Jurnal Ilmiah Ilmu Komputer Vol. 8, No. 2, September 2022 (P) ISSN 2442-451X
Fakultas Ilmu Komputer (O) ISSN 2503-3832
Universitas AL Asyariah Mandar

KLASIFIKASI MAHASISWA BERPOTENSI *DROP OUT* MENGGUNAKAN *ALGORITMA NAIVE BAYES* DAN *DECISION TREE*

Salmawati[1], Yuyun[2], Hazriani[3]

[1,2]Magister Sistem Komputer, Universitas Handayani Makassar, Makassar, Indonesia
[3]Magister Sistem Komputer, Universitas Handayani Makassar, Makassar, Indonesia
[1]salmazhally@gmail.com , [2]yuyunwabula@handayani.ac.id , [3]hazriani@gmail.com

ABSTRAK

Seiring perkembangan dunia pendidikan Indonesia, Perguruan Tinggi Negeri maupun Perguruan Tinggi Swasta bersaing begitu ketat dalam memberikan performanya dalam mencetak lulusan-lulusan berkualitas. Salah satu indikator kegagalan mahasiswa adalah kasus Drop Out. Dengan menggunakan data mining algoritma naive bayes dan decision tree dapat dilakukan klasifikasi mahasiswa berpotensi drop out. Parameter yang digunakan yaitu, jenis kelamin, IPK, SKS, penghasilan orang tua, jenis tempat tinggal dan jenis transformasi. Mahasiswa yang telah lulus/drop out angkatan 2010-2014 sebanyak 1155 data dijadikan sebagai data training dan data testing. Perbandingan algoritma vaive bayes dan decision tree menghasilkan masing-masing nilai akurasi 97,83% dan 99,13%. Dengan nilai rata-rata precision dan recall yaitu 61,54% dan 100% untuk algoritma naive bayes serta 84,61% dan 100% untuk algoritma decision tree. Nilai rata rata f1-score algoritma naive bayes sebesar 76,19% dan decision tree sebesar 91,67%. Algoritma decision tree memiliki hasil peforma lebih tinggi dari algoritma Naive Bayes. Sehingga algoritma decision tree merupakan algoritma paling baik. Decision tree menghasilkan 11 rule dengan atribut IPK sebagai akarnya. Itu artinya IPK merupakan faktor paling berpengaruh dalam mengklasifikasi mahasiswa berpotensi drop out.

Kata Kunci: Klasifikasi, *Drop Out, Naive Bayes, Decision Tree*

ABSTRACT

Along with the development of the world of education in Indonesia, State Universities and Private Universities are competing so tightly in providing their performance in producing quality graduates. One indicator of student failure is the Drop Out case. By using data mining naive Bayes algorithm and decision tree, it is possible to classify students who have the potential to drop out. The parameters used are, gender, IPK, SKS, parents' income, type of residence and type of transformation. 1155 students who have graduated/dropped out of class 2010-2014 are used as training data and testing data. Comparison of vaive bayes algorithm and decision tree yielded accuracy values of 97.83% and 99.13%, respectively. With the average value of precision and recall are 61.54% and 100% for the naive Bayes algorithm and 84.61% and 100% for the decision tree algorithm. The average f1-score value of the Naive Bayes algorithm is 76.19% and the decision tree is 91.67%. The decision tree algorithm has higher performance results than the Naive Bayes algorithm. So the decision tree algorithm is the best algorithm. The decision tree produces 11 rules with the IPK attribute as the root. That means that IPK is the most influential factor in classifying students who have the potential to drop out.

Keywords: Classification, Drop Out, Naive Bayes, Decision Tree

(P) ISSN 2442-451X
(O) ISSN 2503-3832

1. PENDAHULUAN (Times New Roman 10 Bold)

Seiring perkembangan dunia pendidikan Indonesia, Perguruan Tinggi Negeri (PTN) maupun Perguruan Tinggi Swasta (PTS) bersaing begitu ketat dalam memberikan performanya dalam mencetak lulusan-lulusan berkualitas. Maka, pihak perguruan tinggi berupaya dalam meningkatkan kualitas serta memberikan pendidikan terbaik bagi penerima jasanya yaitu mahasiswa [1].

Salah satu masalah apabila ada beberapa mahasiswa yang terlambat lulus atau tidak tepat pada waktunya sehingga menjadi kendala untuk kemajuan dari perguruan tinggi tersebut. Apabila suatu sistem dapat memperkirakan atau memprediksi mahasiswa lulus tepat waktu akan sangat mempermudah bagi pihak kampus dalam mengambil langkah-langkah pencegahan agar tidak terjadi kasus Drop Out [2].

Kasus Drop Out juga bisa terjadi karena beberapa faktor seperti rendahnya kemampuan akademik mahasiswa, faktor biaya dan tempat tinggal saat menempuh pendidikan. Maka dari itu penelitian ini akan melakukan klasifikasi terhadap kasus mahasiswa Drop Out agar dapat diketahui faktor apa yang paling berpengaruh terhadap kasus tersebut sehingga pihak kampus dapat membuat kebijakan untuk mengurangi atau mencegah potensi mahasiswa yang terdeteksi Drop Out [3].

