

PREDIKSI MAHASISWA BERPOTENSI DROP OUT DENGAN METODE ITERATIF DICHOTOMISER 3 (ID3)

Dede Irmayanti, Yusuf Muhyidin, Dede Arif Nurjaman

Program Studi Teknik Informatika, STT Wastukancana

Jl. Cikopak No.53 Sadang Purwakarta

de2irmayanti@gmail.com, yusufshikudo20@gmail.com, arif.dearif17@gmail.com

Abstract - Education is the process of searching for knowledge by teaching and learning process during the education period. The high level of student success and the low level of student failure reflects the quality of a college. The large number of students who fail when completing education can be assessed against the grades of tertiary institutions which should be considered as early as possible students potentially fail. One way that can be used is to use classification data mining methods. The method used in this research is Knowledge Discovery in Database (KDD), the data processing using the id3 algorithm and the application used is Orange. The amount of data used in this study were 532 students of STT Wastukancana class of 2013 and 577 students of class of 2017 as material to make predictions with attributes worth IPS semester 1-4 and Total SKS. The calculation results are evaluated using a confusion matrix. This study produces a decision tree that is used to determine students who are dropped out with a root node is IPS4. Actual number of Drop outs, Predicted Drop out of 150 data, Actual Number of Drop outs, Predicted Not as much as 55 data, Actual Number of No, Predicted Drop out of 24 data, and Actual Number of No, Predicted Not to amount to 238 data. The calculation accuracy is 83,1%, 86.2% precision and 73.2% recall.

Keywords - data mining, KDD, drop out, id3, decision tree.

Abstrak - Pendidikan adalah proses mencari ilmu pengetahuan yang didapatkan dalam proses belajar mengajar selama masa pendidikan yang ditempuh. Tingginya tingkat keberhasilan mahasiswa dan rendahnya tingkat kegagalan mahasiswa merupakan cermin kualitas dari suatu perguruan tinggi. Banyaknya jumlah mahasiswa yang gagal dalam menyelesaikan pendidikan dapat berpengaruh terhadap nilai dari perguruan tinggi tersebut sehingga perlu dilakukan pencegahan sedini mungkin dengan melakukan prediksi terhadap mahasiswa yang berpotensi *Drop out*. Salah satu cara yang dapat digunakan adalah dengan menggunakan metode *data mining* klasifikasi. Metode yang digunakan pada penelitian ini adalah *Knowledge Discovery in Database* (KDD), proses pengolahan datanya menggunakan algoritma id3 aplikasi yang digunakan adalah *Orange*. Jumlah data yang digunakan pada penelitian ini sebanyak 476 mahasiswa Sekolah Tinggi Teknologi Wastukancana Purwakarta angkatan tahun 2013 dan 577 mahasiswa angkatan 2017 sebagai bahan untuk melakukan prediksi dengan atribut yang dinilai adalah nilai IPS semester 1-4 dan Total SKS. Hasil perhitungan dievaluasi akurasi dengan menggunakan *confusion matrix*. Penelitian ini menghasilkan *decision tree* yang digunakan untuk menentukan mahasiswa yang berpotensi *Drop out* dengan node akar adalah IPS4. Jumlah *actual Drop out*, *predicted Drop out* sebanyak 150 data, Jumlah *actual Drop out*, *predicted* Tidak sebanyak 55 data, Jumlah *actual* Tidak, *predicted Drop out* sebanyak 24 data, dan Jumlah *actual* Tidak, *predicted* Tidak sebanyak 238 data. Nilai akurasi perhitungannya sebesar 83,1%, presisi 86,2% dan *recall* 73,2%.

Kata Kunci - data mining, KDD, drop out, id3, decision tree.

I. PENDAHULUAN

Perkembangan teknologi informasi memberikan peluang kepada institusi untuk memanfaatkan teknologi berbasis komputer dalam pengolahan data. Seperti halnya dalam dunia Pendidikan, data akademik dapat diolah menjadi informasi yang sangat berguna untuk menentukan kebijakan-kebijakan maupun strategi dalam peningkatan mutu pendidikan. Informasi dan pengetahuan dapat diperoleh melalui proses pencarian pola dengan menggunakan *data mining*.

Telah banyak dilakukan penelitian mengenai data mining dalam dunia Pendidikan seperti pada penelitian yang dilakukan oleh aprilianto dan yosepta, penelitian ini bertujuan untuk menganalisis pola kelulusan mahasiswa pada sekolah tinggi manajemen informatika dan komputer asia malang dengan menggunakan algoritma

id3. Pada penelitian ini disimpulkan bahwa atribut yang paling menentukan adalah nilai indeks prestasi semester (IPS) pada semester 5[1]. Pada penelitian yang dilakukan oleh Hastuti dan Hidayat tentang analisis algoritma *decision tree* untuk prediksi mahasiswa non aktif dapat disimpulkan bahwa *decision tree* dapat menghasilkan nilai akurasi yang sangat baik sehingga sangat efektif jika digunakan untuk proses pengambilan keputusan[2].

Sekolah Tinggi Teknologi Wastukancana Purwakarta merupakan salah satu perguruan tinggi swasta di purwakarta yang sukses menarik banyak mahasiswa disetiap periodenya. Namun ada beberapa hal yang tidak seimbang antara masuk dan keluar mahasiswa di Sekolah Tinggi Teknologi Wastukancana. Berdasarkan data yang diperoleh dari hasil wawancara, jumlah mahasiswa *Drop out* setiap tahunnya cukup banyak. Data

mahasiswa angkatan tahun 2013-2016 dapat dilihat pada tabel 1.

