

# Desain Aplikasi Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes dan Algoritma C4.5

Suriana, Evi Dianti Bintari, Deby Kurniawan

**Abstraksi**— Kelulusan merupakan permasalahan umum bagi pihak universitas dan mahasiswa, karena kedua pihak tersebut sama-sama tidak dapat memprediksi atau mengelompokkan waktu kelulusan mahasiswa. Dengan adanya masalah ini, perlu untuk menciptakan sistem yang dapat memprediksi atau mengklasifikasikan kelulusan mahasiswa. Pada penelitian ini penulis menggunakan metode Naive Bayes Classifier dan Algoritma C4.5. Penulis menggunakan bahasa pemrograman PHP (*Personal Home Page*) serta MySql sebagai databasenya. Penelitian ini menggunakan beberapa parameter atau kriteria, yaitu IPK (Indeks Prestasi Kumulatif), SKS (Satuan Kredit Semester), Status Sipil, Status Pekerjaan, Waktu Kuliah, dan Pernah Cuti atau Tidaknya mahasiswa. Penelitian ini menggunakan 150 data sebagai Data Latih (*Data Training*) dan dilakukan klasifikasi terhadap Data Uji (*Data Testing*). Hasil dari penelitian ini akan terbagi menjadi 3 kelas klasifikasi kelulusan, yaitu: LULUS CEPAT, TEPAT WAKTU dan TERLAMBAT. Dilakukan uji akurasi sebanyak 5 kali untuk masing-masing metode dan diperoleh rata-rata akurasi untuk metode Naive Bayes Classifier sebesar 69.9 % dan rata-rata akurasi untuk metode Algoritma C4.5 sebesar 70.6 %.

**Kata Kunci**— desain, aplikasi, klasifikasi kelulusan mahasiswa, naive bayes classifier, algoritma c4.5.

## I. PENDAHULUAN

Lulus tepat waktu adalah keinginan seluruh mahasiswa. Tidak hanya itu, lulus tepat waktu adalah keuntungan bagi dua pihak. Pertama, pihak mahasiswa, karena dengan begitu mahasiswa akan mendapatkan pekerjaan dengan lebih mudah karena perusahaan cenderung mencari *fresh graduate*. Selain itu, mahasiswa akan memperoleh keuntungan dari segi biaya, karena lulus tepat waktu sama dengan tidak menambah biaya yang harus dikeluarkan jika harus membayar biaya semester. Kedua, pihak perguruan tinggi, karena seiring tepatnya waktu kelulusan mahasiswa, hal itu akan membantu memajukan kualitas perguruan tinggi tersebut, seperti peningkatan akreditasi.

Sayangnya, waktu kelulusan mahasiswa tidak selalu dapat dideteksi secara dini, sehingga bisa mengakibatkan keterlambatan kelulusan. Hal ini tentunya merugikan kedua pihak. Untuk memecahkan masalah tersebut, perlu adanya suatu sistem atau program yang dapat mengelompokkan golongan mahasiswa berdasarkan waktu kelulusan.

Hingga saat ini di STMIK PPKIA Tarakanita Rahmawati Tarakan masih belum memiliki sistem yang mampu mengelompokkan kelulusan mahasiswanya. Sehingga setiap tahunnya tingkat kelulusan mahasiswa terkadang belum sesuai dengan prediksi kelulusan yang di perkirakan oleh pihak akademik. Hal ini tentu saja mempengaruhi kelancaran kegiatan belajar mengajar maupun tingkat penerimaan mahasiswa baru di Perguruan Tinggi tersebut.

Ada beberapa faktor yang mempengaruhi klasifikasi kelulusan mahasiswa, diantaranya: IPK (Indeks Prestasi Kumulatif), jumlah SKS (Satuan Kredit Semester), status pekerjaan mahasiswa, status sipil mahasiswa, waktu kuliah, dan status keaktifan mahasiswa (apakah pernah melakukan cuti atau tidak).

