(JUSS) Jurnal Sains dan Sistem Informasi ISSN 2614-8277
Vol. 4 No. 2, Desember 2021

KLASIFIKASI MAHASISWA BERPOTENSI *DROP OUT MENGGUNAKAN ALGORITMA DECISION TREE C4.5* DAN *NAIVE BAYES* DI UNIVERSITAS JAMBI

(*Classification of Student Potential to Drop Out Using Decision Tree C4.5 and Naïve Bayes Algorithm in Jambi University*)

Isra Hayati, Jefri Marzal, Edi Saputra*
Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Jambi
*Jl. Jambi-Ma. Bulian KM 15 Mendalo Darat Jambi 36361
email: israhayati06@gmail.com

ABSTRAK

Drop out merupakan salah satu masalah yang pernah dihadapi oleh sebagian mahasiswa dan dapat mempengaruhi mutu pendidikan di suatu perguruan tinggi. Drop out dapat terjadi karena berbagai macam faktor. Salah satu cara untuk mengetahui faktor - faktor yang mempengaruhi mahasiswa *drop out* ialah dengan klasifikasi. Klasifikasi mahasiswa drop out dalam penelitian ini menggunakan beberapa atribut atau variabel, diantaranya Indeks Prestasi Komulatif (IPK), Satuan Kredit Semester (SKS) Total, Asal Sekolah, Pekerjaan Orang Tua, Penghasilan Orang Tua, Biaya Kuliah, Jenis Tempat Tinggal dan Status Mahasiswa. Algoritma klasifikasi yang digunakan yaitu Algoritma *Decision Tree* C4.5 dan Algoritma *Naive Bayes*. Penelitian ini dilakukan untuk membandingkan kinerja dari algoritma *Decision Tree* C4.5 dan *Naive Bayes* sehingga didapat algoritma paling baik dalam mengklasifikasikan mahasiswa berpotensi *drop out*. Hasil yang didapat dalam penelitian ini adalah algoritma *Decision Tree* C4.5 menghasilkan nilai *accuracy* dan *f1-score* sebesar 96,74% dan 86,55% sedangkan algoritma *Naive Bayes* menghasilkan nilai *accuracy* dan *f1-score* sebesar 96,24% dan 82,34%. Dari hasil tersebut dapat disimpulkan bahwa algoritma yang memiliki peforma paling baik dalam mengklasifikasikan mahasiswa *drop out* ialah Algoritma *Decision Tree C4.5*. Hasil penelitian ini diimplementasi menggunakan pohon keputusan dengan SKS Total sebagai *root* yang berarti SKS Total merupakan faktor paling berpengaruh dalam klasifikasi mahasiswa *drop out*. Rekomendasi untuk penelitian selanjutnya dapat membandingkan lebih dari dua algoritma atau menambah variabel lain dalam mengklasifikasikan mahasiswa berpotensi *drop out.*

Kata Kunci: Algoritma *Decision Tree C4.5*, Algoritma *Naïve Bayes*, Klasifikasi Mahasiswa *Drop Out*

I. PENDAHULUAN

Mutu pendidikan di perguruan tinggi erat kaitannya dengan tingginya tingkat keberhasilan mahasiswa dan rendahnya kegagalan mahasiswa. Salah satu indikator kegagalan mahasiswa adalah tidak bisa menyelesaikan studi pada waktu yang ditentukan sehingga mengakibatkan *drop out* (Nasrullah, 2018). Menurut Nurhayati, Kusrini & Luthfi (2015) *drop out* (DO) atau pemberhentian status mahasiswa adalah proses pencabutan status kemahasiswaan atas diri mahasiswa, yang disebabkan oleh hal-hal tertentu yang telah di tentukan oleh universitas atau perguruan tinggi. Potensi *drop out* dapat diminimalisir oleh perguruan tinggi dengan membuat suatu kebijakan atau peraturan mengenai standar pembelajaran maksimum.

Dalam peraturan Menteri Ristek Dikti nomor 44 tahun 2015 standar proses pembelajaran maksimum dalam masa studi adalah 7 tahun, IPK diatas 2.0, dan maksimum sks adalah 144 untuk program Sarjana. Meskipun masa studi maksimum yang ditentukan cukup panjang, memiliki batasan tersendiri dalam menjaga kualitas dan kinerja mahasiswa agar dapat lulus tepat waktu dan terhindar dari kasus *drop out* merupakan hal yang sangat diperlukan oleh perguruan tinggi atau universitas (Atma & Setyanto, 2018).

Universitas Jambi merupakan salah satu perguruan tinggi yang memiliki rata-rata lulusan lebih dari 4 tahun atau diatas 8 semester. Hal ini didasarkan pada data dari Statistik Universitas Jambi tahun 2018 yang mana jumlah mahasiswa baru yang melakukan registrasi ulang sebesar 5.774 mahasiswa, dengan rata-rata jumlah lulusan pertahun sebanyak ±2.700 mahasiswa dan ±3.074

mahasiswa belum lulus tepat waktu (https://dss.unja.ac.id). Dari data tersebut diketahui bahwa jumlah lulusan lebih rendah dibanding dengan jumlah mahasiswa yang terlambat lulus, hal ini mengakibatkan terjadinya penumpukan mahasiswa karena setiap tahun ajaran baru jumlah mahasiswa yang diterima terus bertambah. Karena penumpukan jumlah mahasiswa tersebutlah dapat memicu terjadinya *drop out*.

Wijaya (2018) menyebutkan bahwa kasus *drop out* juga dapat terjadi karena berbagai faktor seperti rendahnya kemampuan akademik, faktor pembiayaan, domisili saat menempuh studi dan lain sebagainya. Untuk itu dalam penelitian ini akan dilakukan klasifikasi pada kasus *drop out* agar dapat diketahui faktor apa yang paling berpengaruh terhadap kasus tersebut sehingga dapat dilakukan kebijakan untuk mencegah ataupun mengurangi potensi mahasiswa yang terdeteksi *drop out*.

