



PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA *DECISION TREE C4.5* UNTUK MEMPREDIKSI MAHASISWA *DROP OUT* DI UNIVERSITAS WIRARAJA

IDDRUS^{1,*}, DEWI WULAN SARI²

¹) Program Studi S1 Informatika Universitas Wiraraja, Sumenep, Indonesia

²) Program Studi S1 Bisnis Digital Universitas Wiraraja, Sumenep, Indonesia

EMAIL: iddrus@wiraraja.ac.id, dewiwulansari@wiraraja.ac.id

Diterima : 11 Juni 2023. Disetujui : 15 Juni 2023. Dipublikasikan : 20 Juni 2023.

ABSTRACT - *Education providers are required to be able to provide good educational services and quality education to each student. Service and quality of education are provided to ensure students succeed in their academics. However, academic failures still occur in some students at various existing universities. One example of academic failure that often occurs is Drop Out (DO). DO is a condition where students are unable to complete their studies during the study period that has been given. DO is a big loss for students and education providers. By utilizing data mining, students should be able to predict from the start who are likely to experience dropouts. Based on these problems, a study was conducted to predict the probability of students dropping out using the Decision Tree algorithm C4.5 model. The test results obtained in this study are accuracy.*

Keywords : *Prediction; Student; Decision Tree C4.5*

ABSTRAK - *Penyelenggara pendidikan dituntut dapat memberikan pelayanan pendidikan dan mutu pendidikan yang baik kepada setiap mahasiswanya. Pelayanan dan mutu pendidikan diberikan guna*

memastikan mahasiswa berhasil dalam akademiknya. Namun, kegagalan akademik tetap terjadi pada beberapa mahasiswa diberbagai universitas yang ada. Salah satu contoh kegagalan akademik yang sering terjadi adalah drop out (DO). DO adalah kondisi dimana mahasiswa tidak mampu menyelesaikan studinya selama masa studi yang telah diberikan. DO menjadi kerugian besar bagi mahasiswa dan penyelenggara pendidikan. Dengan memanfaatkan data mining, seharusnya dapat diprediksi sejak awal mahasiswa yang kemungkinan akan mengalami DO. Berdasarkan masalah tersebut maka dilakukan penelitian untuk memprediksi kemungkinan mahasiswa DO menggunakan model algoritma Decision Tree C4.5 Untuk hasil uji coba yang didapatkan pada penelitian ini yaitu akurasi.

Kata kunci : *Prediksi; Mahasiswa; Pohon Keputusan C4.5*

I. PENDAHULUAN

Penyelenggara pendidikan dituntut dapat memberikan pelayanan pendidikan dan mutu pendidikan yang baik kepada setiap mahasiswanya. Pelayanan dan mutu pendidikan diberikan guna memastikan mahasiswa berhasil dalam akademiknya. Salah satu acuan kualitas perguruan

tinggi dapat dilihat dari *grade* akreditasi yang diberikan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) ataupun Lembaga Akreditasi Mandiri Pendidikan Tinggi (LAM-PT). *Grade* akreditasi yang baik dapat diperoleh perguruan tinggi dengan melengkapi kriteria-kriteria yang telah ditentukan BAN-PT atau LAM-PT. Salah satu kriteria yang harus terpenuhi adalah tingkat kelulusan tepat waktu pada perguruan tinggi.

Dalam pelaksanaannya tidak semua mahasiswa sukses dalam akademiknya yaitu mereka yang mampu lulus tepat waktu. Tidak sedikit mahasiswa yang lulus terlambat atau bahkan mengalami kegagalan akademik. Salah satu contoh kegagalan akademik yang sering terjadi adalah *drop out* (DO). DO menjadi kerugian besar bagi mahasiswa dan penyelenggara pendidikan.

Penyelenggara pendidikan memiliki data lengkap untuk setiap mahasiswa baik data pendaftaran, data akademi dan data pribadi. Data tersebut adalah modal penting untuk dapat digali informasi berharga yang mungkin sebelumnya belum diketahui. Informasi yang didapatkan dari data yang telah diolah dengan teknik tertentu tersebut dapat dijadikan rujukan untuk menyelesaikan atau mencegah berbagai masalah yang akan terjadi pada mahasiswa salah satunya untuk memprediksi sejak awal mahasiswa yang kemungkinan akan mengalami DO. Salah satu teknik pengolahan dan penggalian data yang dapat digunakan adalah *data mining*.