Penelitian tentang perbandingan algoritma Naive Bayes dengan Decision Tree pernah diteliti dalam mengklasifikasi penerimaan mahasisiwa baru yang memperoleh hasil bahwa algoritma Naive Bayes memiliki tingkat akurasi paling baik dibandingkan dengan algoritma Decision Tree [4]. Sedangkan penelitian untuk memprediksi lama studi mahasiswa dan memperoleh hasil bahwa metode Decision Tree memiliki persentase keakuratan yang lebih tinggi dibandingkan Naïve Bayes [5].

Beberapa penelitian tersebut terlihat bahwa algoritma Naive Bayes dan Decision Tree memiliki tingkat akurasi yang baik dalam melakukan klasifikasi. Meskipun kedua algoritma tersebut memiliki kemampuan yang serupa, tetapi pastinya kedua algoritma tersebut memiliki tingkat keakuratan yang berbeda. Maka dari itu perbandingan terhadap kedua algoritma tersebut dibutuhkan untuk memberikan informasi tentang performa yang paling baik sehingga menghasikan klasifikasi yang lebih akurat dalam hal memprediksi mahasiswa berpotensi Drop Out.

Berdasarkan latarbelakang yang telah diuraikan, maka penulis mengambil judul "Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Naive Bayes Dan Decision Tree" yang bertujuan untuk mengetahui algoritma terbaik berdasarkan perfoma tertinggi dalam memprediksi mahasiswa yang berpotensi Drop Out sehingga dapat dijadikan sebagai acuan dalam pengambilan keputusan guna mengurangi tingkat mahasiswa Drop Out.

2. TINJAUAN PUSTAKA (Times New Roman 10 Bold)

2.1. *Klasifikasi*

Suatu teknik dengan melihat atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikanklasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan popular adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk interpretasi seperti Algoritma C4.5, ID3 dan lain-lain [6].

2.2. Drop Out

Salah satu masalah yang dihadapi perguruan tinggi adalah terdapat mahasiswa yang Drop Out (putus kuliah). Drop Out termasuk masalah yang serius dan merupakan situasi yang pernah dihadapi oleh sebagian mahasiswa di perguruan tinggi [7] .

2.3. Algoritma Naive Bayes

Naive bayes merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [8]. Dalam metode ini perhitungan ditunjukkan melalui Persamaan (1) :

$$P(H|X) = \frac{P(X|H)}{P(X)} . P(H) \quad (1)$$

Keterangan
X : Data dengan class yang belum diketahui
H : Hipotesis data merupakan suatu class spesifik
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
P(H) : Probabilitas hipotesis H (prior probabilitas)
P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H
P(X) : Probabilitas X [9]

2.4. Algoritma Decicion Tree

Konsep Decision Tree atau Pohon keputusan adalah mengubah data menjadi aturan - aturan keputusan. Manfaat utama dari penggunaan decision tree adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi simple, sehingga pengambilan keputusan akan lebih memudahkan solusi dari sebuah permasalahan [10]. Decision Tree adalah salah satu metode klasifikasi yang paling populer karena mudah diinterprestasikan oleh manusia [11]. Decision Tree digunakan untuk pengenalan pola dan termasuk dalam pengenalan pola secara statistic. Decision Tree menggunakan 2 perhitungan yang pertama adalah

(P) ISSN 2442-451X
(O) ISSN 2503-3832

perhitungan Gain pada Persamaan (2) dan perhitungan Entropy pada Persamaan (3) [5].

Perhitungan Gain

$$\text{Gain}\ (S, A) = \text{Entropy}\ (S) - \sum_{i=1}^{n} * \text{Entropy}(Si) \quad (2)$$

Keterangan:
S : himpunan
A :atribut
n : jumlah partisi atribut A
|Si| : jumlah kasus pada partikel ke-i
|S| : jumlah kasusdalam S