Tabel 1 Data Mahasiswa Angkatan Tahun 2013-2016

Tahun	Terdaftar	Lulus	Aktif	Drop Out
2013	570	270	42	258
2014	552	265	108	179
2015	633	267	184	182
2016	603	34	413	156

Menurut buku panduan akademik Sekolah Tinggi Teknologi Wastukencana tahun 2014, mahasiswa yang masuk ke dalam kategori *Drop out* (Pemutusan Studi) adalah mahasiswa yang predikat akademiknya <2,00 pada akhir semester IV, jumlah sks nya <44 dan melewati masa maksimal studi yaitu 10 semester untuk Jenjang Diploma-3, dan mahasiswa yang predikat akademiknya <2,00 pada akhir semester VI, jumlah sks nya <72 dan melewati masa maksimal studi yaitu 14 semester untuk Jenjang Strata 1[3].

Berdasarkan permasalahan dan hasil dari beberapa penelitian di atas dapat disimpulkan bahwa penggunaan algoritma id3 dapat digunakan untuk menganalisis data sehingga dapat dilakukan prediksi berdasarkan pola yang terbentuk dari proses data mining. Hal ini menjadi dasar penelitian yang akan dilakukan yaitu mengenai prediksi mahasiswa berpotensi drop out di sekolah tinggi teknologi wastukencana

A. Algoritma Iterative Dichotomiser 3 (ID3)

Algoritma ID3 adalah algoritma pembelajaran pohon keputusan yang sederhana dikembangkan oleh J. Ross Quinlan pada tahun 1993 [4].

Menurut Mcg David, Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun decision tree (pohon keputusan) secara top-down[5]. Karakteristik ID3 dalam membangun pohon keputusan adalah secara top-down dan divide-and-conquer. Top-down artinya pohon keputusan dibangun dari simpul akar ke daun, sementara divide-and-conquer artinya training data secara rekursif dipartisi ke dalam bagian-bagian yang lebih kecil saat pembangunan pohon.

Langkah kerja algoritma *Iterative Dichotomiser 3* adalah:

1. Pilih atribut sebagai akar
Untuk memilih atribut sebagai akar, didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada dan perhitungan entropy.
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.
5. Ekstraksi Rule dari *Decision Tree*

B. Entropy

Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data [6]. Menurut Rokach dan Maimoon, *Information gain* atau biasa disebut *gain* info adalah kriteria pemisahan yang menggunakan pengukuran *entropy*. Untuk mendapatkan *information gain* dari suatu atribut dibutuhkan *entropy* keseluruhan kelas atau *Entropy(S)* [7], secara matematis *entropy* dirumuskan sebagai berikut [8]:

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

Dimana:

- S : himpunan kasus
- n : jumlah partisi S
- pi : proporsi dari Si terhadap S

C. Information Gain

Setelah mendapat nilai *entropy* untuk suatu kumpulan data, maka kita dapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut *information gain*. Secara matematis, *information gain* dari suatu atribut A, dirumuskan sebagai berikut [8]:

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

Dimana:

- S : himpunan kasus
- A : atribut
- V : menyatakan suatu nilai yang mungkin untuk atribut A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S
- Entropy (Si) : *entropy* untuk sampel-sampel yang memiliki nilai i.

D. Confusion Matrix

Confusion Matrix adalah tools yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai actual dan prediksi pada klasifikasi [9].

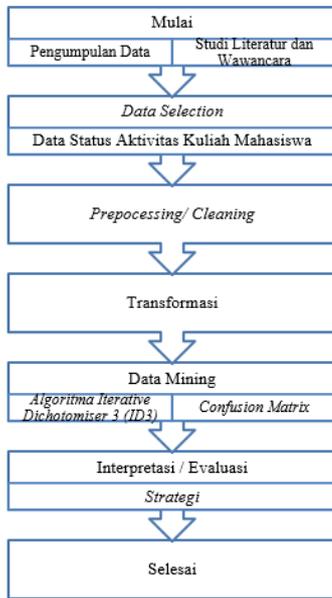
Contoh *confusion matrix* dapat dilihat pada tabel 2.

Tabel 2 *Confusion Matrix*

Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a (true positive - TP)	b (False Negative - FN)
Class = No	c (false positive - FP)	d (true negatif - TN)

II. METODE PENELITIAN

Menurut Larose, data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang sebelumnya yang dapat dipahami dan bermanfaat bagi pemilik data [10]. Proses *Knowledge Discovery Database* (KDD) secara garis besar dapat dijelaskan yaitu sebagai berikut : 1. Pembersihan Data (*Cleaning Data*), 2. Transformasi (*Transformation*), 3. *Data mining*, 4. Interpretasi/ evaluasi (*interpretation/ evaluation*) [11]. Kerangka penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Kerangka Penelitian

Tahapan proses KDD yang dilakukan pada penelitian ini adalah sebagai berikut:

1. *Data Selection*
 Pada tahap ini, penulis melakukan pemilihan (seleksi) data dari sekumpulan data status aktivitas kuliah mahasiswa angkatan 2013 sebanyak 467 orang dan sebanyak 577 data mahasiswa angkatan 2017 sebagai data untuk melakukan prediksi.
2. *Preprocessing / cleaning*
 Pada proses *cleaning*, penulis membuang atribut data yang tidak diperlukan, duplikasi data, dan memeriksa data yang inkonsisten.
3. *Transformation*
 Data-data yang telah melalui proses *selection* dan *preprocessing* tidak bisa langsung digunakan, selanjutnya pada tahap ini dilakukan pembentukan data ke dalam bentuk yang bisa diterapkan untuk proses selanjutnya.
4. *Data mining*
 Pada tahap ini penulis melakukan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan algoritma *Iteratif Dichotomiser (ID3)* dan aplikasi yang digunakan adalah *Orange*.
5. *Interpretation / Evaluation*

Pada tahap ini, penulis menganalisis pola-pola yang telah diidentifikasi oleh aplikasi *Orange* dan kemudian diterjemahkan atau diinterpretasikan ke dalam bentuk yang bisa dimengerti untuk membantu dalam perencanaan strategi bisnis..

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Tahap ini merupakan tahap dimana data yang dibutuhkan dalam penelitian ini dikumpulkan untuk dapat diolah selanjutnya. Data-data yang dikumpulkan adalah data aktifitas kuliah mahasiswa dan data lain yang diperoleh dari informasi dan kutipan, baik dari internet maupun dari literatur yang berkaitan dengan permasalahan yang dihadapi dan diteliti, data yang digunakan merupakan data primer yang dikumpulkan langsung dari proses wawancara di Sekolah Tinggi Teknologi Wastukencana Purwakarta. Sumber data utama yang digunakan dalam penelitian ini adalah dataset aktivitas kuliah mahasiswa angkatan 2013 pada Program Studi yaitu Teknik Informatika, Teknik Industri, dan Teknik Mesin sejumlah 570 record yang terdiri dari 264 orang sudah lulus, 42 orang masih aktif kuliah sampai tahun 2020, 1 orang dikeluarkan, 39 orang dengan status hilang, 206 status mengundurkan diri, mutasi sebanyak 17 orang, dan wafat 1 orang. Contoh data awal yang dikumpulkan dapat dilihat pada tabel 3.

Tabel 3 Data Awal

No	NIM	Nama	Program Studi	Status Mahasiswa	Jenis Pendaftaran	Biaya Masuk	Jenis	Tempat, Tanggal Lahir	Agama	IPS 1	SKS1	IPS2	SKS2	IPS3	SKS3	IPS4	SKS4	Total SKS
1	131251001	Aa Rahmat Fauzi	S1 Teknik Mesin	Mengundurkan diri	Peserta didik baru	0	L	Subang,20-05-1995	Tidak diisi	0	20	0	0	2,6	20	2,55	20	60
2	131251002	Adi Erwin Herawan	S1 Teknik Mesin	Mutasi	Peserta didik baru	0	L	Purwakarta,27-10-1993	Tidak diisi	2,05	20	2,65	20	2,1	20	0	0	60
3	131251003	Agam Gilang Pamungkas	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Purwakarta,31-08-1995	Tidak diisi	2,75	20	2,9	20	3,25	20	3,25	20	80
4	131251004	Alfian Wijaya	S1 Teknik Mesin	Mengundurkan diri	Peserta didik baru	0	L	Karawang,26-12-1994	Tidak diisi	2,65	20	0	0	0	0	0	0	20
5	131251005	Angga Anggara	S1 Teknik Mesin	Mutasi	Peserta didik baru	0	L	Purwakarta,29-06-1994	Tidak diisi	2,4	20	2,35	20	1,95	20	1,3	20	80
6	131251006	Anton Wahyudi Irawan	S1 Teknik Mesin	Hilang	Peserta didik baru	0	L	Subang,13-02-1995	Tidak diisi	2,55	20	3,25	20	2,45	20	1,95	20	80
7	131251007	Arif Ramadhani	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Bandung,04-02-1996	Tidak diisi	3	20	3	20	3	22	3,4	20	82
8	131251008	Bani Ajinudin Noor	S1 Teknik Mesin	Hilang	Peserta didik baru	0	L	Cirebon,14-12-1995	Tidak diisi	2,05	20	1,15	20	1,8	20	0	0	60
9	131251009	Dede Suardi	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Subang,27-07-1994	Tidak diisi	2,5	20	2,95	20	2,8	20	3,3	20	80
10	131251010	Deni Nugraha	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Purwakarta,08-02-1995	Tidak diisi	2,9	20	3,1	20	3,25	20	3,4	20	80
11	131251011	Dhiar Eko Prasetyo	S1 Teknik Mesin	Mengundurkan diri	Peserta didik baru	0	L	Purwakarta,17-03-1995	Tidak diisi	3,05	20	2,85	20	2,2	20	3	20	80
12	131251012	Dimas Ardhiyanto	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Cimahi,04-10-1995	Tidak diisi	2,6	20	2,85	20	2,3	20	2,8	20	80
13	131251013	Dinidn Maryadin	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Purwakarta,04-12-1995	Tidak diisi	3	20	2,95	20	2,65	20	2,45	20	80
14	131251014	Endang Kusnandar	S1 Teknik Mesin	Mutasi	Peserta didik baru	0	L	Purwakarta,03-02-1995	Tidak diisi	2,35	20	1,45	20	0	0	0	0	40
15	131251015	Fajar Alamsyah	S1 Teknik Mesin	Lulus	Peserta didik baru	0	L	Tangerang,21-10-1995	Tidak diisi	2,6	20	2,95	20	2,35	20	2,4	20	80

Penjelasan atribut data dapat dilihat pada table 4.