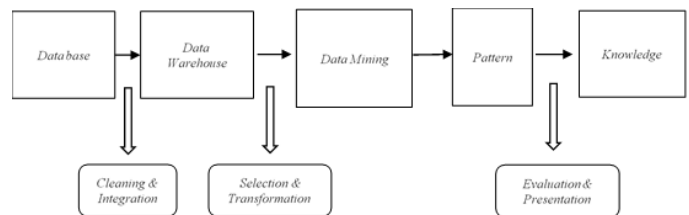
Data mining adalah analisa terhadap data yang berjumlah besar untuk menemukan hubungan yang jelas sehingga dapat ditarik kesimpulan dan berguna bagi pemilik data. Salah satu teknik yang dapat digunakan dalam *data mining* adalah prediksi. Salah satu model prediksi yang dapat digunakan algoritma *Decision Tree C4.5*. Berdasarkan pemaparan di atas, maka menjadi penting bagi peneliti untuk melakukan penelitian terhadap mahasiswa di Universitas Wiraraja untuk memprediksi kemungkinan mahasiswa DO dengan menggunakan teknik *data mining* yaitu menggunakan algoritma *Decision Tree C4.5*.

II. TEORI DASAR

2.1 Data Mining

Data mining adalah analisa terhadap data yang berjumlah besar untuk menemukan hubungan yang jelas sehingga dapat ditarik kesimpulan dan berguna bagi pemilik data. Pada proses data yang jumlahnya besar memerlukan *data mining* sehingga diharapkan memperoleh hasil berupa informasi baru terkait dari dan dapat dimanfaatkan sebagai sumber untuk pengambilan keputusan selanjutnya. Banyak istilah lain yang digunakan untuk *data mining*, seperti *analisa data/pola*, *arkeologi data*, *pengerukan data*, *ekstraksi pengetahuan* dan lainnya.

Menurut Han (2012:7) proses *data mining* merupakan urutan iteratif dari langkah-langkah yang ditunjukkan pada Gambar 1.



Gambar 1.
Proses *Data Mining*

Data mining memiliki beberapa teknik pengolahan data untuk tujuan dan kegunaan berbeda. Berikut beberapa teknik dalam *data mining*:

1. Regresi / Estimasi (*Regression/Estimation*)
Memprediksi label target numerik dari suatu titik data (*data point*). Contoh algoritma dari regresi adalah *Linear Regression*, *Logistic Regression*.
2. Prediksi (*Prediction*)
Menemukan pola tertentu dari suatu data. Pola tersebut dapat diketahui dari variabel-variabel yang ada pada data. Pola yang didapat bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya.
3. Klasifikasi (*Classification*)
Memprediksi apakah suatu data termasuk ke dalam salah satu kelas data yang telah ditetapkan. Contoh algoritma klasifikasi adalah *Decision Tree*, *Neural Network*, *Bayesian Model*, *K-Nearest Neighbour*.
4. *Association Analysis*
Mengidentifikasi hubungan dalam satu set item berdasarkan data transaksi. Contoh

algoritmanya adalah *Frequent Pattern Growth (FP-Growth)*, *Apriori Algorithm*.

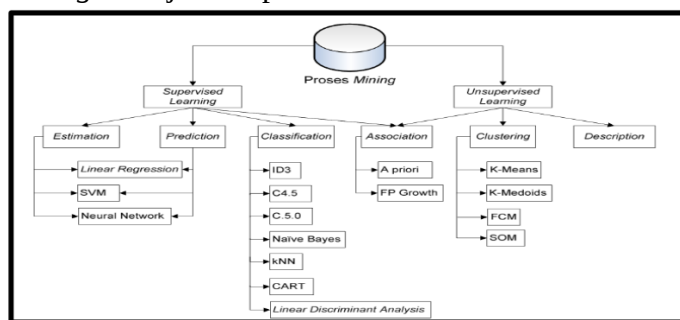
5. Klasterisasi (*Clustering*)

Mengidentifikasi kelompok data dalam dataset berdasarkan atribut dalam dataset. Contoh algoritmanya adalah *K-Means*, *Density-Based Clustering*.

6. *Description*

Memahami lebih jauh tentang data yang diamati. Sehingga diharap mampu mengetahui perilaku dari data tersebut yang nantinya bisa digunakan untuk mengetahui karakteristik dari data.

Beberapa metode menurut Ridwan dkk, yang bisa dimanfaatkan berdasarkan pengelompokan data *mining* ditunjukkan pada Gambar 2.



Gambar 2.
Metode Data Mining

2.2 Pohon Keputusan (*Decision Tree*)

Pohon keputusan (*Decision Tree*) adalah Sebuah metode kecerdasan buatan yang berbentuk pohon dimana pada setiap cabang memiliki pilihan serta alternatif dengan daun yang menyatakan keputusan yang dipilih. *Decision Tree* banyak digunakan karena dapat menggabungkan sebuah pola/pengetahuan/ informasi kedalam sebuah bentuk pohon keputusan. *Decision Tree* terdiri dari kumpulan *node* (simpul) yang dihubungkan oleh cabang. Terdapat tiga jenis *node* pada *Decision Tree* :

1. *Root node*, merupakan simpul yang tidak memiliki input tetapi memiliki output yang lebih dari satu.
2. *Internal node*, memiliki satu input dan memiliki output lebih dari dua
3. *Leaf* atau *terminal node*, mempunyai satu input dan tidak mempunyai output.