Menghitung Nilai Entropy

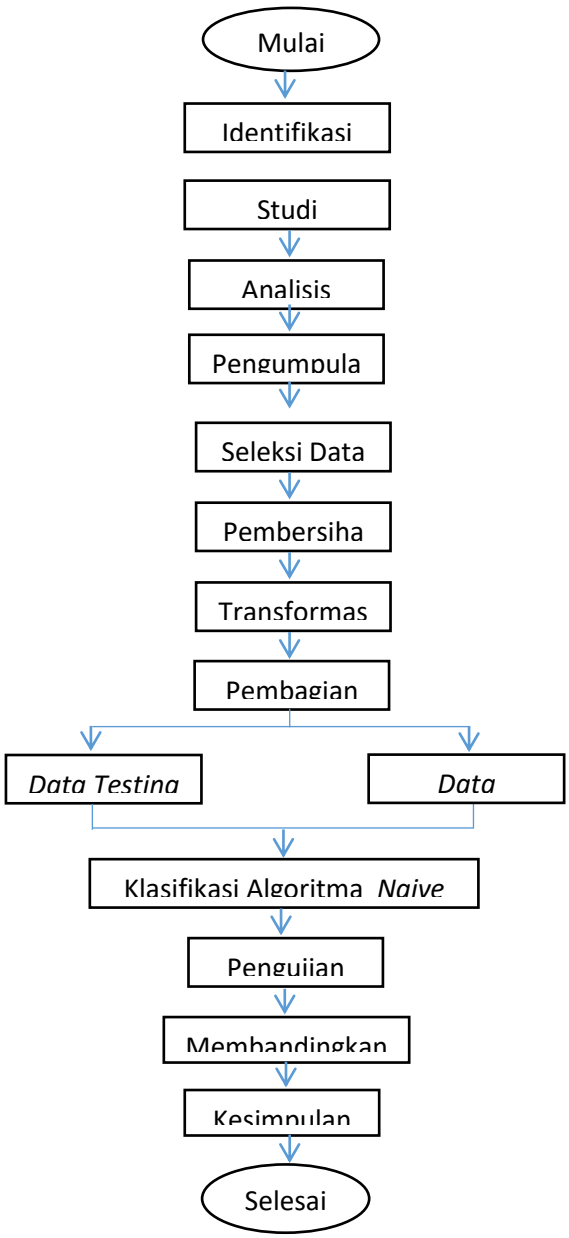
$$\text{Entropy}\ (S) = \sum_{i=1}^{n} - pi * log2\ pi \quad (3)$$

Keterangan :
S : himpunan kasus
A : fitur
n : jumlah partisi S
pi : Proporsi dari Si terhadap S

3. METODE YANG DIUSULKAN

Penelitian ini akan melalui beberapa tahap, dimana setiap tahap berdekatan dan saling mempengaruhi satu sama lain. Tahapan-tahapan penelitian yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 3.1. Tahap awal yang dilakukan adalah proses identifikasi masalah dengan melakukan wawancara secara langsung dengan ketua Program Studi Sistem Informasi dan Ketua Jurusan Program Studi Teknik Informatika. Selanjutnya pada tahap studi literatur dilakukan pencarian berbagai literatur yang nantinya dapat membantu dan mendukung dalam melaksanakan penelitian. Pada tahap analisis kebutuhan dilakukan pengidentifikasian untuk mengetahui apa saja kebutuhan fungsional dari sistem berdasarkan preferensi pengguna dalam kasus ini yakni ketua Program Studi Studi.

Pada proses pengumpulan data dilakukan dengan menggunakan metode dokumentasi dengan mengambil data akademik dan demografi mahasiswa di Universitas Al Asyariah mandar. Setelah data diperoleh, dilakukan pengolahan data yang terdiri dari tiga proses yaitu seleksi data, pembersihan data, dan yang terakhir transformasi data. Langkah selanjutnya yaitu pembagian data, pada tahap ini data akan dibagi menjadi dua yaitu Data Training yang merupakan data latih dan Data Testing yang merupakan data uji. Kemudian dilakukan pengimplementasian algoritma Naive Bayes dan Decision Tree untuk klasifikasi data. Selanjutnya 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. Setelah semua tahap selesai dilakukan, maka dapat ditarik kesimpulan dari penelitian tersebut.



Gambar 3.1 Tahapan Penelitian

4. HASIL PENELITIAN

Data yang dikumpulkan merupakan data dari Sistem Informasi Akademik (SIAKAD) Universitas Al Asyariah Mandar yang diperoleh dari bagian Akademik Sistem Informasi dan Kemahasiswaan dengan jumlah data sebesar 1.296 data. Data yang diperoleh merupakan data

(P) ISSN 2442-451X
(O) ISSN 2503-3832

mahasiswa Fakultas Agama Islam dan Fakultas Keguruan dan Ilmu Pendidikan angkatan tahun 2010 – 2014 yang berstatus lulus/DO. Data dalam penelitian ini merupakan data awal sebelum dilakukan praproses data.

4.1. Praproses Data

Praproses data merupakan proses mempersiapkan data sebelum dilakukannya proses klasifikasi. Praproses data dalam penelitian ini terdiri dari data selection, data cleaning, dan data transformation.

1. Data Selection

Data Selection atau seleksi data merupakan proses pemilihan data yang benar- benar diperlukan dan sesuai untuk proses klasifikasi. Dalam penelitian ini akan ditentukan atribut target yaitu status mahasiswa yang dikelompokan kedalam dua class yaitu DO dan Tidak DO. Data yang diperlukan untuk menentukan atribut status mahasiswa ialah jenis kelamin, IPK, SKS, penghasilan orang tua, jenis tempat tinggal dan jenis trasnportasi.