Tabel 4 Keterangan Atribut

No.	Atribut	Keterangan
1.	NO	Nomor urut data
2.	NIM	Nomor Induk Mahasiswa
3.	NIK	Nomor Induk Kependudukan
4.	Nama	Nama Mahasiswa
5.	Program Studi	Program Studi Teknik Infomatika, Teknik Industri, Teknik mesin, Manajemen Industri dan Teknik Tekstil.
6.	Status Mahasiswa	Aktifitas kuliah mahasiswa
7.	Jenis Pendaftaran	Peserta didik baru / konversi
8.	Biaya Masuk	Biaya yang dikeluarkan diawal pendaftaran
9.	Jenis Kelamin	Laki-laki dan Perempuan
10.	Tempat Tanggal Lahir	Tempat dan Tanggal Lahir Mahasiswa
11.	Agama	Agama Mahasiswa
12.	Alamat	Alamat Mahasiswa
13.	IPS1	Indek Prestasi Semester 1
14.	SKS1	Satuan Kredit Semester 1
15.	IPS2	Indek Prestasi Semester 2
16.	SKS2	Satuan Kredit Semester 2
17.	IPS3	Indek Prestasi Semester 3
18.	SKS3	Satuan Kredit Semester 3
19.	IPS4	Indek Prestasi Semester 4
20.	SKS4	Satuan Kredit Semester 4
21.	Total SKS	Total Satuan Kredit Semester 1-4

B. Proses KDD

Selanjutnya akan dilakukan proses analisis data dengan tahapan sebagai berikut:

1. Data Selection

Dari 21 atribut data, yang akan digunakan pada proses klasifikasi ini hanya 6 atribut saja yaitu **IPS1, IPS2, IPS3, IPS4, Total SKS dan Status Mahasiswa**. Contoh data hasil seleksi yang akan digunakan untuk proses *data mining* dapat dilihat pada tabel 5.

Tabel 5 Tabel Data Selection

IPS 1	IPS 2	IPS 3	IPS 4	Total SKS	Status Mahasiswa
0	0	2,6	2,55	60	Mengundurkan diri
2,05	2,65	2,1	0	60	Mutasi
2,75	2,9	3,25	3,25	80	Lulus
2,65	0	0	0	20	Mengundurkan diri

IPS 1	IPS 2	IPS 3	IPS 4	Total SKS	Status Mahasiswa
					an diri
2,4	2,35	1,95	1,3	80	Mutasi
2,55	3,25	2,45	1,95	80	Hilang
3	3	3	3,4	82	Lulus
2,05	1,15	1,8	0	60	Hilang
2,5	2,95	2,8	3,3	80	Lulus
2,9	3,1	3,25	3,4	80	Lulus
3,05	2,85	2,2	3	80	Mengundurkan diri
2,6	2,85	2,3	2,8	80	Lulus
3	2,95	2,65	2,45	80	Lulus
2,35	1,45	0	0	40	Mutasi
2,6	2,95	2,35	2,4	80	Lulus
3,2	2,65	2,85	3,4	80	Lulus
2,65	2,85	2,25	2,05	80	Lulus
2,4	0	0	0	20	Mengundurkan diri
3,2	3,2	2,85	2,65	80	Lulus
3	3,15	3,25	3	82	Lulus

Tahap selanjutnya, beberapa atribut dilakukan proses klasifikasi data, diantaranya IPS, Total SKS dan Status mahasiswa. Klasifikasi atribut nilai data mahasiswa dapat dilihat pada tabel 6.

Tabel 6 Klasifikasi Atribut Nilai Data Mahasiswa

Atribut	Range	Nilai Atribut
IPS1, IPS2, IPS3, IPS4	0,00-1,30 1,31-2,60 2,61-4,00	Kurang Cukup Baik
Total SKS	-	<72 ≥72
Status Mahasiswa	-	Drop out Tidak