Salah satu penelitian yang berkaitan yaitu penelitian menurut Nasrullah tahun 2018 dalam klasifikasi mahasiswa berpotensi *drop out* menggunakan algoritma C4.5 yang terdiri dari 9 atribut penelitian. Dalam penelitian tersebut atribut paling berpengaruh terhadap potensi *drop out* ialah nilai IPS semester IV dengan menghasilkan 17 *rule* mahasiswa berpotensi *drop out*. Penelitian lain yang serupa juga dilakukan oleh Ade Putra (2017) mengenai prediksi mahasiswa *drop out* menggunakan algoritma C4.5 dengan 4 atribut penelitian. Hasil penelitian menunjukkan bahwa mahasiswa yang memiliki jumlah SKS dan jumlah semester yang tidak baik berpotensi mengalami *drop out.* Kemudian pada penelitian Wahyuni, Saragih & Perangin-Angin (2018) dalam menganalisa mahasiswa *drop out* menggunakan algoritma *decision tree* C4.5 dengan 6 atribut penelitian. Atribut paling berpengaruh terhadap mahasiswa yang mengalami *drop out* ialah asal sekolah.

Dari beberapa penelitian diatas dapat dilihat bahwa rata-rata algoritma klasifikasi yang digunakan ialah algoritma *Decision Tree C4.5*. Padahal menurut penelitian Yahya & Jananto (2019) algoritma yang memiliki kemampuan serupa (sama baiknya) dengan *decision tree C4.5* ialah *naive bayes*. Untuk itu penulis tertarik untuk melakukan perbandingan terhadap kedua algoritma tersebut.

Penelitian mengenai perbandingan algoritma *decision tree C4.5* dan *naive bayes* telah dilakukan oleh Suyadi, Setyanto & Alfattah (2017) dalam klasifikasi penerimaan mahasiswa baru tingkat universitas yang memperoleh hasil bahwa algoritma *naive bayes* memiliki tingkat akurasi paling baik. Sedangkan pada penelitian Anam & Santoso (2018) dalam klasifikasi penerima beasiswa, algoritma *decision tree* C4.5 memiliki tingkat akurasi paling baik dibandingkan algoritma *naive bayes.* Meski memiliki kemampuan yang serupa tetapi kedua algoritma tersebut pastinya memiliki keakuratan yang berbeda. Oleh karena itu, perbandingan terhadap algoritma C4.5 dan *naive bayes* dibutuhkan untuk dapat memberikan informasi mengenai performa paling baik dari algoritma tersebut sehingga klasifikasi yang dihasilkan lebih akurat.

Dalam penelitian ini, penulis akan melakukan "Klasifikasi Mahasiswa Berpotensi *Drop Out* Menggunakan Algoritma *Decision Tree* C4.5 dan *Naive Bayes* di Universitas Jambi" yang bertujuan untuk mengetahui algoritma terbaik berdasarkan hasil peforma tertinggi dalam mengklasifikasikan mahasiswa berpotensi *drop out* sehingga dapat dijadikan acuan untuk mengurangi mahasiswa *drop out*.

II. METODOLOGI PENELITIAN

Penelitian dilakukan di Universitas Jambi yang beralamat di Jl. Lintas Jambi- Muara Bulian KM. 15 Mendalo Indah, Muaro Jambi, Jambi.

Bahan Penelitian

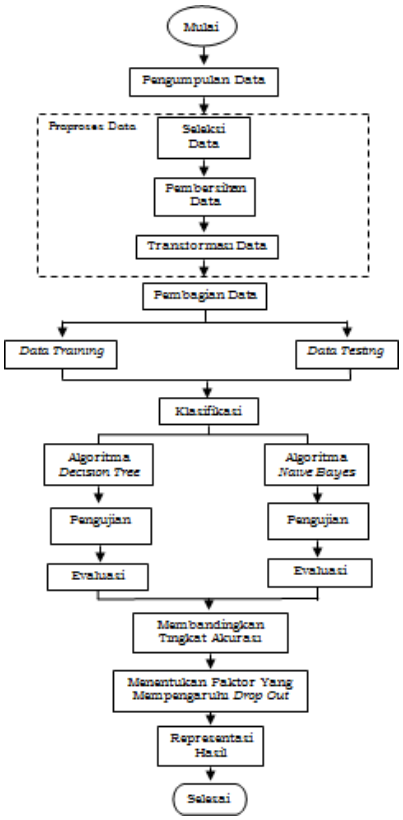
1. Studi literatur terkait penelitian tentang metode klasifikasi, algoritma *decision tree* C4.5 dan algoritma *naive bayes.*
2. Pengumpulan data akademik melalui Sistem Informasi Akademik (SIAKAD) yang diperoleh dari LPTIK Universitas Jambi.

Alat Penelitian

1. Perangkat keras: Laptop dengan RAM 4 GB
2. Perangkat lunak: Sistem Operasi *Windows* 8.1, *Python* 3.6.8, *Jupyter Notebook*, *Sublime Text* 3, dan *Microsoft Office* 2007

 ISSN 2614-8277

Kerangka Kerja Penelitian



Gambar 1. Kerangka Penelitian

a. Pengumpulan data
Data yang dikumpulkan diperoleh dari LPTIK Universitas Jambi yang terdiri dari data akademik mahasiswa S1 tahun 2018 pada semester genap angkatn 2012, 2015 dan 2017.

b. Praproses Data
Praproses data yang dilakukan terdiri dari:
- *Data Selection*
Dalam penelitian ini akan dilakukan beberapa proses seleksi data diantaranya menentukan atribut target, mengambil karakter awal data pada atribut asal sekolah, kategorisasi data yang memiliki banyak variansi data dan menghapus atribut-atribut yang tidak diperlukan.
- *Data Cleaning*
Dalam peneltian ini terdapat data kosong (*missing value)* sehingga perlu dihilangkan (dihapus). Untuk proses penghapusan data *missing value* akan dilakukan menggunakan *syntax python*.
- *Data Transformation*
Pada penelitian ini ada beberapa data yang merupakan data kategorikal sehingga perlu diubah kedalam bentuk angka agar dapat diproses pada tahap klasifikasi. Pengubahan data tersebut menggunakan pelabelan pada *python* yaitu *Label Encoder*.

c. Pembagian Data
Pada tahap ini data akan dibagi menjadi 2 bagian yaitu *data training* dan data *testing* dengan proporsi 80% : 20%.