Pada setiap *Decision Tree*, setiap *leaf* memiliki sebuah nama kelas. *Root node* dan *internal node* berisi aturan yang digunakan untuk membedakan data yang memiliki karakter berbeda.

2.3 Algoritma *Decision Tree* C4.5

Terdapat beberapa algoritma yang dapat dilakukan untuk membuat suatu pohon keputusan (*Decision Tree*) salah satu yang sering digunakan adalah algoritma C4.5. Algoritma C4.5 dikembangkan oleh Ross Quinlan sebagai pengembangan dari algoritma ID3. Algoritma C4.5 adalah pembentukan pohon keputusan berdasarkan *training* data. Kelebihan dari algoritma ini yaitu dapat mengatasi *missing value*, dapat mengolah data numerik serta kategori *noisy* data. Dalam membuat pohon keputusan dengan algoritma C4.5 diperlukan langkah langkah sebagai berikut .

1. Siapkan data *training*

Data *training* adalah data histori yang telah disiapkan sebelumnya dan telah dikelompokkan kedalam bentuk kelas-kelas tertentu.

2. Menentukan akar dari pohon

Akar akan ditentukan melalui atribut yang telah dipilih dengan cara menghitung nilai *gain* dari setiap atribut, nilai *gain* yang paling tinggi nantinya akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut perlu menghitung terlebih dahulu nilai *entropy*. Untuk menghitung nilai *entropy* memiliki dapat digunakan rumus sebagai berikut

$$Entropy(N) = \sum_{i=1}^P p_i * \log_2 p_i \quad (1)$$

Dimana :

N : Himpunan kasus

P : Jumlah partisi N

Pi : Proporsi dari Ni terhadap N

3. Setelah menghitung *entropy* setiap kasus maka selanjutnya menghitung nilai *gain* menggunakan rumus sebagai berikut :

$$Gain(N,a) = Entropy(N) - \sum_{i=1}^P \frac{|N_i|}{|N|} * Entropy(N_i) \quad (2)$$

Dimana :

N : Himpunan kasus

A : Atribut

P : Jumlah partisi atribut A

|Ni| : jumlah kasus pada partisi ke i

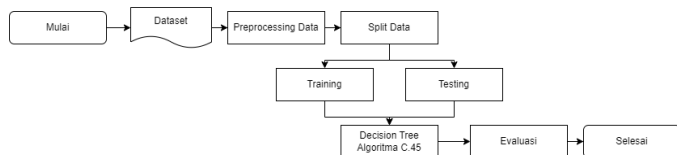
|N| : Jumlah kasus dalam N

4. Setelah selesai lalu ulangi langkah ke 2 dan ke 3 sampai semua atribut dihitung dan memiliki informasi dan yang akan dijadikan sebagai *node* (akar).

III. RANCANGAN SISTEM

3.1 Alur Penelitian

Tahapan penelitian yang dilakukan digambarkan pada gambar di bawah ini:



Gambar 3.
Alur Penelitian

Pada gambar diatas merupakan tahapan penelitian yang dilakukan dengan menggunakan *Rapid Miner* dimana pertama yang dilakukan adalah pengumpulan data. Data yang digunakan adalah data bersumber dari bagian akademik Universitas Wiraraja. Data yang digunakan meliputi data akademik, pendaftaran dan pribadi mahasiswa yang berjumlah 992. Data yang didapatkan akan dilakukan proses *preprocessing* data terlebih dahulu sebelum diolah. Setelah data selesai proses *preprocessing* dilakukan pembagi data menjadi 80% data *training* dan 20% data *testing*, selanjutnya metode yang digunakan dalam perhitungan adalah algoritma C4.5 yang nantinya akan menghasilkan prediksi, pohon keputusan dan tingkat akurasi dari data yang diolah.

3.2 Perhitungan Algoritma Decision Tree C4.5

Algoritma *Decision Tree C4.5* menghasilkan sebuah diagram alir yang terdiri dari *root*, *root node*, *internal node*, dan *leaf node*. Penentuan *root*, *root node*, *internal node*, dan *leaf node* dilakukan dengan cara perhitungan nilai *entropy* dan *gain* pada semua atribut yang digunakan. Nilai *entropy* dan *gain* dihitung menggunakan persamaan (1) dan persamaan (2) diatas. Berikut ini adalah perhitungan nilai *entropy* dan *gain*.