2. Preprocessing/cleaning

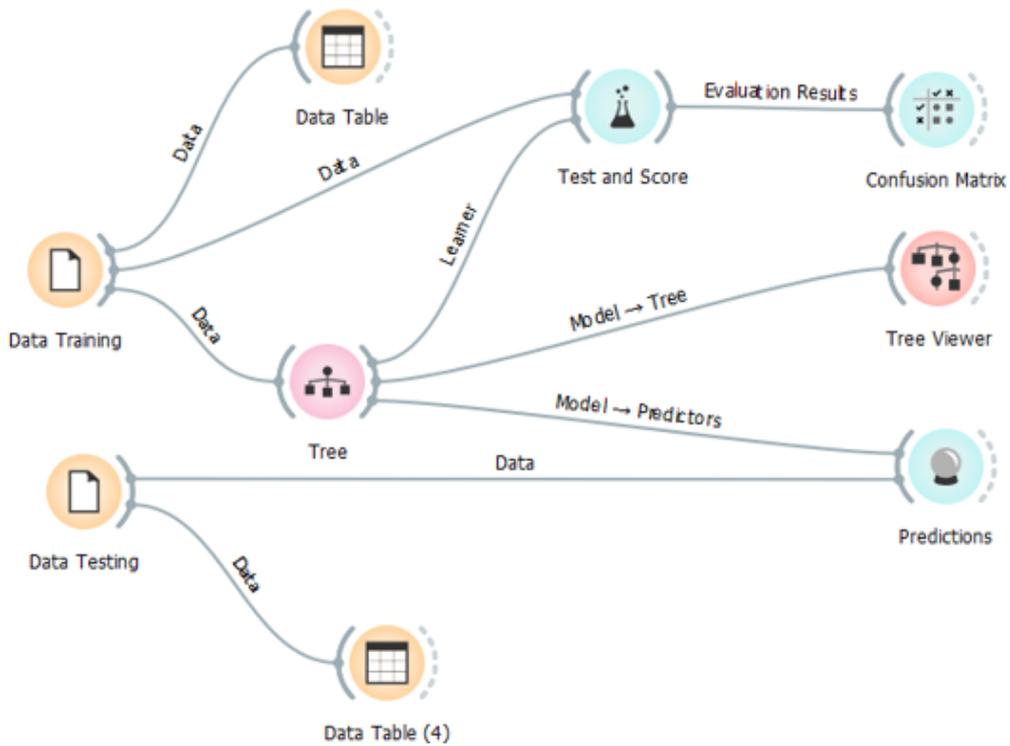
Data *cleaning* merupakan tahap *preprocessing* yang dilakukan untuk menghapus redundansi data, menghilangkan data mahasiswa konversi, mutasi dan wafat, merubah status mahasiswa mengundurkan diri, hilang, dikeluarkan menjadi **Drop out**, mengganti status aktif dan lulus menjadi **Tidak**

3. Transformasi

Setelah tahap *preprocessing/cleaning*, data yang didapat dari dataset bertipe numerik, sedangkan pengujian ini memerlukan data tipe kategori. Teknik

2. Membuat Pemodelan

Setelah data diupload, selanjutnya dilakukan pemodelan seperti yang terlihat pada gambar 4.



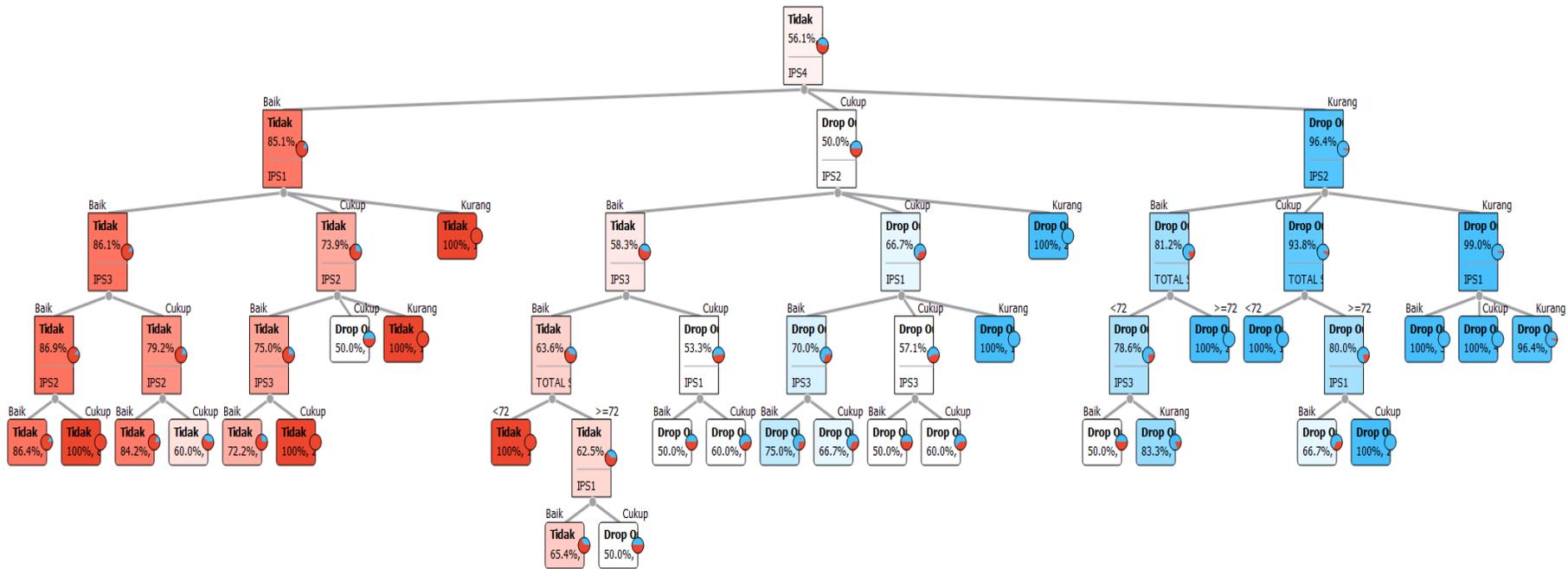
Gambar 4 Pemodelan Menggunakan Aplikasi Orange

Berikut penjelasan pemodelan yang digunakan:

- a. *Data Training*
Data training merupakan dataset yang terdiri dari 467 mahasiswa angkatan 2013.
- b. *Data Testing*
Data testing merupakan dataset yang terdiri dari 577 mahasiswa angkatan 2017.
- c. *Data Table*
Data table berfungsi untuk menampilkan data yang telah diupload ke aplikasi.
- d. *Tree*
Tree merupakan model yang digunakan pada penelitian ini (*Decision Tree* Algoritma ID3)
- e. *Tree Viewer*
Tree Viewer berfungsi untuk menampilkan *tree* yang terbentuk dari hasil pengolahan data.
- f. *Prediction*
Prediction berfungsi untuk menampilkan hasil prediksi dari pengolahan 577 data mahasiswa angkatan 2017, berdasarkan rules yang dihasilkan oleh perhitungan dari 467 data.
- g. *Test and Score* dan *Confusion Matrix* berfungsi untuk mengevaluasi hasil pemodelan *Decision Tree*.

3. Proses Pembentukan *Decision Tree*

Gambar 5 menunjukkan *Decision tree* hasil pengolahan data menggunakan aplikasi orange.



Gambar 5 Decision Tree

Proses pengolahan data menghasilkan rules sebagai berikut:

Tree

1. IPS4 = Baik, IPS1 = Baik, IPS3 = Baik, IPS2 = Baik, : 86,4% **Tidak** { }
2. IPS4 = Baik, IPS1 = Baik, IPS3 = Baik, IPS2 = Cukup, : 100% **Tidak** { }
3. IPS4 = Baik, IPS1 = Baik, IPS3 = Cukup, IPS2 = Baik, : 84,2% **Tidak** { }
4. IPS4 = Baik, IPS1 = Baik, IPS3 = Cukup, IPS2 = Cukup, : 60% **Tidak** { }
5. IPS4 = Baik, IPS1 = Cukup, IPS2 = Baik, IPS3 = Baik, : 72,2% **Tidak** { }
6. IPS4 = Baik, IPS1 = Cukup, IPS2 = Baik, IPS3 = Cukup, : 100% **Tidak** { }
7. IPS4 = Baik, IPS1 = Cukup, IPS2 = Cukup, : 50% **Drop out** { }
8. IPS4 = Baik, IPS1 = Cukup, IPS2 = Kurang, : 100% **Tidak** { }
9. IPS4 = Baik, IPS1 = Kurang, : 100% **Tidak** { }
10. IPS4 = Cukup, IPS2 = Baik, IPS3 = Baik, Total Kasus = <72 : 100% **Tidak** { }
11. IPS4 = Cukup, IPS2 = Baik, IPS3 = Baik, Total Kasus = >=72, IPS1 = Baik, : 65,4% **Tidak** { }
12. IPS4 = Cukup, IPS2 = Baik, IPS3 = Baik, Total Kasus = >=72, IPS1 = Cukup, : 50 % **Drop out** { }
13. IPS4 = Cukup, IPS2 = Baik, IPS3 = Cukup, IPS1 = Baik, : 50% **Drop out** { }
14. IPS4 = Cukup, IPS2 = Baik, IPS3 = Baik, IPS1 = Cukup, : 60% **Drop out** { }
15. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Baik, IPS3 = Baik, : 75% **Drop out** { }
16. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Baik, IPS3 = Cukup, : 66,7% **Drop out** { }
17. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Cukup, IPS3 = Baik, : 50% **Drop out** { }
18. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Cukup, IPS3 = Cukup, : 60% **Drop out** { }
19. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Kurang, : 100% **Drop out** { }
20. IPS4 = Cukup, IPS2 = Kurang, : 100% **Drop out** { }
21. IPS4 = Kurang, IPS2 = Baik, Total Kasus = <72, IPS3 = Baik, : 50% **Drop out** { }
22. IPS4 = Kurang, IPS2 = Baik, Total Kasus = <72, IPS3 = Kurang, : 83,3% **Drop out** { }
23. IPS4 = Kurang, IPS2 = Baik, Total Kasus = >=72, : 100% **Drop out** { }
24. IPS4 = Kurang, IPS2 = Cukup, Total Kasus = <72, : 100% **Drop out** { }
25. IPS4 = Kurang, IPS2 = Cukup, Total Kasus = >=72, IPS1 = Baik : 66,7% **Drop out** { }
26. IPS4 = Kurang, IPS2 = Cukup, Total Kasus = >=72, IPS1 = Cukup : 100% **Drop out** { }
27. IPS4 = Kurang, IPS2 = Kurang, IPS1 = Baik : 100% **Drop out** { }

28. IPS4 = Kurang, IPS2 = Kurang, IPS1 = Cukup: 100% **Drop out** { }
29. IPS4 = Kurang, IPS2 = Kurang, IPS1 = Baik : 96,4% **Drop out** { }

Berdasarkan rules yang terbentuk, dapat diketahui bahwa mahasiswa berpotensi *Drop out* adalah mahasiswa dengan:

1. IPS4 = Baik, IPS1 = Cukup, IPS2 = Cukup, : 50% **Drop out** { }
2. IPS4 = Cukup, IPS2 = Baik, IPS3 = Baik, Total Kasus = >=72, IPS1 = Cukup, : 50 % **Drop out** { }
3. IPS4 = Cukup, IPS2 = Baik, IPS3 = Cukup, IPS1 = Baik, : 50% **Drop out** { }
4. IPS4 = Cukup, IPS2 = Baik, IPS3 = Baik, IPS1 = Cukup, : 60% **Drop out** { }
5. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Baik, IPS3 = Baik, : 75% **Drop out** { }
6. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Baik, IPS3 = Cukup, : 66,7% **Drop out** { }
7. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Cukup, IPS3 = Baik, : 50% **Drop out** { }
8. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Cukup, IPS3 = Cukup, : 60% **Drop out** { }
9. IPS4 = Cukup, IPS2 = Cukup, IPS1 = Kurang, : 100% **Drop out** { }
10. IPS4 = Cukup, IPS2 = Kurang, : 100% **Drop out** { }
11. IPS4 = Kurang, IPS2 = Baik, Total Kasus = <72, IPS3 = Baik, : 50% **Drop out** { }
12. IPS4 = Kurang, IPS2 = Baik, Total Kasus = <72, IPS3 = Kurang, : 83,3% **Drop out** { }
13. IPS4 = Kurang, IPS2 = Baik, Total Kasus = >=72, : 100% **Drop out** { }
14. IPS4 = Kurang, IPS2 = Cukup, Total Kasus = <72, : 100% **Drop out** { }
15. IPS4 = Kurang, IPS2 = Cukup, Total Kasus = >=72, IPS1 = Baik : 66,7% **Drop out** { }
16. IPS4 = Kurang, IPS2 = Cukup, Total Kasus = >=72, IPS1 = Cukup : 100% **Drop out** { }
17. IPS4 = Kurang, IPS2 = Kurang, IPS1 = Baik : 100% **Drop out** { }
18. IPS4 = Kurang, IPS2 = Kurang, IPS1 = Cukup: 100% **Drop out** { }
19. IPS4 = Kurang, IPS2 = Kurang, IPS1 = Baik : 96,4% **Drop out** { }

Setelah dilakukan proses pemodelan, selanjutnya akan dilakukan proses evaluasi untuk membuktikan keakuratan terhadap *tree* yang terbentuk.

		Predicted		
		Drop Out	Tidak	Σ
Actual	Drop Out	150	55	205
	Tidak	24	238	262
Σ		174	293	467

Gambar 6 Confussion Matrix

Hasil evaluasi dengan menggunakan *confusion matrix* menunjukkan bahwa:

1. Jumlah actual *Drop out*, predicted *Drop out* sebanyak 150 data
2. Jumlah actual *Drop out*, predicted Tidak sebanyak 55 data
3. Jumlah actual Tidak, predicted *Drop out* sebanyak 24 data
4. Jumlah actual Tidak, predicted Tidak sebanyak 238 data

Nilai akurasi dari perhitungannya adalah:

1. $Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$
 $= \frac{150 + 238}{150 + 238 + 55 + 24} \times 100\%$
 $= 83,1 \%$
2. $Presisi = \frac{TP}{TP+FP} \times 100\%$
 $= \frac{150}{150 + 24} \times 100\%$
 $= 86,2 \%$
3. $Recall = \frac{TP}{TP+FN} \times 100\%$
 $= \frac{150}{150 + 55} \times 100\%$
 $= 73,2 \%$

D. Analisis Prediksi Data

Setelah dilakukan pembentukan *Tree*, selanjutnya akan dilakukan proses prediksi data mahasiswa angkatan 2017 sebanyak 577. Pada proses ini dianalisa apakah hasil prediksi sesuai dengan *rules* yang terbentuk pada *Decision Tree*. Dataset hasil prediksi dapat dilihat pada table 7.

Tabel 2 Dataset Hasil Prediksi

NIM	Nama	IPS 1	IPS 2	IPS 3	IPS 4	TOTAL SKS	Prediksi
17115 1001	Abizar Al-Afkar Fauzan	Kurang	Kurang	Baik	Kurang	>=72	Drop out
17115 1002	Acep Restu Rusma	Kurang	Kurang	Cukup	Kurang	>=72	Drop out

NIM	Nama	IPS 1	IPS 2	IPS 3	IPS 4	TOTAL SKS	Prediksi
	na						
17115 1003	Adam Sundana	Kurang	Kurang	Kurang	Kurang	<72	Drop out
17115 1004	Adi Sumarna	Kurang	Kurang	Cukup	Cukup	>=72	Drop out
17115 1005	Adli Handika	Kurang	Kurang	Kurang	Kurang	<72	Drop out
17115 1006	Adrian Pratama	Kurang	Kurang	Cukup	Kurang	>=72	Drop out
17115 1007	Adytio Handianto	Kurang	Kurang	Cukup	Cukup	>=72	Drop out
17115 1008	Afifah Nuraini	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1009	Afrizal yuriz Ruhiat Nurfirdaus	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1010	Agung Gumelar	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1011	Agus Ahmad Kosim	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1012	Ahmad Yusup	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1013	Aji Mauludi	Kurang	Kurang	Cukup	Cukup	>=72	Drop out
17115 1014	Aji Prasetyo	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1015	Aldi Septianto	Kurang	Kurang	Cukup	Kurang	<72	Drop out
17115 1016	Aldian Yusuf Gunawan	Kurang	Kurang	Cukup	Baik	>=72	Tidak
17115 1017	Alin Asry Nugraha	Kurang	Kurang	Baik	Baik	>=72	Tidak
17115 1018	Andi Adryan	Kurang	Kurang	Baik	Kurang	<72	Drop out
17115 1019	Andi Sophia n	Kurang	Kurang	Kurang	Kurang	<72	Drop out

Berdasarkan hasil prediksi 577 data, dapat diketahui bahwa jumlah mahasiswa berpotensi *Drop out* sebanyak 281 orang dan mahasiswa yang tidak berpotensi *Drop out* sebanyak 296 orang. Bagi mahasiswa yang diprediksi *Drop out*, selanjutnya dilakukan bimbingan khusus untuk menghindari kemungkinan *Drop out* terjadi.