2. Data Cleaning

Data cleaning merupakan tahap preprocessing yang dilakukan untuk menghapus redudansi data, menghilangkan data mahasiswa yang mengundurkan diri, putus sekolah, wafat, mutasi, aktif dan lainnya, merubah status mahasiswa dikeluarkan menjadi DO dan status mahasiswa luus menjadi Tidak DO. Dari seluruh data yang diperoleh, terdapat 141 data yang di cleaning antara lain, mengundurkan diri sebanyak 75 data, putus sekolah 19 data, wafat 2 data, mutasi 30 data, aktif 9 odata dan lainnya 6 data. Sehingga total data setelah dilakukan Cleaning Data menjadi 1155 data.

3. Data Transformation

Data transformation merupakan proses mengubah data menjadi bentuk yang di perlukan. Dalam penelitian ini, peneliti mengubah data IPK, SKS, penghasilan orangtua, jenis tempat tinggal dan jenis transpotrasi.

4.2. Pembagian Data

Setelah dilakukan praproses data, selanjutnya akan dilakukan pembagian data. Pada tahap ini, data akan dibagi menjadi dua bagian yaitu data training dan data testing. Pembagian data training dan data testing ini berdasarkan atribut target yang telah memiliki class data. Data training merupakan data yang digunakan untuk melatih algoritma. Tujuannya agar algoritma dapat mempelajari pola dari data yang diberikan. Sedangkan data testing merupakan data yang digunakan untuk melihat performa dari algoritma yang telah dilatih. Dalam penelitian ini data akan di bagi dengan proporsi 80% data training dan 20% data testing. Berikut jumlah data setelah dilakukan pembagian.

Tabel 4.1 Pembagian Data

Klasifikasi	Jumlah Data	*Data Training* (80%)	*Data Testing* (20%)
DO	67	54	13
Tidak DO	1088	871	217
Total	**1155**	**925**	**230**

Dari Tabel 4.1 diketahui bahwa jumlah data training yang dihasilkan sebanyak 925 data dan jumlah data testing sebanyak 230 data.

4.3. Algoritma Naive Bayes

Perhitungan pada algoritma naive bayes dilakukan dengan beberapa tahap, mulai dari menentukkan variabel sampai dengan penarikan kesimpulan. Adapun tahapan-tahapan perhitungannya adalah sebagai berikut:

1. Menetukkan Data Training

Data Training digunakan sebagai data acuan dalam perhitungan ini. Data yang digunakan sebanyak 925 data mahasiswa Unasman.

Tabel 4.2 Sampel Data Training

No.	Jenis Kelamin	IPK	SKS	Penghasilan Orang Tua	Jenis Tempat Tinggal	Jenis Transportasi	Status Drop Out
1	Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO
2	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
3	Perempuan	Cukup	Memenuhi	3	1	0	Tidak DO
4	Perempuan	Baik	Memenuhi	2	1	1	Tidak DO
5	Laki-Laki	Baik	Memenuhi	3	1	1	Tidak DO
6	Perempuan	Baik	Memenuhi	2	0	2	Tidak DO
7	Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO
8	Perempuan	Baik	Memenuhi	3	0	2	Tidak DO
9	Perempuan	Cukup	Memenuhi	2	1	1	Tidak DO
10	Perempuan	Cukup	Memenuhi	3	0	2	Tidak DO
11	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
12	Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO
13	Perempuan	Baik	Memenuhi	3	0	0	Tidak DO
14	Perempuan	Cukup	Memenuhi	2	2	2	Tidak DO
15	Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO
16	Perempuan	Baik	Memenuhi	3	2	2	Tidak DO
17	Perempuan	Baik	Memenuhi	2	3	1	Tidak DO
18	Perempuan	Kurang	Memenuhi	2	0	2	DO
19	Perempuan	Baik	Memenuhi	2	0	0	Tidak DO
20	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
21	Perempuan	Cukup	Memenuhi	2	0	2	Tidak DO
22	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
23	Perempuan	Kurang	Tidak Memenuhi	2	2	2	DO
24	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
25	Perempuan	Baik	Memenuhi	3	1	0	Tidak DO
26	Laki-Laki	Cukup	Memenuhi	2	0	0	Tidak DO
27	Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO
28	Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO
29	Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO
30	Perempuan	Baik	Memenuhi	2	1	1	Tidak DO
31	Perempuan	Cukup	Memenuhi	3	1	0	Tidak DO
32	Perempuan	Baik	Memenuhi	3	0	0	Tidak DO
33	Perempuan	Cukup	Memenuhi	3	1	0	Tidak DO
34	Perempuan	Cukup	Memenuhi	2	2	2	Tidak DO
35	Perempuan	Baik	Memenuhi	3	0	0	Tidak DO
36	Laki-Laki	Baik	Memenuhi	3	0	0	Tidak DO
37	Laki-Laki	Cukup	Memenuhi	3	0	2	Tidak DO
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
925	Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO

2. Menentukkan Data Testing

Setelah proses training dilakukan pada sebuah algoritma machine learning, tahap selanjutnya adalah melakukan evaluasi terhadap performa algoritma tersebut atau biasa disebut testing. Pada proses testing, performa algoritma akan diuji menggunakan data testing, dimana data testing dan data training merupakan data yang berbeda. Data yang digunakan dalam data testing adalah sebanyak 230 data mahasiswa Unasman.