Menghitung *entropy* total :

$$Entropy[total] = \left(-\frac{768}{992} \times \log_2\left(\frac{768}{992}\right)\right) + \left(-\frac{224}{992} \times \log_2\left(\frac{224}{992}\right)\right) = 0.770629069$$

Menghitung *entropy* dan *gain* penghasilan_ortu

$$Entropy[penghasilan_ortu - \text{Kurang dari Rp.500.000}] = \left(-\frac{198}{260} \times \log_2\left(\frac{198}{260}\right)\right) + \left(-\frac{62}{260} \times \log_2\left(\frac{62}{260}\right)\right) = 0.792472498$$

$$Entropy[penghasilan_ortu - > \text{Rp. 500.000 s/d Rp. 1.000.000}] = \left(-\frac{311}{410} \times \log_2\left(\frac{311}{410}\right)\right) + \left(-\frac{99}{410} \times \log_2\left(\frac{99}{410}\right)\right) = 0.797465429$$

$$Entropy[penghasilan_ortu - > \text{Rp. 1.000.000 s/d 2.000.000}] = \left(-\frac{114}{135} \times \log_2\left(\frac{114}{135}\right)\right) + \left(-\frac{21}{135} \times \log_2\left(\frac{21}{135}\right)\right) = 0.623570208$$

$$Entropy[penghasilan_ortu - > \text{Rp. 2.000.000 s/d Rp. 5.000.000}] = \left(-\frac{135}{168} \times \log_2\left(\frac{135}{168}\right)\right) + \left(-\frac{33}{168} \times \log_2\left(\frac{33}{168}\right)\right) = 0.714727473$$

$$Entropy[penghasilan_ortu - > \text{Rp. 5000.000 Lebih}] = \left(-\frac{10}{19} \times \log_2\left(\frac{10}{19}\right)\right) + \left(-\frac{9}{19} \times \log_2\left(\frac{9}{19}\right)\right) = 0.998000884$$

$$Gain(Total, penghasilan_ortu) = 0.770629069 - \left(\frac{260}{992} \times 0.792472498\right) - \left(\frac{410}{992} \times 0.797465429\right) - \left(\frac{135}{992} \times 0.623570208\right) - \left(\frac{168}{992} \times 0.714727473\right) - \left(\frac{19}{992} \times 0.998000884\right)$$

$$= 0.770629069 - 0.207704485 - 0.329597607 - 0.084860865 - 0.121042556 - 0.019114936 = 0.00830862$$

Menghitung *entropy* dan *gain* penghasilan_ortu

$$Entropy[pekerjaan_ortu - \text{Wiraswasta}] = \left(-\frac{478}{635} \times \log_2\left(\frac{478}{635}\right)\right) + \left(-\frac{157}{635} \times \log_2\left(\frac{157}{635}\right)\right) = 0.806880826$$

$$Entropy[pekerjaan_ortu - \text{Guru/Dosen Negeri}] = \left(-\frac{8}{10} \times \log_2\left(\frac{8}{10}\right)\right) + \left(-\frac{2}{10} \times \log_2\left(\frac{2}{10}\right)\right) = 0.721928095$$

$$Entropy[pekerjaan_ortu - \text{Guru/Dosen Swasta}] = \left(-\frac{6}{8} \times \log_2\left(\frac{6}{8}\right)\right) + \left(-\frac{2}{8} \times \log_2\left(\frac{2}{8}\right)\right) = 0.811278124$$

$$Entropy[pekerjaan_ortu - \text{Pegawai Negeri Sipil}] = \left(-\frac{165}{195} \times \log_2\left(\frac{165}{195}\right)\right) + \left(-\frac{30}{195} \times \log_2\left(\frac{30}{195}\right)\right) = 0.619382195$$

$$Entropy[pekerjaan_ortu - \text{Pegawai Swasta}] = \left(-\frac{65}{82} \times \log_2\left(\frac{65}{82}\right)\right) + \left(-\frac{17}{82} \times \log_2\left(\frac{17}{82}\right)\right) = 0.736323027$$

$$Entropy[pekerjaan_ortu - \text{Pensiunan PNS/TNI/POLRI}] = \left(-\frac{14}{16} \times \log_2\left(\frac{14}{16}\right)\right) + \left(-\frac{2}{16} \times \log_2\left(\frac{2}{16}\right)\right) = 0.543564443$$

$$Entropy[pekerjaan_ortu - \text{Pensiunan Swasta}] = \left(-\frac{21}{26} \times \log_2\left(\frac{21}{26}\right)\right) + \left(-\frac{5}{26} \times \log_2\left(\frac{5}{26}\right)\right) = 0.706274089$$

$$Entropy[pekerjaan_ortu - \text{Tidak Bekerja}] = \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} \times \log_2\left(\frac{4}{6}\right)\right) = 0.918295834$$