d. Klasifikasi Data
Pada proses ini dilakukan klasifikasi data menggunakan algoritma *Decision tree* C4.5 dan algoritma *Naive bayes*.

e. Pengujian
Pengujian yang dilakukan dalam penelitian ini menggunakan metode *k-folds cross validation* dengan k=10 (10-*fold cross validation)*. Data akan dibagi menjadi 10 bagian secara acak sehingga didapat akurasi optimal dari data.

f. Evaluasi
Evaluasi hasil klasifikasi dari masing-masing algoritma menggunakan teknik *confussion matrix* dengan menghasilkan nilai peforma berupa *accuracy*, *precision*, *recall* dan *f1-score*.

g. Membandingkan Performa Algoritma
Perbandingan performa algoritma dilakukan dengan melihat nilai *accuracy*, *precision*, *recall* serta *f1-score* tertinggi dari evaluasi hasil klasifikasi menggunakan *confussion matrix,* sehingga didapat algoritma terbaik.

h. Menentukan Faktor Yang Mempengaruhi *Drop Out*
Faktor yang mempengaruhi mahasiswa *drop out* akan ditentukan bardasarkan implementasi algoritma dengan hasil performa paling baik.

i. Representasi Hasil
Representasi hasil penelitian mengenai klasifikasi mahasiswa berpotensi *drop out* dengan akurasi algoritma terbaik di implementasi menggunakan aplikasi sederhana berbasis web.

Objek Penelitian dan Variabel Penelitian

Objek dalam penelitian ini ialah data mahasiswa dari Sistem Informasi Akademik (SIAKAD) yang diperoleh dari LPTIK Universitas Jambi. Variabel penelitian yang digunakan yaitu data akademik mahasiswa S1 angkatan 2012, 2015 dan 2017 yang berada pada semester genap tahun 2018 yang terdiri dari data Indeks Prestasi Komulatif (IPK), Satuan Kredit Semester (SKS) Total, Asal

Sekolah, Pekerjaan Orang Tua, Pendapatan Orang Tua, Biaya Kuliah, Jenis Tempat Tinggal dan Status Mahasiswa.

Sumber Data

Data dalam penelitian ini merupakan data sekunder yang diperoleh dari Sistem Informasi Akademik (SIAKAD) Universitas Jambi angkatan 2012, 2015 dan 2017. Data yang digunakan didasarkan pada Peraturan Akademik Universitas Jambi yaitu Peraturan Rektor Nomor 02 Tahun 2017 mengenai evaluasi hasil belajar. Evaluasi dilakukan pada:

- IPK <2,00 pada semester 4 dengan minimal 40 SKS
- IPK <2,00 pada semester 8 dengan minimal 85 SKS, dan
- Lebih dari 14 semester.

Data yang diambil adalah data mahasiswa program S1 karena memiliki jumlah mahasiswa paling tinggi dibanding jenjang program yang lain.

Analisis Data

Pada penelitian ini akan dilakukan analisis pada performa atau kinerja dari algoritma *decision tree* c4.5 dan algoritma *naïve bayes* menggunakan metode *confussion matrix*. *Confussian matrix* akan menghasilkan nilai TP (*True Positive*) yaitu data yang terklasifikasi sebagai data positif yang sebenarnya memang benar data positif, TN (*True Negative*) yaitu data yang terklasifikasi sebagai data negative yang sebenarnya memang benar adalah data negative, FP (*False Positive*) yaitu data yang terklasifikasi sebagai data positif yang sebenarnya adalah data negative dan FN (*False Negative*) yaitu data yang terklasifikasi sebagai data negative yang sebenarnya adalah data positif. Dari hasil data TP, TN, FP dan FN akan dihasilkan peforma/kinerja algoritma berupa nilai *Accuracy*, *Precision*, *Recall* dan *F1-Score* yang akan menjadi ukuran dalam menentukan algoritma mana yang paling baik dalam mengklasifikasi mahasiswa berpotensi *dropout*.

III. HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang dikumpulkan merupakan data Sistem Informasi Akademik (SIAKAD) Universitas Jambi yang diperoleh dari Lembaga Pengembangan Teknologi Informasi dan Komunikasi (LPTIK) dengan jumlah data sebesar 8.706 data. Data yang diperoleh merupakan data mahasiswa program S1 angkatan 2012, 2015 dan 2017. Data dalam penelitian ini merupakan data awal sebelum dilakukan praproses data yang mana terdiri dari 7 atribut diantaranya id_mhs_pt, angkatan, ipk, sks_total, asal_sekolah, pekerjaan_ayah, penghasilan_ayah, ukt, dan jns_tinggal.

Praproses Data

1. *Data Selection*

Data selection atau seleksi data merupakan proses pemilihan data yang sesuai untuk proses klasifikasi, yang mana dalam proses ini akan dilakukan beberapa hal diantaranya:

Menentukan Atribut Target

Dalam penelitian ini akan ditentukan atribut target yaitu status mahasiswa yang dikelompokan kedalam dua *class* yaitu DO dan Tidak DO. Atribut status akan diperolah dari aturan evaluasi hasil belajar sesuai dengan peraturan rektor tahun 2017. Dalam peraturan tersebut disebutkan bahwa mahasiswa dengan ipk <2 pada semester 4 minimal 40 sks dan pada semester 8 minimal 85 sks serta mahasiswa yang berada lebih dari 14 semester berpotensi *drop out*. Data yang diperlukan untuk menentukan atribut status mahasiswa ialah angkatan, ipk dan sks total dengan menggunakan rumus pada excel. Contoh rumus yang digunakan yaitu:

$$= IF(C2 < 2, "DO", IF(B2 = 2012, "DO", IF(AND(B2 = 2017, D2 <= 40), "DO", IF(AND(B2 = 2015, D2 <= 85), "DO", "TIDAK DO"))))$$ **(1)**

Contoh data sebelum dan setelah ditentukan *class* targetnya dapat dilihat pada tabel berikut:

id_mhs_pt	angkatan	ipk	sks_total	asal_sekolah	pekerjaan_ayah	penghasilan_ayah	ukt	jns_tinggal
6279	2015	3.28	138	SMAN 8 MERANGIN	Petani/Nelayan	Kurang dari Rp.500.000	NULL	Kost
7152	2015	2.41	77	SMAN 3 KERINCI	Petani/Nelayan/Peternak	Tidak ada penghasilan	Rp. 1.000.000 - Rp. 1.999.999	Bersama ortu