(P) ISSN 2442-451X
(O) ISSN 2503-3832

Tabel 4.3 Data Testing

No.	Jenis Kelamin	IPK	SKS	Penghasilan Orang Tua	Jenis tempat Tinggal	Jenis Transportasi	Status Drop Out
1	Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO
2	Perempuan	Kurang	Memenuhi	2	0	2	DO
3	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
4	Perempuan	Cukup	Memenuhi	3	1	0	Tidak DO
5	Laki-Laki	Kurang	Tidak Memenuhi	2	1	1	DO
6	Perempuan	Baik	Memenuhi	2	1	1	Tidak DO
7	Laki-Laki	Baik	Memenuhi	3	1	1	Tidak DO
8	Perempuan	Baik	Memenuhi	2	0	2	Tidak DO
9	Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO
10	Perempuan	Kurang	Memenuhi	2	0	2	DO
11	Perempuan	Baik	Memenuhi	3	0	2	Tidak DO
12	Perempuan	Cukup	Memenuhi	2	1	1	Tidak DO
13	Perempuan	Cukup	Memenuhi	3	0	2	Tidak DO
14	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
15	Perempuan	Kurang	Tidak Memenuhi	2	1	1	DO
16	Perempuan	Kurang	Tidak Memenuhi	2	0	2	DO
17	Laki-Laki	Cukup	Tidak Memenuhi	2	0	1	DO
18	Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO
19	Laki-Laki	Kurang	Tidak Memenuhi	3	2	2	DO
20	Perempuan	Baik	Memenuhi	3	0	0	Tidak DO
21	Perempuan	Cukup	Memenuhi	2	2	2	Tidak DO
22	Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO
23	Perempuan	Baik	Memenuhi	3	2	2	Tidak DO
24	Laki-Laki	Kurang	Memenuhi	3	0	0	DO
25	Laki-Laki	Kurang	Tidak Memenuhi	2	1	1	DO
26	Perempuan	Kurang	Memenuhi	2	2	2	DO
27	Perempuan	Kurang	Tidak Memenuhi	3	1	1	DO
28	Perempuan	Cukup	Memenuhi	3	1	1	DO
29	Perempuan	Kurang	Tidak Memenuhi	3	0	0	DO
30	Perempuan	Baik	Memenuhi	2	3	1	Tidak DO
31	Perempuan	Baik	Memenuhi	2	0	0	Tidak DO
32	Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
230	Perempuan	Baik	Memenuhi	2	1	1	Tidak DO

3. Mencari Kelas Prediksi

Setelah nilai probabilitas Tidak DO dan DO dihitung, tahap berikutnya adalah membandingan kedua nilai tersebut. Jika nilai probabilitas Tidak DO lebih besar dari probabilitas DO, maka data akan diklasifikasikan sebagai data Tidak DO. Sebaliknya jika probabilitas DO lebih besar dari probabilitas Tidak DO, maka data akan diklasifikasikan sebagai data DO.

Tabel 4.4. Kelas Prediksi

Jenis Kelamin	IPK	SKS	Penghasilan Orang Tua	Jenis tempat Tinggal	Jenis Transportasi	Status Drop Out	Class Prediction
Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO	Tidak DO
Perempuan	Kurang	Memenuhi	2	0	2	DO	Tidak DO
Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO	Tidak DO
Perempuan	Cukup	Memenuhi	3	1	0	Tidak DO	Tidak DO
Laki-Laki	Kurang	Tidak Memenuhi	2	1	1	DO	DO
Perempuan	Baik	Memenuhi	2	1	1	Tidak DO	Tidak DO
Laki-Laki	Baik	Memenuhi	3	1	1	Tidak DO	Tidak DO
Perempuan	Baik	Memenuhi	2	0	2	Tidak DO	Tidak DO
Perempuan	Cukup	Memenuhi	3	0	0	Tidak DO	Tidak DO
Perempuan	Kurang	Memenuhi	2	0	2	DO	Tidak DO
Perempuan	Baik	Memenuhi	3	0	2	Tidak DO	Tidak DO
Perempuan	Cukup	Memenuhi	2	1	1	Tidak DO	Tidak DO
Perempuan	Cukup	Memenuhi	3	0	2	Tidak DO	Tidak DO
Perempuan	Cukup	Memenuhi	3	2	0	Tidak DO	Tidak DO
Perempuan	Kurang	Tidak Memenuhi	2	1	1	DO	DO
Perempuan	Kurang	Tidak Memenuhi	2	0	2	DO	DO
Laki-Laki	Cukup	Tidak Memenuhi	2	0	1	DO	DO
Perempuan	Cukup	Memenuhi	2	0	0	Tidak DO	Tidak DO
Laki-Laki	Kurang	Tidak Memenuhi	3	2	2	DO	DO
Perempuan	Baik	Memenuhi	3	0	0	Tidak DO	Tidak DO
Perempuan	Cukup	Memenuhi	2	2	2	Tidak DO	Tidak DO