Oleh karena itu, penulis terdorong untuk membuat Desain Aplikasi Klasifikasi Kelulusan Mahasiswa pada STMIK PPKIA Tarakanita Rahmawati Tarakan. Pada penelitian ini, penulis menggunakan metode Naive Bayes dan Algoritma C4.5. Dengan menggunakan metode ini akan dilakukan penilaian dengan menggunakan Metode Naive Bayes yang dibandingkan dengan penilaian menggunakan Algoritma C4.5. Penilaian tersebut digunakan untuk membagi kelulusan mahasiswa dalam 3 (tiga) kelas, yaitu: Lulus Cepat, Tepat Waktu, dan Terlambat. Proses klasifikasi tersebut dilakukan dengan menggunakan beberapa variabel, antara lain IPK (Indeks Prestasi Kumulatif), jumlah SKS (Satuan Kredit Semester), status pekerjaan mahasiswa, status sipil mahasiswa, waktu kuliah, dan status keaktifan mahasiswa (apakah pernah melakukan cuti atau tidak). Kemudian dilakukan analisa perbandingan hasil dan akurasi dari kedua metode tersebut yang nantinya akan digunakan sebagai acuan untuk mengambil langkah yang tepat dalam mencapai tingkat kelulusan optimal yang diinginkan.

## II. TINJAUAN PUSTAKA

Desain merupakan kerangka bentuk, rancangan [1]. Desain adalah suatu sistem yang berlaku untuk segala jenis perancangan dimana titik beratnya adalah melihat segala sesuatu persoalan tidak secara terpisah atau tersendiri, melainkan sebagai suatu kesatuan dimana satu masalah dengan lainnya saling terkait.

Aplikasi berasal dari Bahasa Inggris "*Application*" yang bermakna penerapan atau penggunaan. Pengertian aplikasi secara umum adalah alat terapan yang difungsikan secara

khusus dan terpadu sesuai kemampuan yang dimilikinya aplikasi merupakan suatu perangkat komputer yang siap pakai bagi user. Bila dimaknai secara istilah, aplikasi adalah suatu program yang siap untuk digunakan yang dibuat untuk melaksanakan suatu fungsi bagi pengguna jasa aplikasi serta penggunaan aplikasi lain yang dapat digunakan oleh suatu sasaran yang akan dituju. Menurut sebuah kamus komputer, aplikasi mempunyai arti yaitu pemecahan masalah yang menggunakan salah satu teknik pemrosesan data aplikasi yang biasanya berpaku pada sebuah komputasi yang diinginkan maupun pemrosesan data yang diharapkan.

Klasifikasi (*Classification*) adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Klasifikasi juga dapat diartikan menentukan sebuah *record* data baru ke salah satu dari beberapa kategori (atau kelas) yang telah didefinisikan sebelumnya [2]. Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasikan suatu spesies tertentu. Komponen-komponen utama dari proses klasifikasi antara lain:

1. Kelas, merupakan variabel tidak bebas yang merupakan label dari hasil klasifikasi.
2. Prediktor, merupakan variabel bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasikan.
3. Set data pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan prediktor untuk dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat.
4. et data uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat [3].

Proses klasifikasi biasanya dibagi menjadi dua fase yakni *learning* dan *testing*. Pada fase *learning* sebagian data yang telah diketahui kelas datanya diumpamakan untuk membentuk model perkiraan. Kemudian pada fase *testing* model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

Metode-metode yang telah dikembangkan untuk menyelesaikan kasus klasifikasi antara lain Pohon Keputusan, Naive Bayes, Jaringan Syaraf Tiruan, Analisis Statistik, Algoritma Genetik, Rought Sets, K-Nearest Neighbour dan metode berbasis aturan

Lulus adalah berhasil (dalam ujian), dapat melalui dengan baik (dalam menghadapi segala cobaan). Kelulusan juga berarti hal (keadaan) lulus (ujian dan sebagainya) [4]. Sedangkan mahasiswa adalah peserta didik yang terdaftar dan belajar di perguruan tinggi tertentu [5]. Secara lebih singkat mahasiswa yaitu suatu kelompok dalam masyarakat yang memperoleh status karena ikatan dengan perguruan tinggi, universitas, institut ataupun akademi.