Gambar 2. Contoh Data Sebelum Mendapat *Class* Target

id_mhs_pt	angkatan	ipk	sks_total	asal_sekolah	pekerjaan_ayah	penghasilan_ayah	ukt	jns_tinggal	status
6279	2015	3.28	138	SMAN 8 MERANGIN	Petani/Nelayan	Kurang dari Rp.500.000	NULL	Kost	Tidak DO
7220	2015	1.74	61	SMAN 11 KOTA JAMBI	Buruh	Tidak ada penghasilan	Rp. 2.000.000 - Rp. 4.999.999	Kost	DO

Gambar 3. Contoh Data Setelah Mendapat *Class* Target

Pengelompokan Jenis Asal Sekolah

Pada atribut asal_sekolah data yang diperoleh berupa jenis dan nama sekolah dari masing-masing wilayah, seperti: SMAN 8 MERANGIN, SMAN 11 KOTA JAMBI, dan sebagainya. Dalam penelitian ini, data yang digunakan pada atribut asal_sekolah hanya merujuk pada jenis sekolah. Contohnya seperti pada SMAN 8 MERANGIN menjadi SMAN yang berarti sekolah menengah atas negeri. Pengelompokan jenis asal sekolah dilakukan dengan menggunakan rumus excel, sebagai berikut.

$$= IFERROR(LEFT(E2, FIND(" ", E2) - 1), "") \quad \textbf{(2)}$$

Mengkategorikan Atribut IPK, SKS Total dan UKT

Pada atribut ipk, sks_total dan ukt, data yang dimiliki terdiri dari banyak variasi data sehingga perlu dikelompokan kedalam beberapa kategori. Berikut kategori data pada atribut ipk, sks_total dan ukt.

a. IPK

Pada atribut ipk, data akan dibagi menjadi 5 (lima) kategori diantaranya sebagai berikut.

Tabel 1. Kategori Pada Atribut ipk

IPK	Kategori
0.00 – 1.99	Tidak Memuaskan
2.00 – 2.75	Cukup Memuaskan
2.76 – 3.24	Memuaskan
3.25 – 3.79	Sangat Memuaskan
3.80 – 4.00	Dengan Pujian

Sumber: Pengolahan data, 2020

b. SKS Total

Data sks total yang digunakan dalam penelitian ini merupakan jumlah sks pada semester genap tahun 2018 yang mana terdiri dari sks total pada semester 4 untuk angkatan 2017, semester 8 untuk angkatan 2015 dan semester 14 untuk angkatan 2012. Untuk atribut sks total akan dibagi menjadi 2 (dua) kategori diantaranya.

Tabel 2. Kategori Pada Atribut sks_total

Semester	SKS Total	Kategori
4 (empat)	0-40	Tidak Memenuhi
	>40	Memenuhi
8 (delapan)	0-85	Tidak Memenuhi
	>85	Memenuhi
14 (empat belas)	<144	Tidak Memenuhi
	144	Memenuhi

Sumber: Pengolahan data, 2020

c. UKT

Pada atribut ukt terdapat data bernilai NULL yang mana dalam penelitian ini data NULL merupakan mahasiswa yang mendapatkan beasiswa yang berarti uktnya kosong atau biayanya sebesar Rp. 0 sehingga data NULL akan diubah menjadi 0 (biaya ukt Rp. 0). Untuk kategori pada atribut ukt dibagi menjadi 6 (enam) yaitu.

Tabel 3. Kategori Pada Atribut ukt

UKT	Kategori
0	Beasiswa
<1000000	Kurang dari Rp. 1.000.000
1000000 – 1999999	Rp. 1.000.000 - Rp. 1.999.999
2000000 – 4999999	Rp. 2.000.000 - Rp. 4.999.999
5000000 – 10000000	Rp. 5.000.000 - Rp. 10.000.000
>10000000	Lebih dari Rp. 10.000.000

Sumber: Pengolahan data, 2020

Menghapus Atribut id_mhs_pt dan angkatan

Pada penelitian ini, atribut id_mhs_pt dan atribut angkatan tidak digunakan dalam proses klasifikasi sehingga atribut tersebut akan dihapus. Atribut id_mhs_pt dan angkatan akan dihapus menggunakan *syntax python*.

Jadi, atribut yang digunakan untuk proses klasifikasi terdiri dari 8 atribut termasuk atribut target, diantaranya ipk, sks_total, asal_sekolah, pekerjaan_ayah, penghasilan_ayah, ukt, jns_tinggal dan status.

2. *Data Cleaning*

Data cleaning atau pembersihan data pada penelitian ini dilakukan dengan menghapus data-data *missing value*. Dari data yang

diperoleh terdapat data *missing value* pada beberapa atribut penelitian, diantaranya sebagai berikut:

Tabel 4. Jumlah Data *Missing Value*

Atribut	Jumlah Data *Missing Value*
ipk	0
sks_total	0
asal_sekolah	1654
pekerjaan_ayah	481
penghasilan_ayah	1059
ukt	0
jns_tinggal	0
status	0

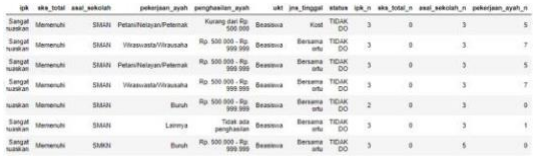
Sumber: Pengolahan data, 2020

Dari 8 atribut yang digunakan, 3 diantaranya memiliki data *missing value* yaitu atribut asal_sekolah, atribut pekerjaan_ayah dan atribut penghasilan_ayah. Untuk menghapus data *missing value* digunakan *syntax python* yang mana *syntax* tersebut akan menghapus setiap baris data *missing value* pada setiap atribut. Data bersih yang diperoleh setelah menghilangkan data *missing value* yaitu berjumlah 5.978 data.

data.

3. *Data Transformation*

Data dalam penelitian ini terdiri dari data *categorical* dan *numerical* sehingga perlu diubah ke dalam bentuk data yang sama. Data akan diubah menggunakan fitur *LabelEncoder* pada modul *sklearn python*. Data yang akan diubah dimasukan ke dalam kolom baru yaitu kolom ipk_n, sks_total_n, asal_sekolah_n, pekerjaan_ayah_n, penghasilan_ayah_n, ukt_n, jns_tinggal_n dan status_n.