$$Entropy[pekerjaan_ortu - \text{TNI/POLRI}] = \left(-\frac{9}{14} \times \log_2\left(\frac{9}{14}\right)\right) + \left(-\frac{5}{14} \times \log_2\left(\frac{5}{14}\right)\right) = 0.940285959$$

$$Gain(Total, pekerjaan_ortu) = 0.770629069 - \left(\frac{635}{992} \times 0.806880826\right) - \left(\frac{10}{992} \times 0.721928095\right) - \left(\frac{8}{992} \times 0.811278124\right) - \left(\frac{195}{992} \times 0.619382195\right) - \left(\frac{82}{992} \times 0.736323027\right) - \left(\frac{16}{992} \times 0.543564443\right) - \left(\frac{26}{992} \times 0.706274089\right) - \left(\frac{6}{992} \times 0.918295834\right) - \left(\frac{14}{992} \times 0.940285959\right)$$

$$= 0.770629069 - 0.516501335 - 0.007277501 - 0.006542566$$

$$- 0.121753556 - 0.060865412 - 0.008767168 - 0.018511216$$

$$- 0.005554209 - 0.013270165 = 0.011585942$$

Menghitung *entropy* dan *gain* jenis_kelamin

$$Entropy[jenis_kelamin - L] =$$

$$\left(-\frac{414}{550} \times \log_2\left(\frac{414}{550}\right)\right) + \left(-\frac{136}{550} \times \log_2\left(\frac{136}{550}\right)\right) = 0.806926814$$

$$Entropy[jenis_kelamin - P] =$$

$$\left(-\frac{354}{442} \times \log_2\left(\frac{354}{442}\right)\right) + \left(-\frac{88}{442} \times \log_2\left(\frac{88}{442}\right)\right) = 0.720114444$$

$$Gain(Total, jenis_kelamin) =$$

$$= 0.770629069 - \left(\frac{550}{992} \times 0.806926814\right) - \left(\frac{442}{992} \times 0.720114444\right)$$

$$= 0.770629069 - 0.447388859 - 0.320857444 = 0.011585942$$

Menghitung *entropy* dan *gain* Asal_sklh

$$Entropy[Asal_sklh - SMA] =$$

$$\left(-\frac{485}{617} \times \log_2\left(\frac{485}{617}\right)\right) + \left(-\frac{132}{617} \times \log_2\left(\frac{132}{617}\right)\right) = 0.748943732$$

$$Entropy[Asal_sklh - SMK] =$$

$$\left(-\frac{113}{147} \times \log_2\left(\frac{113}{147}\right)\right) + \left(-\frac{34}{147} \times \log_2\left(\frac{34}{147}\right)\right) = 0.780257655$$

$$Entropy[Asal_sklh - MA] =$$

$$\left(-\frac{170}{228} \times \log_2\left(\frac{170}{228}\right)\right) + \left(-\frac{58}{228} \times \log_2\left(\frac{58}{228}\right)\right) = 0.818155993$$

$$Gain(Total, Asal_sklh) =$$

$$= 0.770629069 - \left(\frac{617}{992} \times 0.748943732\right) - \left(\frac{147}{992} \times 0.780257655\right) -$$

$$\left(\frac{228}{992} \times 0.818155993\right)$$

$$= 0.770629069 - 0.465824882 - 0.115622858 - 0.188043918 = 0.002382767$$

Menghitung *entropy* dan *gain* IP

$$Entropy[IP - Sangat Rendah] =$$

$$\left(-\frac{28}{52} \times \log_2\left(\frac{28}{52}\right)\right) + \left(-\frac{24}{52} \times \log_2\left(\frac{24}{52}\right)\right) = 0.995727452$$

$$Entropy[IP - Rendah] =$$

$$\left(-\frac{34}{58} \times \log_2\left(\frac{34}{58}\right)\right) + \left(-\frac{24}{58} \times \log_2\left(\frac{24}{58}\right)\right) = 0.978449329$$

$$Entropy[IP - Cukup] =$$

$$\left(-\frac{185}{245} \times \log_2\left(\frac{185}{245}\right)\right) + \left(-\frac{59}{245} \times \log_2\left(\frac{59}{245}\right)\right) = 0.796395168$$

$$Entropy[IP - Tinggi] =$$

$$\left(-\frac{520}{637} \times \log_2\left(\frac{520}{637}\right)\right) + \left(-\frac{117}{637} \times \log_2\left(\frac{117}{637}\right)\right) = 0.688047624$$

$$Gain(Total, IP) =$$

$$= 0.770629069 - \left(\frac{52}{992} \times 0.995727452\right) - \left(\frac{58}{992} \times 0.978449329\right) -$$