Berdasarkan data yang ada, didapatkan bahwa ada 5 data yang sebenarnya DO tapi hasil prediksinya Tidak DO. Hal ini terjadi dikarenakan penelitian ini menggunakan parameter IPK semester IV. Mahasiswa yang sebenarnya DO tetapi hasil prediksi tidak DO disebabkan karena pada awal semester sampai semester IV mahasiswa tersebut aktif mengikuti proses perkuliahan dan semester berikutnya mahasiswa tersebut tidak lagi aktif samapi batas masa studi selesai sehingga hasilnya mahasiswa tersebut di drop out.

4. Menghitung Counfusion Matriks

Setelah data dihitung, selanjutnya akan dilakukan pengujian data. Tahap pengujian dilakukan untuk mengecek kembali tentang keakuratan penghitungan data. Pengujian dilakukan dengan cara melakukan pembandingan hasil akhir prediksi dengan data yang sesungguhnya terjadi dilapangan. Adapun metode yang digunakan untuk melakukan pengujian adalah metode confusion matrix . Ada empat nilai yang dihasilkan di dalam tabel confusion matrix, di antaranya True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN).

Tabel 4.5 Counfusion Matriks

	Prediksi	
Aktual	DO	Tidak DO
DO	8	5
Tidak DO	0	217

Berdasarkan hasil pengujian diatas dapat disimpulkan bahwa hasil keakuratan sebesar 97,83%, sensitivity 100% dan presisi data terhadap pengujian sebesar 61,54%, specificity 97,75% dan F1-Score sebesar 76,19%. Tingkat akurasinya tidak 100% disebabkan ada beberapa data yang setelah diprediksi hasilnya tidak sama dengan hasil yang sesungguhnya.

4.4. Algoritma Decision Tree

Langkah pertama adalah memilih atribut yang akan dijadikan akar (root node) dengan menghitung nilai gain yang paling tinggi. Sebelumnya yang akan dihitung adalah nilai entropy semua data. Prorses perhitungan entropy dan gain dilakukan untuk menentukkan akar (root) dari pohon keputusan dalam mengklasifikasi mahasiswa berpotensi Drop Out.

1. Entropi

Entropi adalah nilai informasi yang menyatakan ukuran ketidakpastian (impurity) dari attribut dari suatu kumpulan obyek data dalam satuan bit. Berikut adalah hasil perhitungan entropy.

Tabel 4.6. Menghitung Entropy

ATRIBUT		JUMLAH DATA	TIDAK DO	DO	ENTROPY
TOTAL		230	217	13	0,31347990
JENIS KELAMIN					
	Laki-Laki	41	36	5	0,53494370
	Perempuan	189	181	8	0,25286820
IPK					
	Baik	90	90	0	0
	Cukup	127	125	2	0,11684976
	Kurang	13	2	11	0,61938219
SKS					
	Memenuhi	222	217	5	0,15537867
	Tidak Memenuhi	8	0	8	0
PENGHASILAN ORANG TUA					
	0	0	0	0	0
	1	0	0	0	0
	2	99	91	8	0,40502013
	3	131	126	5	0,23382826
	4	0	0	0	0
	5	0	0	0	0
JENIS TEMPAT TINGGAL					
	0	125	119	6	0,27783957
	1	70	65	5	0,37123233
	2	34	32	2	0,32275696
	3	1	1	0	0
	4	0	0	0	0
	5	0	0	0	0
JENIS TRANSPORTASI					
	0	110	108	4	0,19985852
	1	58	52	6	0,47983202
	2	62	57	3	0,32293807
	3	0	0	0	0
	4	0	0	0	0

(P) ISSN 2442-451X
(O) ISSN 2503-3832

2. Gain

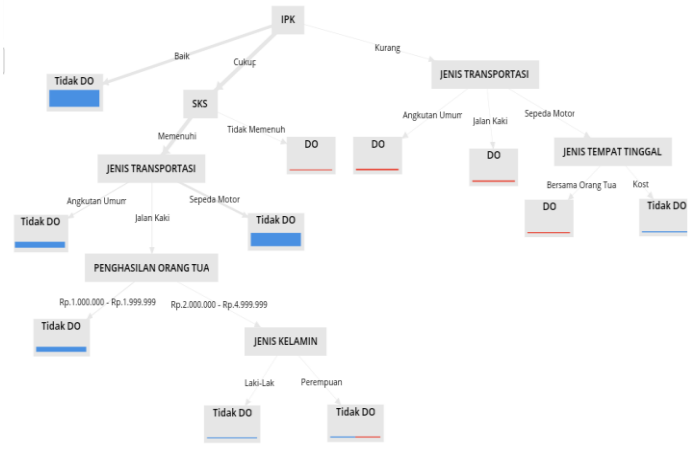
Setelah menghitung entropy kemudian menghitung nilai gain setiap atribut. Information gain adalah ukuran efektifitas suatu atribut dalam mengklasifikasikan data. Gain digunakan untuk menentukan urutan atribut dimana attribut yang memiliki nilai Information gain terbesar yang dipilih. Berikut hasil perhitungan nilai gain.