Gambar 4. Data Yang Telah di Transformasi dan Dimasukan Dalam Kolom Baru

Setelah data diubah, kolom pada data awal (belum diberi label) tidak diperlukan lagi sehingga kolom tersebut harus dihapus.

Pembagian Data

Setelah dilakukan praproses data, selanjutnya akan dilakukan pembagian data. Data akan dibagi menjadi *data training* dan *data testing* dengan proporsi 80% : 20%. Berikut jumlah data setelah dilakukan pembagian.

Tabel 5. Pembagian Data

Klasifikasi	Jumlah Data	*Data Training* (80%)	*Data Testing* (20%)
DO	460	370	90
Tidak DO	5.518	4.412	1.106
Total	**5.978**	**4.782**	**1.196**

Sumber: Pengolahan data, 2020

Klasifikasi Menggunakan Algoritma *Decision Tree C4.5* dan *Naïve Bayes*

Setelah data dibagi menjadi *data training* dan *data testing*, selanjutnya ialah proses klasifikasi data. Pada tahap klasifikasi data, akan dilakukan pembangunan model menggunakan algoritma *decision tree* c4.5 dan algoritma *naive bayes*. Hasil akurasi klasifikasi yang diperoleh dari algoritma *decision tree* c4.5 dan algoritma *naive bayes* sebagai berikut.

Tabel 6. Akurasi

Algoritma	Akurasi (%)
Decision Tree C4.5	96,82%
Naive Bayes	96,24%

Sumber: Pengolahan data, 2020

Pengujian Menggunakan *Cross Validation*

Setelah didapat hasil akurasi dari proses klasifikasi, selanjutnya ialah melakukan pengujian menggunakan *cross validation*. Jenis pengujian *cross validation* yang digunakan ialah *k-fold cross validation* dimana k=10 yang berarti data akan dibagi menjadi 10 bagian dengan 9 bagian sebagai *data training* dan 1 bagian sebagai *data testing* (10-*fold cross validation*). *K-fold cross validation* adalah salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan melakukan perulangan dengan mengacak atribut masukan sehingga sistem teruji untuk beberapa atribut input yang acak (Liklikwatil, Noersasongko & Supriyanto, 2018).

1. Pengujian Algoritma *Decision Tree C4.5*

Tabel 7. Pengujian *10-Fold Cross Validation* Algoritma *Decision Tree* C4.5

K-Fold	Akurasi
1	97,5 %
2	97,5 %
3	97,5 %

 ISSN 2614-8277

4	95,83 %
5	97,5 %
6	95 %
7	99,15 %
8	95,79 %
9	95,79%
10	95,79 %
Rata-Rata	**96,74%**

Sumber: Pengolahan data, 2020

Berdasarkan hasil dari pengujian diatas dapat dilihat bahwa algoritma *decision tree* c4.5 menghasilkan rata-rata akurasi dari setiap *fold* sebesar 96,74%.

2. Pengujian Algoritma *Naive Bayes*

Tabel 8. Pengujian *10-Fold Cross Validation* Algoritma *Naive Bayes*

K-Fold	Akurasi
1	95,83 %
2	97,5 %
3	98,33 %
4	95,83 %
5	96,67 %
6	95,83 %
7	96,64 %
8	94,96 %
9	94,96 %
10	95,79 %
Rata-Rata	**96,24%**

Sumber: Pengolahan data, 2020

Berdasarkan hasil dari pengujian diatas dapat dilihat bahwa algoritma *naïve bayes* menghasilkan rata-rata akurasi dari setiap *fold* sebesar 96,24%.

Evaluasi Hasil Klasifikasi

Tahap selanjutnya dalam penelitian ini ialah evaluasi hasil klasifikasi pada data yang telah dilakukan pengujian menggunakan k-*fold cross validation*. Model evaluasi yang digunakan ialah *confussion matrix* yang menghasilkan nilai TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*). Selain itu, *confussion matrix* juga menghasilkan ukuran peforma dari klasifikasi berupa nilai akurasi, *precision* dan *recall*

1. Evaluasi Hasil Klasifikasi Algoritma *Decision Tree C4.5*

Hasil evaluasi klasifikasi menggunakan algoritma *decision tree c4.5* dapat dilihat pada tabel berikut.

Tabel 9. *Confussion Matrix* Pada Algoritma *Decision Tree* C4.5

Data Sebenarnya	Klasifikasi	
	DO	**Tidak DO**
DO	58 (TP)	32 (FN)
Tidak DO	7 (FP)	1.099 (TN)

Dari hasil *confussion matrix* yang telah diperoleh, dihasilkan nilai akurasi, *precision,recall* dan *f1-score* dengan hasil sebagai berikut.

```
from sklearn.metrics import classification_report
print (classification_report(y_test, y_hsl, digits=4))
```

```
              precision    recall  f1-score   support

           0     0.8923    0.6444    0.7484        90
           1     0.9717    0.9937    0.9826      1106

    accuracy                         0.9674      1196
   macro avg     0.9320    0.8191    0.8655      1196
weighted avg     0.9657    0.9674    0.9649      1196
```

Gambar 5. Hasil Pengukuran Kinerja Algoritma *Decision Tree C4.5*

Berdasarkan gambar di atas hasil akurasi yang didapat sebesar 96,74% dengan nilai *precision* untuk data DO dan Tidak DO masing-masing sebesar 89,23% dan 97,17%.. Nilai *recall* dari data DO dan Tidak DO yaitu 64,44% dan 99,37%. Dan nilai *f1-score* dari data DO dan Tidak DO yaitu 74,84% dan 98,26%. Untuk nilai *precision, recall* dan *f1-score* secara keseluruhan dapat dilihat pada nilai rata-rata macro (*macro average*) dari *precision, recall* dan *f1-score.* Nilai rata-rata yang dihasilkan yaitu *precision* sebesar 93,20%, *recall* sebesar 81,91% dan *f1-score*
sebesar 86,55%

2. Evaluasi Hasil Klasifikasi Algoritma *Naive Bayes*

Evaluasi hasil klasifikasi pada algoritma *naïve bayes* juga menggunakan metode *confussion matrix*, hasil evaluasi dapat dilihat pada tabel berikut.

