *Memprediksi Tingkat Kelulusan Peserta Kursus Pertama dengan Algoritma Data Mining C4.5 dan RapidMiner.. JUISI*, 06(No. 01, Februari 2020).
- LPK SHINJU, 2021. *Data Siswa*, Semarang: Prima Mahargya.
- Novichasari, S. I., 2013. *Penentuan Kelayakan Kredit Bank Menggunakan Naive Bayes Classifier Berbasis Particle Swarm Optimization*, s.l.: s.n.
- Pintek.id, 2020. *Lebih Bersaing! Upgrade Fasilitas LPK bersama Pintek*. [Online] Available at: <https://pintek.id/blog/lembaga-pelatihan-kerja/> [Accessed 25 01 2022].
- Putri, R. P. S., 2018. *Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Teknik Informatika*, s.l.: s.n.
- Riadi, M., 2019. *Pengertian, Fungsi, Proses dan Tahapan Data Mining*. [Online] Available at: <https://www.kajianpustaka.com/2017/09/data-mining.html> [Accessed 28 12 2021].
- Safitri, N., 2018. *Penerapan Data Mining Untuk Prediksi Kelulusan Siswa Menggunakan Algoritma Naive Bayes Pada Sma Negeri 2 Sukatani.*, s.l.: s.n.
- Samsudiney, 2019. *Penjelasan Sederhana tentang Apa Itu SVM?*. [Online] Available at: <https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02> [Accessed 29 12 2021].
- Sano, A. V. D. S. M., 2019. *Definisi, Karakteristik, dan Manfaat Data Mining -Seri Data Mining for Business Intelligence (2)*. [Online] Available at: <https://binus.ac.id/malang/2019/01/definisi-karakteristik-dan-manfaat-data-mining-seri-data-mining-for-business-intelligence-2/> [Accessed 20 12 2021].

- Setiawan, R., 2021. *Apa itu Data Mining dan Bagaimana Metodenya?*. [Online] Available at: <https://www.dicoding.com/blog/apa-itu-data-mining/> [Accessed 30 12 2021].
- Siti Mutmainnah, G. A. H. A. A. F., 2020. *Optimasi Algoritma C4.5 Menggunakan Teknik Bagging Pada Data Kadar Karat Emas. Universitas Muhamadiyah Jember.*
- Yudhoyono, D. H. S. B., 2006. *Peraturan Pemerintah Republik Indonesia Nomor 31 Tahun 2006 Tentang Sistem Pelatihan Kerja Nasional..* Jakarta: Menteri Hukum Dan Hak Asasi Manusia Republik Indonesia..
- Yuliana, T., 2019. *Artificial Neural Network (ANN) in R.* [Online] Available at: <https://medium.com/@tyasyuliana/artificial-neural-network-ann-in-r-56a78b53ecf8> [Accessed 29 12 2021].
- Yuliana, T., 2019. *Artificial Neural Network (ANN) in R.* [Online] Available at: <https://medium.com/@tyasyuliana/artificial-neural-network-ann-in-r-56a78b53ecf8> [Accessed 29 12 2021].