Tabel 4.7 Menghitung Gain

ATRIBUT		JUMLAH DATA	TIDAK DO	DO	ENTROPY	GAIN
TOTAL		230	217	13	0,31347990	
JENIS KELAMIN						0,010330320
	Laki-Laki	41	36	5	0,53494370	
	Perempuan	189	181	8	0,25286620	
IPK						0,213949956
	Baik	90	90	0	0	
	Cukup	127	125	2	0,11684976	
	Kurang	13	2	11	0,61938219	
SKS						0,163505711
	Memenuhi	222	217	5	0,15537867	
	Tidak Memenuhi	8	0	8	0	
PENGHASILAN ORANG TUA						0,005964706
	0	0	0	0	0	
	1	0	0	0	0	
	2	99	91	8	0,40502013	
	3	131	126	5	0,23382826	
	4	0	0	0	0	
	5	0	0	0	0	
JENIS TEMPAT TINGGAL						0,001784486
	0	125	119	6	0,27783957	
	1	70	65	5	0,37123233	
	2	34	32	2	0,32275696	
	3	1	1	0	0	
	4	0	0	0	0	
	5	0	0	0	0	
JENIS TRANSPORTASI						0,009841403
	0	110	108	4	0,19985852	
	1	58	52	6	0,47983202	
	2	62	57	3	0,32293807	
	3	0	0	0	0	
	4	0	0	0	0	

3. Pohon Keputusan

Setelah nilai entropy dan gain dihitung, kemudian dilakukan uji pohon keputusan dengan menggunakan RapidMiner. Hasil uji pohon keputusan yang dilakukan oleh algoritma Decision Tree C4.5 di RapidMiner Studio menunjukkan bahwa variabel IPK menjadi root.



Gambar 4.1. Hasil Pohon Keputusan

Berdasarkan hasil tersebut, atribut IPK merupakan hal yang sangat berpengaruh dalam mengklasifikasi mahasiswa berpotensi Drop Out. Dapat dilihat bahwa mahasiswa yang memiliki IPK dengan kategori Baik tidak berpotensi Drop Out.

Adapun rules atau aturan yang dihasilkan dari uji pohon keputusan dapat dilihat pada Tabel 4.8 dimana diperoleh 11 rules atau aturan untuk klasifikasi mahasiswa berpotensi Drop Out yaitu:

Tabel 4.8 Rule Prediksi Algoritma Decision Tree

Aturan	Deskripsi
Rule 1	IF IPK="Baik", Then mahasiswa tersebut tidak berpotensi Tidak DO
Rule 2	IF IPK="Cukup" And SKS="tidak memenuhi", Then mahasiswa tersebut berpotensi DO
Rule 3	IF IPK="Kurang" And JenisTransportasi="Angkutan Umum" Then Mahasiswa tersebut berpotensi DO
Rule 4	IF IPK="Kurang" And Jenis Transportasi="Jalan Kaki" Then Mahasiswa tersebut berpotensi DO
Rule 5	IF IPK="Cukup" And SKS="Memenuhi" And Jenis Transportasi="Angkutan Umum" Then mahasiswa tersebut memiliki kemungkinan untuk Tidak DO
Rule 6	IF IPK="Cukup" And SKS="Memenuhi" And Jenis Transportasi="Sepea Motor" Then Mahasiswa tersebut berpotensi Tidak DO
Rule 7	IF IPK="Kurang" And Jenis Transportasi="Sepeda Motor" And Jenis Tempat Tinggal="Bersama Orangtua" Then Mahasiswa tersebut berpotensi DO
Rule 8	IF IPK="Kurang" And Jenis Transportasi="Sepeda Motor" And Jenis Tempat Tinggal="Kost" Then Mahasiswa tersebut berpotensi Tidak DO
Rule 9	IF IPK="Cukup" And SKS="Memenuhi" And Jenis Transportasi="Jalan Kaki" And Penghasilan Orangtua="Rp.1.000.000 – Rp.1.999.999' Then mahasiswa tersebut memiliki kemungkinan untuk Tidak DO
Rule 10	IF IPK="Cukup" And SKS="Memenuhi" And Jenis Transportasi="Jalan Kaki" And Penghasilan Orangtua="Rp.2.000.000 – Rp.4.999.999' And Jenis Kelamin="Laki-Laki" Then mahasiswa tersebut memiliki kemungkinan untuk Tidak DO
Rule 11	IF IPK="Cukup" And SKS="Memenuhi" And Jenis Transportasi="Jalan Kaki" And Penghasilan Orangtua="Rp.2.000.000 – Rp.4.999.999' And Jenis Kelamin="Perempuan" Then mahasiswa tersebut memiliki kemungkinan untuk Tidak DO