Tabel 10. *Confussion Matrix* Pada Algoritma *Naive Bayes*

Data Sebenarnya	Klasifikasi	
	DO	**Tidak DO**
DO	45 (TP)	45 (FN)
Tidak DO	0 (FP)	1.106 (TN)

Sumber: Pengolahan data, 2020

Pengukuran kinerja algoritma *naive bayes* dapat dilihat pada nilai akurasi, *precision, recall,* dan *f1-score* yang dihasilkan

 ISSN 2614-8277

berdasarkan data dari *confussion matrix*.

```
from sklearn.metrics import classification_report
print (classification_report(y_test, y_pred, digits=4))
```

	precision	recall	f1-score	support
0	1.0000	0.5000	0.6667	90
1	0.9609	1.0000	0.9801	1106
accuracy			0.9624	1196
macro avg	0.9805	0.7500	0.8234	1196
weighted avg	0.9638	0.9624	0.9565	1196

Gambar 6. Hasil Pengukuran Kinerja Algoritma *Naive Bayes*

Hasil akurasi algoritma *naïve bayes* yang didapat sebesar 96,24%. Nilai *precision* untuk data DO dan Tidak DO masing-masing sebesar 100% dan 96,09%. Nilai *recall* dari data DO dan Tidak DO yang dihasilkan yaitu 50% dan 100%. Selanjutnya untuk nilai *f1-score* dari data DO dan Tidak DO yang didapat yaitu 66,67% dan 98,01%. Nilai rata-rata (*macro average*) dari *precision*, *recall* dan *f1-score* masing-masing sebesar 98,05%, 75%, dan 82,34%.

Perbandingan Algoritma *Decision Tree* C4.5 dan *Naive Bayes*

Dari hasil evaluasi klasifikasi, dihasilkan peforma/kinerja dari algoritma *decision tree* c4.5 dan *naive bayes* berupa rata-rata nilai *accuracy*, *precision, recall* dan *f1-score*. Berikut masing-masing hasil pengukuran kinerja dari tiap algoritma.

Tabel 11. Hasil Pengukuran Peforma Klasifikasi Algoritma *Decision Tree C4.5* dan *Naive Bayes*

Algoritma Klasifikasi	Akurasi (%)	*Precision* (%)	*Recall* (%)	*F1-Score* (%)
Decision Tree C4.5	96,74%	93,20%	81,91 %	86,55 %
Naive Bayes	96,24%	98,05%	75%	82,34 %

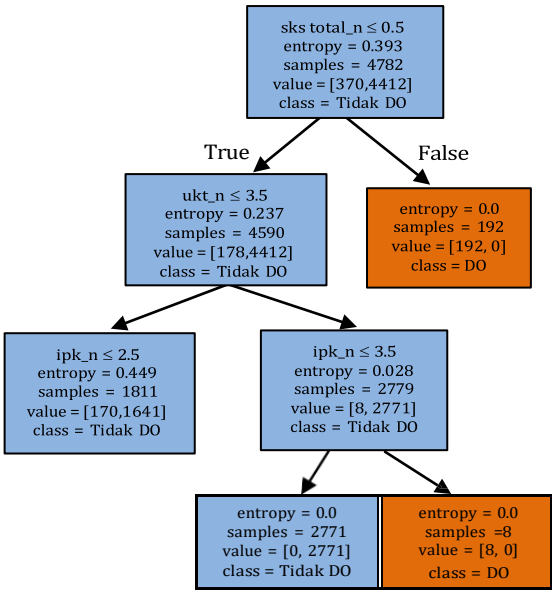
Sumber: Pengolahan data, 2020

Hasil pengukuran kinerja dari algoritma diatas menghasilkan nilai akurasi sebesar 96,74% *precision* sebesar 93,20%, *recall* sebesar 81,91% dan *f1-score* sebesar 86,55% pada algoritma *decision tree* c4.5. Sedangkan pada algoritma *naïve bayes* nilai akurasi yang dihasilkan yaitu sebesar 96,24%, *precision* sebesar 98,05%, *recall* sebesar 75%, dan *f1-score* sebesar 82,34%. Dari hasil tersebut dapat dilihat bahwa peforma dari algoritma *decision tree* c4.5 memiliki hasil yang lebih baik dibandingkan dengan algoritma *naïve bayes*. Sehingga algoritma dengan klasifikasi paling baik adalah algoritma *decision tree c4.5*.

Implementasi Hasil Berdasarkan Algoritma Terbaik

1. Visualisasi *Decision Tree* (Pohon Keputusan)

Dalam klasifikasi menggunakan algoritma *decision tree* c4.5 dihasilkan pohon keputusan yang mana atribut sks (satuan kredit semester) total adalah akar dari pohon keputusan tersebut. Itu artinya sks total paling berpengaruh dalam klasifikasi mahasiswa berpotensi *drop out* dilanjutkan dengan ukt dan ipk. Contoh dari hasil pohon keputusan dapat dilihat pada gambar berikut.



Gambar 7. Pohon Keputusan

Faktor yang paling berpengaruh dalam klasifikasi mahasiswa *drop out* juga dapat ditentukan dengan menggunakan perhitungan nilai *gain*. Berikut ini nilai gain dari masing-masing atribut:

Ipk	ipk_n	DO	Tidak Do	Total	*Entropy*	*Gain*
Cukup memuaskan	0	18	56	74	0.800392	0.076367
Dengan pujian	1	0	37	37	0	
Memuaskan	2	27	417	444	0.330655	
Sangat memuaskan	3	8	620	628	0.098447	
Tidak memuaskan	4	12	1	13	0.391244	
sks_total	**sks_total_n**	**DO**	**Tidak DO**	**Total**	***Entropy***	***Gain***
Tidak memenuhi	1	45	0	45	0	0.182901
Memenuhi	0	20	1131	1151	0.126444	
asal_sekolah	**asal_sekolah_n**	**DO**	**Tidak DO**	**Total**	***Entropy***	***Gain***