Untuk dapat mencapai gelar kesarjanaan pada program studi Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, setiap mahasiswa dinyatakan lulus, apabila telah melaksanakan pengambilan mata kuliah sesuai total SKS yang dibebankan kepadanya. Beban studi mahasiswa dalam satu semester adalah jumlah sks yang dapat diambil oleh mahasiswa

dalam semester yang bersangkutan. Beban studi program sarjana adalah 144-147 SKS (Satuan Kredit Semester) atau lebih termasuk skripsi atau tugas akhir, dan dijadwalkan untuk 8 (delapan) semester. Dalam pelaksanaan studi dapat ditempuh dalam waktu kurang dari 8 semester dan selama-lamanya 12 (dua belas) semester. Mahasiswa wajib mengikuti Ujian Tengah Semester (UTS) dan Ujian Akhir Semester (UAS) yang dilaksanakan pada setiap mata kuliah yang diambil. Mahasiswa jenjang Strata 1 (S1) diwajibkan untuk melakukan penyelesaian Tugas Akhir/Skripsi sebagai penentu kriteria kelulusan. Mahasiswa yang telah menyelesaikan Tugas Akhir/Skripsi harus melakukan seminar hasil didepan mahasiswa lain. Setelah itu mahasiswa diwajibkan mengikuti ujian komprehensiv dan seminar akhir. Mahasiswa dinyatakan lulus dalam penyelesaian studinya, apabila telah memenuhi ketentuan tersebut serta tidak melewati batas masa studi yang ditentukan [6].

Naive Bayes Classifier (NBC) adalah salah satu metode pada bidang keilmuan *machine learning* yang menggunakan perhitungan probabilitas. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [7].

NBC juga merupakan pengklasifikasian statistic yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. NBC didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decition tree* dan *neural network*. NBC terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar [8]. Dalam terminologi sederhana, NBC mengasumsikan bahwa kehadiran (atau ketiadaan) fitur tertentu dari suatu kelas tidak berhubungan dengan kehadiran (atau ketiadaan) fitur lainnya. Sebuah keuntungan dari NBC yaitu NBC hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata varian dan variabel) diasumsikan independen, hanya varian dari variabel-variabel untuk setiap kelas yang perlu ditentukan. Metode NBC ini akan digunakan untuk mengolah data-data mahasiswa sehingga kelulusan mahasiswa dapat diklasifikasikan. Berikut dasar dari teorema NBC yang digunakan.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

X : data dengan class yang belum diketahui

H : hipotesis data X merupaka suatu class spesifik

P(H|X) : probabilitas hipotesis H berdasar kondisi X

P(H) : probabilitas hipotesis H

P(X|H) : probabilitas X berdasar kondisi pada hipotesis H

P(X) : probabilitas dari X.

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Algoritma C4.5 merupakan kelompok algoritma Decision Tree. Algoritma ini mempunyai input berupa *training samples* dan *samples*. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan samples merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [9].

Pohon keputusan sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Pohon keputusan juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, sangat baik sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain [10].

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, temperature, kelembaban dan angin. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut cuaca mempunyai instance berupa cerah, berawan dan hujan [11]. Proses pada pohon keputusan adalah: mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi aturan dan menyederhanakan aturan. Untuk memudahkan penjelasan mengenai algoritma C4.5.

### III. ANALISA DAN DESAIN SISTEM

#### A. Penerapan Metode

Aplikasi klasifikasi kelulusan mahasiswa dengan menerapkan metode Naïve Bayes Classifier dan Algoritma C4.5 ini dirancang menggunakan Bahasa Pemrograman PHP (*Personal Home Page*) dan MySQL sebagai databasenya.