E. Evaluasi

Berdasarkan hasil pengolahan *data mining*, dapat dilakukan perbaikan-perbaikan strategi untuk mengurangi jumlah mahasiswa berpotensi *Drop out* di STT Wastukencana, diantaranya:

1. Melakukan prediksi data mahasiswa setiap angkatan, sehingga dapat diketahui apakah mahasiswa tersebut akan menyelesaikan pendidikan hingga lulus atau akan berpotensi *Drop out*.
2. Menambah fitur prediksi pada sistem informasi akademik, dimana dosen wali dan mahasiswa dapat melihat grafik perkembangan akademik mahasiswa. Sehingga bagi mahasiswa yang diprediksi *Drop out* akan diberikan pembinaan untuk menghindari potensi tersebut.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terhadap 467 data mahasiswa angkatan 2013, konstruksi Pohon Keputusan (*Decision tree*) yang terbentuk dengan menggunakan Algoritma ID3 menghasilkan pohon sebanyak 50 simpul dengan IPS4 terpilih sebagai akarnya. Pengukuran kinerja klasifikasi menunjukkan bahwa akurasi mencapai 83,1%, *precision* 86,2% dan *recall* sebesar 73,2%. Sedangkan berdasarkan hasil prediksi dari 577 data mahasiswa angkatan 2017, dapat diketahui bahwa jumlah mahasiswa berpotensi *Drop out* sebanyak 281 orang dan mahasiswa yang tidak berpotensi *Drop out* sebanyak 296 orang. Hasil dari prediksi data ini dapat dijadikan salah satu bahan untuk meminimalisir potensi mahasiswa *Drop out* di STT Wastukencana.

Saran untuk penelitian ini yaitu dapat dilakukan penelitian lain dengan menggunakan algoritma yang berbeda seperti algoritma Naïve Bayes. Sehingga diharapkan nilai akurasi akan semakin baik. Selain itu, atribut yang digunakan pada perhitungan dapat ditambah seperti jumlah penghasilan, tempat tinggal dan latar belakang sekolah, sehingga prediksi akan semakin bervariasi tidak hanya dari segi akademik, tapi juga dari segi ekonomi dan sosial.

DAFTAR PUSTAKA

- [1] D. A. Yosepta and Tria Aprilianto, "Analisa Pola Kelulusan Mahasiswa Pada Sekolah Tinggi Manajemen Informatika & Komputer Asia Malang Dengan Menggunakan algoritma Iterative Dichotomiser 3 (ID3)," *Positif*, vol. 3, no. 1, pp. 47–55, 2017.
- [2] K. Hastuti and E. Y. Hidayat, "Analisis Algoritma Decision Tree untuk Prediksi Mahasiswa Non Aktif," *Semantik*, vol. 2013, no. November, pp. 211–216, 2013.
- [3] STT Wastukencana, *Buku Panduan Akademik*. Purwakarta, 2014.
- [4] D. K. Widiyati, M. Wati, and H. S. Pakpahan, "Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," *J. Rekayasa Teknol. Inf.*, vol. 2, no. 2, p. 125, 2018, doi: 10.30872/jurti.v2i2.1864.
- [5] I. Srimenganti, I. Taufik, and E. Mulyana, "Implementasi Algoritma Decision Tree (ID3) Untuk Penyakit Campak," *Semin. Nas. Tek. Elektro*, pp. 235–242, 2018.
- [6] T. Thi Bi Dan, S. Widya Sihwi, and R. Anggrainingsih, "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret," *J. Teknol. Inf. ITSmart*, vol. 4, no. 2, p. 84, 2016, doi: 10.20961/its.v4i2.1770.
- [7] A. E. Tyasti, D. Ispriyanti, and A. Hoyyi, "Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam Medis," *Gaussian*, vol. 4, no. Dm, pp. 237–246, 2015.
- [8] T. D. Utama, S. W. Sihwi, and A. Doewes, "Implementasi Algoritma Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha Uns," *J. Itsmart*, vol. 3, no. 2, pp. 74–82, 2014.
- [9] Dwi Untari, "Data Mining untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5," *Fak. Ilmu Komput. Univ. Dian Nuswantoro*, 2010.
- [10] F. Pratiwi, "SATIN – Sains dan Teknologi Informasi Analisa Pemilihan Bahasa Pemograman dalam Penyelesaian Tugas Akhir Mahasiswa," vol. 4, no. 2, 2018.
- [11] Triase dan Samsudin, "Implementasi Data Mining Dalam Mengklasifikasikan Ukt (Uang Kuliah Tunggal) Pada Uin Sumatera Utara Medan," *J. Teknol. Inf. Jurti*, vol. 4, no. 2, pp. 370–376, 2020.