 ISSN 2614-8277

LAINNYA	0	4	5	9	0.991076	0.009324
MAN	1	2	72	74	0.179256	
MAS	2	2	16	18	0.503258	
SMAN	3	47	856	903	0.295037	
SMAS	4	5	101	106	0.274251	
SMKN	5	5	69	74	0.35678	
SMKS	6	0	12	12	0	

pekerjaan_ayah	**pekerjaan_ayah_n**	**DO**	**Tidak DO**	**Total**	***Entropy***	***Gain***
Buruh	0	2	156	158	0.097941	0.023619
Lainnya	1	0	81	81	0	
PNS/TNI/Polri	2	6	146	152	0.239874	
Pedagang	3	0	19	19	0	
Pegawai/Karyawan	4	18	145	163	0.50121	
Petani/Nelayan /Peternak	5	13	345	358	0.225123	
Tidak Bekerja	6	4	59	63	0.341154	
Wiraswasta/ Wirausaha	7	22	180	202	0.496622	

penghasilan_ayah	**penghasilan_ayah_n**	**DO**	**Tidak DO**	**Total**	***Entropy***	***Gain***
Dibawah Rp. 500.000	0	1	37	38	0.175565	0.040932
Lebih dari Rp. 20.000.000	1	1	1	2	1	
Rp. 1.000.000 – Rp. 1.999.999	2	9	282	291	0.199024	
Rp. 2.000.000 – Rp. 4.999.999	3	10	227	237	0.252263	
Rp. 5.000.000 – Rp. 20.000.000	4	0	52	52	0	
Rp. 500.000 – Rp. 999.999	5	8	379	387	0.145196	
Tidak ada penghasilan	6	36	153	189	0.702467	

ukt	**ukt_n**	**DO**	**Tidak DO**	**Total**	***Entropy***	***Gain***
Beasiswa	0	4	269	273	0.110254	0.091845
Kurang dari Rp. 1.000.000	1	28	15	43	0.933025	
Lebih dari Rp. 10.000.000	2	0	18	18	0	
Rp. 1.000.000 – Rp. 1.999.999	3	20	166	186	0.492411	
Rp. 2.000.000 – Rp. 4.999.999	4	11	543	554	0.140629	
Rp. 5.000.000 – Rp. 10.000.000	5	2	120	122	0.120681	

jns_tinggal	**jns_tinggal_n**	**DO**	**Tidak DO**	**Total**	***Entropy***	***Gain***
Bersama ortu	0	50	856	906	0.308038	0.00032
Kost	1	12	236	248	0.279506	
Lainnya	2	1	17	18	0.309543	
Wali	3	2	22	24	0.413817	

Pada tabel diatas, dapat dilihat bahwa atribut sks_total memiliki nilai *gain* paling tinggi yaitu 0.182901 sehingga atribut sks total merupakan faktor paling berpengaruh terhadap klasifikasi mahasiswa berpotensi *drop out*.

2. *Decision Rules* (Aturan - Aturan Keputusan)

Dari pohon keputusan yang telah dihasilkan, terbentuk 17 *decision rules* atau aturan-aturan keputusan. Contoh *decision rule* yang terbentuk diantaranya sebagai berikut.

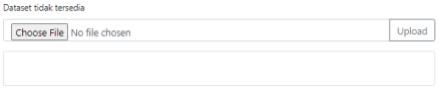
(R1) IF (sks total = Tidak memenuhi) THEN status = DO

(R2) IF (sks total = Memenuhi) AND ukt ≠ Rp.1.000.000 – Rp. 1.999.999 AND ipk ≠ Sangat memuaskan THEN status = DO

(R3) IF sks total = Memenuhi AND ukt ≠ Rp.1.000.000 – Rp. 1.999.999 AND ipk = Sangat memuaskan THEN status = Tidak DO

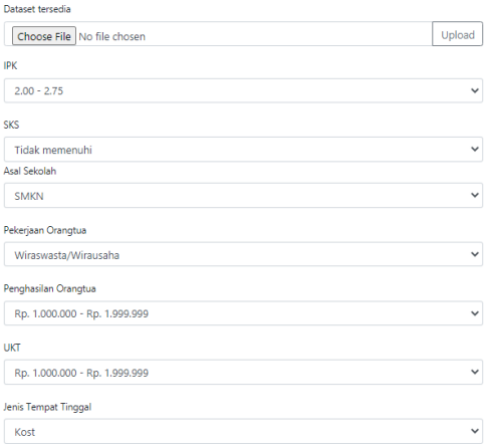
3. Aplikasi Sederhana

Berikut merupakan tampilan atau *interface* dari aplikasi klasifikasi mahasiswa berpotensi *drop out.* Aplikasi tersebut merupakan aplikasi sederhana berbasis web yang digunakan sebagai contoh dalam melakukan klasifikasi mahasiswa yang berpotensi *drop out* dengan menggunakan beberapa variabel/parameter yang telah ditentukan. Dalam aplikasi ini algoritma yang digunakan merupakan algoritma *decision tree c4.5* yang mana dalam hasil penelitian merupakan algoritma yang memiliki ketepatan yang sangat baik.

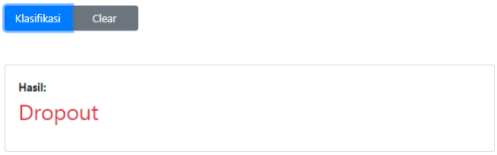


Gambar 8. Tampilan Awal Sebelum Mengupload Dataset

 ISSN 2614-8277



Gambar 9. Tampilan Proses Klasifikasi



Gambar 10. Tampilan Hasil Klasifikasi

Pembahasan

Hasil penelitian menunjukkan bahwa pada proses klasifikasi yang telah dilakukan pengujian menggunakan *10-fold cross validation*, *decision tree c4.5* memiliki akurasi yang lebih tinggi yaitu sebesar 96,74% dibandingkan dengan algoritma *naive bayes* yang memiliki akurasi sebesar 96,24%. Sehingga terbukti bahwa algoritma *decision tree c4.5* merupakan algoritma yang paling baik dalam melakukan klasifikasi data. Hal ini sejalan dengan penelitian Anam & Santoso (2018) yang menyatakan bahwa algoritma *decision tree* merupakan algoritma paling baik dengan akurasi yang lebih tinggi dibandingkan dengan algoritma *naive bayes*.