$$\left(\frac{245}{992} \times 0.796395168\right) - \left(\frac{245}{992} \times 0.796395168\right)$$

$$= 0.770629069 - 0.052195391 - 0.057207723 - 0.196690339 - 0.441820903$$

$$= 0.001137412$$

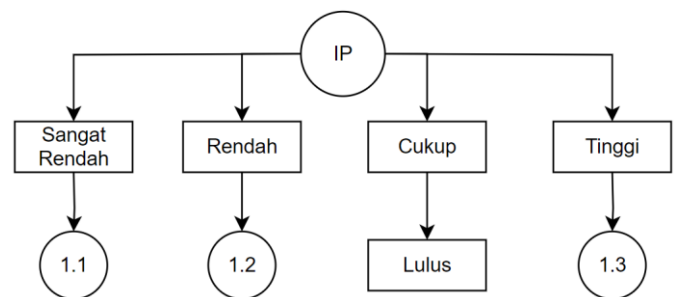
Hasil perhitungan *node* 1 ditunjukkan pada tabel 1. Selanjutnya untuk *node* 2 dan berikutnya dilakukan perhitungan menggunakan cara yang sama.

Tabel 1.

Hasil perhitungan *entropy* dan *gain* untuk *node* 1

Node Akar I	Jumlah	L	DO	Entropy	Gain
Total	992	768	224	0.7706	
PENGHASILAN					0.0083086
Kurang dari Rp. 500.000	260	198	62	0.7925	
> Rp. 500.000 s/d Rp. 1.000.000	410	311	99	0.7975	
> Rp. 1.000.000 s/d 2.000.000	135	114	21	0.6236	
> Rp. 2.000.000 s/d Rp. 5.000.000	168	135	33	0.7147	
> Rp. 5000.000 Lebih	19	10	9	0.998	
PEKERJAAN ORTU					0.0115859
Wiraswasta	635	478	157	0.8069	
Guru/Dosen Negeri	10	8	2	0.7219	
Guru/Dosen Swasta	8	6	2	0.8113	
Pegawai Negeri Sipil	195	165	30	0.6194	
Pegawai Swasta	82	65	17	0.7363	
Pensiunan PNS/TNI/POLRI	16	14	2	0.5436	
Pensiunan Swasta	26	21	5	0.7063	
Tidak Bekerja	6	2	4	0.9183	
TNI / POLRI	14	9	5	0.9403	
Jenis kelamin					0.0023828
Laki	550	414	136	0.8069	
Perempuan	442	354	88	0.7201	
ASAL SMA					0.0011374
SMA	617	485	132	0.7489	
SMK	147	113	34	0.7803	
MA	228	170	58	0.8182	
IP					0.0227147
0,00 - 0,99 = Sangat Rendah	52	28	24	0.9957	
1,00 - 1,99 = Rendah	58	34	24	0.9784	
2,00 - 2,99 = Cukup	245	186	59	0.7964	
3,00 - 4,00 = Tinggi	637	520	117	0.688	

Dari hasil perhitungan pada table 1 dapat digambarkan pohon keputusan sementara seperti gambar berikut:



Gambar 4.

Pohon Keputusan Sementara

Selanjutnya pohon keputusan sementara dilengkapi pada bagian *node* selanjutnya dengan cara melakukan perhitungan *entropy* dan *gain* dengan cara yang sama. Sehingga jika semua proses perhitungan selesai dan semua nilai *entropy* dan *gain* didapat.

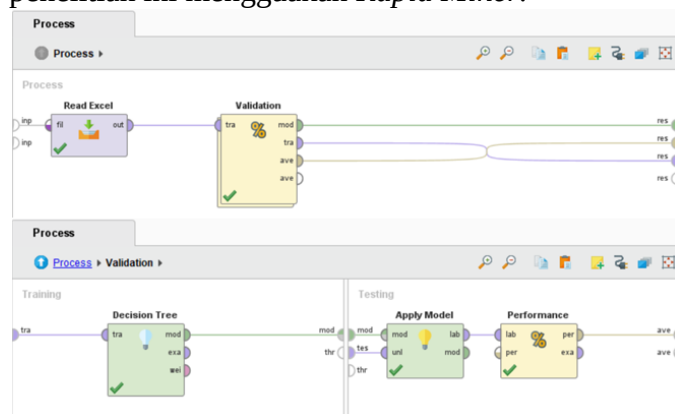
IV. IMPLEMENTASI

Pada penelitian ini model dibangun menggunakan bantuan *software Rapid Miner*.

Model yang dibangun pada *Rapid Miner* dalam penelitian ini menggunakan beberapa operator sebagai berikut.