Hasil pelatihan algoritma Decision Tree adalah pohon keputusan yang kemudian digunakan untuk menentukan data DO atau Tidak DO. Selanjutnya akan dilakukan pengujian. Pada tahap pengujian ini, data yang digunakan adalah Data testing. Langkah selanjutnya adalah mengukur performa hasil klasifikasi dengan menggunakan Confusion Matrix. Dan didapatkan hasil keakuratan sebesar 99,13%, sensitivity 100% dan presisi data terhadap pengujian sebesar 84,61%, specificity 99,87% dan F1-Score sebesar 91,67%.

5. KESIMPULAN

Kesimpulan yang dapat diperoleh dari penelitian ini adalah Perbandingan algoritma Naive Bayes dan Decision Tree menghasilkan masing-masing nilai akurasi 97,83% dan 99,13%. Dengan nilai rata-rata precision dan recall yaitu 61,54% dan 100% untuk algoritma Naive Bayes serta 84,61% dan 100% untuk algoritma Decision Tree. Nilai rata rata f1-score algoritma Naive bayes sebesar 76,19% dan Decision Tree sebesar 91,67%. Dari hasil yang didapat, algoritma Decision Tree C4.5 memiliki hasil peforma lebih tinggi dari algoritma Naive Bayes. Sehingga algoritma Decision Tree merupakan algoritma paling baik dalam mengklasifikasi mahasiswa berpotensi drop out di kampus Unasman. Dalam mengklasifikasi mahasiswa berpotensi drop out, diimplementasi sebuah pohon keputusan yang menghasilkan 11 rule (aturan) dengan atribut IPK sebagai akarnya. Itu artinya IPK merupakan

(P) ISSN 2442-451X
(O) ISSN 2503-3832

faktor paling berpengaruh dalam klasifikasi mahasiswa berpotensi drop out.

Daftar Pustaka

[1] Soni ahmad.,"Peranan Perguruan Tinggi Dalam Meningkatkan Kualitas Pendidikan Di Indonesia Untuk Menghadapi Asean Community 201533," 2015.

[2] T. Sinta Peringkat *et al.*, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," [Online]. Available: www.bri-institute.ac.id.

[3] inggit hasta Wijaya, "Prediksi Mahasiswa Drop Out Berdasarkan Klasifikasi Administratif," *Simki-Techsain*, vol. 02, no. 01, pp. 1–8, 2018.

[4] S. Suyadi, A. Setyanto, and H. Al Fattah, "Analisis Perbandingan Algoritma Decision Tree (C4.5) Dan K-Naive Bayes Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas," *Indones. J. Appl. Informatics*, vol. 2, no. 1, p. 59, 2017, doi: 10.20961/ijai.v2i1.13258.

[5] I. C. Wibowo, A. C. Fauzan, M. D. P. Yustiana, and F. A. Qhabib, "Komparasi Algoritma Naive Bayes dan Decision Tree Untuk Memprediksi Lama Studi Mahasiswa," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 1, no. 2, pp. 65–74, Dec. 2019, doi: 10.28926/ilkomnika.v1i2.21.

[6] F. Rizki, A. Faisol, and F. Santi Wahyuni, "Penerapan Metode Naive Bayes Untuk Memprediksi Penjualan Pada Ud. Hikmah Pasuruan Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 1, pp. 26–34, 2020, doi: 10.36040/jati.v4i1.2379.

[7] I. F. Yang, B. Dengan, M. Putus, K. Di, and I. P. B. Angkatan, "Identifikasi Faktor-Faktor Yang Berhubungan Dengan Mahasiswa Putus Kuliah Di Ipb Angkatan 2008 Menggunakan Analisis Survival," *Xplore J. Stat.*, vol. 1, no. 2, pp. 2–7, 2013, doi: 10.29244/xplore.v1i2.12404.

[8] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, "Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)," *J. Saintikom*, vol. 15, no. 2, pp. 81–92, 2016.

[9] A. Fathan Hidayatullah, M. Rifqi Ma, and arif Program Studi Manajemen Informatika STMIK Jenderal Achmad Yani Yogyakarta Jl Ringroad Barat, "Penerapan Text Mining dalam Klasifikasi Judul Skripsi," *Semin. Nas. Apl. Teknol. Inf. Agustus*, pp. 1907–5022, 2016.

[10] Rismayanti, "Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan)," *J. Sist. Inf.*, vol. 02, no. 01, pp. 16–24, 2018.

[11] G. D. M. Zulma and N. Chamidah, "Perbandingan Metode Klasifikasi Naive Bayes, Decision Tree Dan K-Nearest Neighbor Pada Data Log Firewall," *Senamika*, no. April, pp. 679–688, 2021, [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1396.