Peforma dari algoritma *decision tree c4.5* dan algoritma *naïve bayes* juga diukur dengan menggunakan nilai *precision*, *recall* dan *f1-score*. Pada algoritma *decision tree* nilai rata-rata *precision* yang didapat sebesar 93,20% sedangkan pada algoritma *naive bayes* nilai rata-rata *precision* lebih tinggi yaitu 98,05%. Untuk nilai rata-rata *recall* algoritma *decision tree* memiliki hasil yang lebih tinggi yaitu sebesar 81,91% dan algoritma *naïve bayes* sebesar 75%. Dan untuk nilai rata-rata *f1-score* pada algoritma *decision tree c4.5* memberikan hasil lebih tinggi yaitu sebesar 86,55% sedangkan *naïve bayes* sebesar 82,34%. Menurut Abdurrohman, Dini dan Muharram (2019) algoritma atau model terbaik adalah model dengan *f1-score* terbaik, karena hal tersebut menentukan nilai *precision* dan *recall* dari model yang bersangkutan. Sehingga algoritma paling baik dalam melakukan klasifikasi ialah algoritma *Decision Tree* C4.5.

Berdasarkan hasil pohon keputusan yang terbentuk dari algoritm *decision tree c4.5* yang merupakan algoritma dengan peforma paling baik, didapatkan hasil bahwa sks total merupakan *root* dari pohon keputusan yang berarti sks total merupakan faktor yang paling berpengaruh dalam klasifikasi mahasiswa berpotensi *drop out* dengan menghasilkan 17 *rule* (aturan keputusan).

IV. KESIMPULAN

Dari penelitian yang telah dilakukan didapat kesimpulan sebagai berikut:

1. Perbandingan algoritma *Decision Tree* C4.5 dan *Naive Bayes* menghasilkan nilai akurasi masing-masing sebesar 96,74% dan 96,24%. Dengan nilai rata-rata *precision* dan *recall* yaitu 93,20% dan 81,91% untuk algoritma *Decision Tree* C4.5 serta 98,05% dan 75% untuk algoritma *Naïve Bayes*. Nilai rata-rata *f1-score algoritma decision tree c4.5* sebesar 86,55% dan *naive bayes* sebesar 82,34%. Dari hasil yang didapat, algoritma *Decision Tree* C4.5 memiliki hasil peforma lebih tinggi dari algoritma *Naive Bayes* sehingga algoritma *Decision Tree* C4.5 merupakan algoritma paling baik dalam mengklasifikasi mahasiswa berpotensi *drop out* di Universitas Jambi.
2. Berdasarkan algoritma *Decision Tree* C4.5 yang merupakan algoritma paling baik dalam mengklasifikasi mahasiswa berpotensi *drop out*, diimplementasi sebuah pohon keputusan yang menghasilkan 17 *rule* (aturan) dengan atribut sks total sebagai akarnya. Itu artinya sks (satuan kredit semester) total merupakan faktor yang paling berpengaruh dalam klasifikasi mahasiswa berpotensi *drop out*. Hal ini juga dapat dilihat pada nilai *information gain* yang dihasilkan pada masing-masing atribut yang mana nilai *gain* tertinggi terdapat pada atribut sks total sehingga terbukti bahwa atribut sks total merupakan faktor paling berpengaruh dalam klasifikasi mahasiswa berpotensi *drop out*.

V. REFERENSI

Abdurrohman, H., Dini, R., & Muharram, A.

P. (2019). *Evaluasi Performa Metode Deep Learning Untuk Klasifikasi Citra Lesi Kulit the Ham10000*. *January*, 63–68. https://doi.org/10.5614/sniko.2018.10

Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal Ilmiah Ilmu-Ilmu Teknik*, *8*(1), 13–19. https://ejournal.upm.ac.id/index.php/energy/article/view/111/449

Atma, Y. D., & Setyanto, A. (2018). Perbandingan algoritma c4.5 dan k-nn dalam identifikasi mahasiswa berpotensi drop out. *Metik Jurnal*, *2*(2), 31–37.

Lembaga Pengembangan Teknologi Informasi dan Komunikasi Universitas Jambi. 2019. Statistik Universitas Jambi. (https://dss.unja.ac.id, diakses 22 September 2019)

Liklikwatil, R. D., Noersasongko, E., & Supriyanto, C. (2018). Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet. *E-Jurnal JUSITI (Jurnal Sistem Informasi Dan Teknologi Informasi)*, *7–2*(2), 172–182. https://doi.org/10.36774/jusiti.v7i2.252

Nasrullah, A. H. (2018). Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, *10*(2), 244–250. https://doi.org/10.33096/ilkom.v10i2.300.244-250

Nurhayati, S., Kusrini, & Luthfi, E. T. (2015). Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector. *Prediksi Menggunakan SVM*, *3*(6), 82–93.

Peraturan Akademik Universitas Jambi. 2018. Peraturan Rektor Nomor 02 Tahun 2017 (https://www.unja.ac.id/wp-content/uploads/2018/05/PERATURAN-REKTOR-NO-2-TH-2017.pdf, diakses 31 Oktober 2019).

Putra, A. (2017). Solusi Prediksi Mahasiswa Drop Out Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma. *SIMETRIS*, *8*(1), 177–184.

Suyadi, S., Setyanto, A., & Fattah, H. Al. (2017). Analisis Perbandingan Algoritma Decision Tree (C4.5) Dan K-Naive Bayes Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas. *Indonesian Journal of Applied Informatics*, *2*(1), 59. https://doi.org/10.20961/ijai.v2i1.13258

Wahyuni, S., Saragih, kana S., & Perangin-angin, M. I. (2018). Implemntasi Metode Decision Tree C4.5 Untuk Menganalisa Mahasiswa Dop Out. *Ethos*, *6*(1), 42–51.

Wijaya, inggit hasta. (2018). Prediksi Mahasiswa Drop Out Berdasarkan Klasifikasi Administratif. *Simki-Techsain*, *02*(01), 1–8.

Yahya, N., & Jananto, A. (2019). Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang). *Prosiding SENDI*, *2014*, 978–979. https://www.unisbank.ac.id/ojs/index.php/sendi_u/article/view/7389/2369