1. Operator **Read Excel** digunakan untuk membaca file *excel* yang berisi *dataset* yang telah dilakukan proses *preprocessing* sebelumnya.
2. Operator **Validation** digunakan untuk proses pembagian data latih dan data uji dengan perbandingan 80:20 dan menggunakan teknik pembagian *linier sampling*.
3. Operator **Algoritma Decision Tree** digunakan untuk pengolahan data.
4. Operator **Apply Model** digunakan untuk penerapan model algoritma yang digunakan.
5. Operator **Performance** digunakan untuk mengevaluasi dan melihat performa pengolahan data dari algoritma *Decision Tree* yang digunakan.

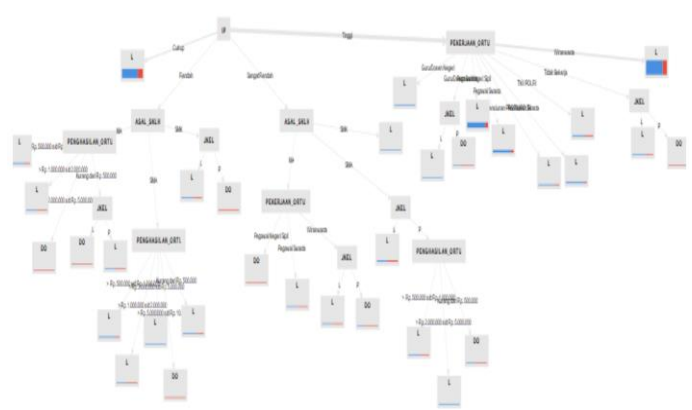
Berikut model yang dibangun dalam penelitian ini menggunakan *Rapid Miner*.



Gambar 5.

Model penelitian pada *Rapid Miner*

Hasil pengujian model yang dibangun menggunakan *algoritma Decision Tree C4.5* dengan bantuan software *Rapid Miner* berupa *Decision Tree* dan aturan atau *Rule Decision Tree*. *Decision Tree* dan aturan atau *rule* yang dihasilkan dapat dilihat pada gambar berikut ini.



Gambar 6.

Decision Tree pada *Rapid Miner*

Tree

```

IP = Cukup: L {L=186, DO=59}
IP = Rendah
  ASAL_SKLH = MA
    PENGHASILAN_ORTU = >= Rp. 500.000 s/d Rp. 1.000.000: L {L=6, DO=2}
    PENGHASILAN_ORTU = > Rp. 1.000.000 s/d 2.000.000: L {L=2, DO=2}
    PENGHASILAN_ORTU = > Rp. 2.000.000 s/d Rp. 5.000.000: DO {L=0, DO=2}
  ASAL_SKLH = SMA
    PENGHASILAN_ORTU = < Rp. 500.000: JKEL = L: DO {L=0, DO=3}
    PENGHASILAN_ORTU = < Rp. 500.000: JKEL = P: L {L=1, DO=1}
    PENGHASILAN_ORTU = >= Rp. 500.000 s/d Rp. 1.000.000: L {L=9, DO=2}
    PENGHASILAN_ORTU = > Rp. 1.000.000 s/d 2.000.000: L {L=2, DO=2}
    PENGHASILAN_ORTU = > Rp. 2.000.000 s/d Rp. 5.000.000: L {L=5, DO=0}
    PENGHASILAN_ORTU = > Rp. 5.000.000 s/d Rp. 10.000.000: DO {L=0, DO=2}
  ASAL_SKLH = SMK
    PENGHASILAN_ORTU = < Rp. 500.000: L {L=5, DO=2}
    PENGHASILAN_ORTU = >= Rp. 500.000 s/d Rp. 1.000.000: L {L=3, DO=2}
    PENGHASILAN_ORTU = > Rp. 1.000.000 s/d 2.000.000: L {L=4, DO=0}
    PENGHASILAN_ORTU = > Rp. 2.000.000 s/d Rp. 5.000.000: L {L=4, DO=0}
    PENGHASILAN_ORTU = > Rp. 5.000.000 s/d Rp. 10.000.000: DO {L=2, DO=3}
  ASAL_SKLH = SMA
    JKEL = L: L {L=11, DO=11}
    JKEL = P
      PENGHASILAN_ORTU = < Rp. 500.000 s/d Rp. 1.000.000: L {L=3, DO=2}
      PENGHASILAN_ORTU = >= Rp. 500.000 s/d Rp. 1.000.000: L {L=4, DO=0}
      PENGHASILAN_ORTU = > Rp. 1.000.000 s/d 2.000.000: L {L=4, DO=0}
      PENGHASILAN_ORTU = > Rp. 2.000.000 s/d Rp. 5.000.000: L {L=4, DO=0}
      PENGHASILAN_ORTU = > Rp. 5.000.000 s/d Rp. 10.000.000: DO {L=2, DO=3}
      ASAL_SKLH = SMK: L {L=3, DO=0}
IP = Tinggi
  PEKERJAAN_ORTU = Guru/Dosen Negeri: L {L=4, DO=0}
  PEKERJAAN_ORTU = Guru/Dosen Swasta
    JKEL = L: L {L=3, DO=0}
    JKEL = P: DO {L=1, DO=2}
  PEKERJAAN_ORTU = Pegawai Negeri Sipil: L {L=116, DO=16}
  PEKERJAAN_ORTU = Pegawai Swasta: L {L=46, DO=7}
  PEKERJAAN_ORTU = Pensiunan PNS/TNI/POLRI: L {L=7, DO=2}
  PEKERJAAN_ORTU = Pensiunan Swasta: L {L=15, DO=2}
  PEKERJAAN_ORTU = TNI / POLRI: L {L=8, DO=5}
  PEKERJAAN_ORTU = Tidak Bekerja
    JKEL = L: L {L=2, DO=1}
    JKEL = P: DO {L=0, DO=2}
  PEKERJAAN_ORTU = Wiraswasta: L {L=318, DO=80}
  
```

Gambar 7.

Aturan atau *Rule Decision Tree* pada *Rapid Miner*

Hasil pengujian algoritma *Decision Tree C4.5* dengan data *training* sebanyak 80% dan data *testing* sebanyak 20% mendapatkan hasil *class recall true* L sebesar 96,41% dan *true DO* sebesar 3,23%. Hasil dari *class precision* sebesar *pred L* sebesar 84,29% dan *pred DO* sebesar 14,29%. Hasil tersebut berdasarkan *Confusion Matrix* yang ditunjukkan pada gambar berikut:

accuracy: 81.82%			
	true L	true DO	class precision
pred L	151	30	84.29%
pred DO	6	1	14.29%
class recall	95.41%	3.23%	

Gambar 8.
Confusion Matrix

Gambar 8 menunjukkan nilai akurasi sebesar 81,82% yang menunjukkan klasifikasi yang dilakukan termasuk dalam *good classification*, Menurut (Gorunescu, 2011) tingkat akurasi dengan *range accuracy* bernilai 0.80 – 0.90 = *good classification*.

V. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, penulis menarik beberapa kesimpulan diantaranya:

1. Data *mining* dengan algoritma *Decision Tree C4.5* dapat diterapkan untuk mengklasifikasi prediksi mahasiswa *drop out* pada Universitas Wiraraja.
2. Hasil pengujian menggunakan algoritma *Decision Tree C4.5* menggunakan *Rapid Miner* memiliki nilai *accuracy* sebesar 81,82% termasuk dalam *good classification* dengan *classification_error* sebesar 18.18%.

REFERENSI

- Kemenristekdikti. (2015). Standar Nasional Pendidikan Tinggi. Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 44 Tahun 2015 Tentang.
- Han, J., Pei, J. and Kamber, M., (2011). *Data Mining: Concepts And Techniques. Third Edition*, Elsevier, USA.
- Ridwan, M., Suyono, H. and Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier. ECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*. 7(1).59-64.

E. Pramadani, H. Sunandar, and Y. Hasan. (2018). Implementasi Data *Mining* Penjualan Koran Dengan Metode *C4.5* (STUDI KASUS : PT . Media Massa Cahaya). 6(1).11–15.

A. W. Wardhana, E. Patimah, A. I. Shafarindu, Y. M. Siahaan, B. V. Haekal, and D. S. Prasvita. (2021). Klasifikasi Data Penjualan pada Supermarket dengan Metode *Decision Tree*. Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (Senamika).2(1), 660-667.

Iddrus and Junaedi, H. (2022) .Prediksi Kelulusan Mahasiswa Menggunakan *Support Vector Machine* Berbasis *Particle Swarm Optimization*. SMARTICS Journal. Malang, Jawa Timur, Indonesia. 8(1).1–7.

Dewi, K. R., Mauladi, K.F., Masrurroh. (2020). Analisa Algoritma *C4.5* untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri”. Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi). 4(3). 109-114.

Siahaan, N. V. B., Poningsih, P., Suhendro, D., Hartama, D., & Suhada, S. (2022). Penerapan Data *Mining* Algoritma *C4.5* Terhadap Prediksi Faktor Menurunnya Hasil Panen Padi. Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika), 7(1), 27.

Anwar, F. F., Jaya, A. I., & Abu, M. (2022). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* dengan Penerapan Algoritma *C4.5*. Jurnal Ilmiah Matematika Dan Terapan, 19(1), 19–28.

Sunanto, N., & Falah, G. (2022). Penerapan Algoritma *C4.5* Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes. Rabbit: Jurnal Teknologi Dan Sistem Informasi Univrab, 7(2), 208–2