Studi kasus penelitian ini ialah Klasifikasi Kelulusan Mahasiswa pada STMIK PPKIA Tarakanita Rahmawati Tarakan menggunakan Metode Naïve Bayes Classifier dan Algoritma C4.5. Untuk sampel studi kasus akan menggunakan 30 data testing atau data uji dari 150 data training atau data latih.

1. Menentukan kriteria penilaian, yaitu :

TABLE I  
 TABEL BOBOT KRITERIA

Nama Kriteria
IPK (Indeks Prestasi Kumulatif)
SKS (Satuan Kredit Semester)
Status Sipil
Status Pekerjaan
Waktu Kuliah
Status Keaktifan (Pernah Cuti atau Tidaknya Mahasiswa)

2. Selanjutnya menentukan sub kriteria, dimana pada studi kasus ini masing-masing kriteria memiliki sub kriteria, yaitu:

TABLE II  
 TABLE SUB KRITERIA

Kriteria	Sub Kriteria	Keterangan
IPK (Indeks Prestasi Kumulatif)	M	3,51 – 4,00
	B	3,01 – 3,50
	C	2,51 – 3,00
SKS (Satuan Kredit Semester)	143	-
	144	-
	145	-
	146	-
	147	-
	148	-
	149	-
Status Sipil	150	-
	M	Menikah
Status Pekerjaan	B	Belum Menikah
	TB	Bekerja
Waktu Kuliah	P	Tidak Bekerja
	M	Pagi
Status Keaktifan	P	Malam
	T	Pernah Cuti
		Tidak Pernah Cuti

TABLE III  
 TABLE DATA LATIH

Nama	I P K	SKS	Status Sipil	Status Pekerja an	Waktu Kuliah	Status Keaktif an	Kelulus
Data Latih 1	M	147	BM	B	P	T	Tepat Waktu
Data Latih 2	C	147	BM	TB	P	T	Tepat Waktu
Data Latih 3	B	147	BM	TB	P	T	Tepat Waktu
Data Latih 4	B	149	BM	TB	P	T	Tepat Waktu
Data Latih 5	B	147	BM	B	P	T	Tepat Waktu
Data Latih 6	B	147	BM	B	P	T	Tepat Waktu
Data Latih 7	B	147	BM	B	M	T	Tepat Waktu
Data Latih 8	B	147	BM	TB	P	T	Tepat Waktu
Data Latih 9	B	147	BM	TB	P	T	Tepat Waktu
Data Latih 10	C	147	BM	TB	P	T	Tepat Waktu
....	..	...	...	...	...	...	...
Data Latih 51	B	147	BM	TB	P	T	Tepat Waktu
Data Latih 52	B	145	BM	B	M	T	Tepat Waktu
Data Latih 53	C	145	BM	B	P	T	Tepat Waktu
Data Latih 54	C	144	BM	B	M	T	Terlambat
Data Latih 55	M	147	BM	B	M	T	Tepat Waktu
Data Latih 56	C	145	BM	B	M	P	Terlambat

3. Menentukan data mahasiswa yang akan diklasifikasikan. Berikut adalah data latih yang digunakan dalam penelitian ini, dapat dilihat pada Tabel III.

Data Latih 57	C	148	M	B	M	P	Terlambat
Data Latih 58	B	148	M	TB	P	T	Tepat Waktu
Data Latih 59	M	149	BM	TB	P	T	Lulus Cepat
Data Latih 60	B	149	BM	B	M	T	Lulus Cepat

Dari data latih yang ada pada Tabel III akan dilakukan proses klasifikasi menggunakan metode Naïve Bayes Classifier. Langkah- langkah yang dilakukan dalam proses klasifikasi tersebut, antara lain:

1. Menghitung jumlah data per kelas.

Dari 150 data latih yang ada pada Tabel 3.3, dilakukan perhitungan jumlah mahasiswa untuk masing-masing kelas berdasarkan klasifikasi yang terbentuk (*prior probability*) seperti dibawah ini:

- $P(\text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{\text{jumlah "Lulus Cepat"}}{\text{pada kolom Kelulusan}} = \frac{54}{150} = 0.36$
- $P(\text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{\text{jumlah "Tepat Waktu"}}{\text{pada kolom Kelulusan}} = \frac{59}{150} = 0.393$
- $P(\text{Kelulusan} = \text{"Terlambat"}) = \frac{\text{jumlah "Terlambat"}}{\text{pada kolom Kelulusan}} = \frac{37}{150} = 0.25$

2. Menghitung jumlah kasus yang sama pada setiap kriteria berdasarkan kelas Kelulusan seperti dibawah ini:

- $P(\text{IPK} = \text{"M"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{14}{54} = 0.25926$
- $P(\text{IPK} = \text{"M"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{4}{59} = 0.0678$
- $P(\text{IPK} = \text{"M"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{0}{37} = 0$
- $P(\text{IPK} = \text{"B"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{38}{54} = 0.7037$
- $P(\text{IPK} = \text{"B"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{44}{59} = 0.74576$
- $P(\text{IPK} = \text{"B"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{18}{37} = 0.48649$
- $P(\text{IPK} = \text{"C"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{2}{54} = 0.03704$
- $P(\text{IPK} = \text{"C"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{11}{59} = 0.18644$
- $P(\text{IPK} = \text{"C"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{19}{37} = 0.51351$
- $P(\text{SKS} = \text{"143"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{0}{54} = 0$
- $P(\text{SKS} = \text{"143"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{0}{59} = 0$
- $P(\text{SKS} = \text{"143"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{1}{37} = 0.02703$
- $P(\text{SKS} = \text{"144"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{0}{54} = 0$
- $P(\text{SKS} = \text{"144"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{1}{59} = 0.01695$
- $P(\text{SKS} = \text{"144"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{1}{37} = 0.02703$
- $P(\text{SKS} = \text{"145"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{15}{54} = 0.27778$
- $P(\text{SKS} = \text{"145"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{4}{59} = 0.0678$
- $P(\text{SKS} = \text{"145"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{5}{37} = 0.13514$
- $P(\text{SKS} = \text{"146"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{16}{54} = 0.27778$
- $P(\text{SKS} = \text{"146"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{11}{59} = 0.0678$
- $P(\text{SKS} = \text{"146"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{13}{37} = 0.13514$
- $P(\text{SKS} = \text{"147"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{19}{54} = 0.35185$
- $P(\text{SKS} = \text{"147"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{22}{59} = 0.37288$
- $P(\text{SKS} = \text{"147"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{8}{37} = 0.21622$
- $P(\text{SKS} = \text{"148"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{0}{54} = 0$
- $P(\text{SKS} = \text{"148"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{9}{59} = 0.15254$
- $P(\text{SKS} = \text{"148"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{2}{37} = 0.08108$
- $P(\text{SKS} = \text{"149"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{3}{54} = 0.05556$

- $P(\text{SKS} = \text{"149"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{10}{59} = 0.16949$
- $P(\text{SKS} = \text{"149"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{1}{37} = 0.02703$
- $P(\text{SKS} = \text{"150"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{1}{54} = 0.01852$
- $P(\text{SKS} = \text{"150"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{2}{59} = 0.0339$
- $P(\text{SKS} = \text{"150"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{5}{37} = 0.13514$
- $P(\text{Status Sipil} = \text{"M"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{6}{54} = 0.11111$
- $P(\text{Status Sipil} = \text{"M"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{4}{59} = 0.0678$
- $P(\text{Status Sipil} = \text{"M"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{9}{37} = 0.24324$
- $P(\text{Status Sipil} = \text{"BM"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{48}{54} = 0.89$
- $P(\text{Status Sipil} = \text{"BM"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{55}{59} = 0.9322$
- $P(\text{Status Sipil} = \text{"BM"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{28}{37} = 0.75676$
- $P(\text{Status Pekerjaan} = \text{"B"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{33}{54} = 0.611$
- $P(\text{Status Pekerjaan} = \text{"B"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{26}{59} = 0.44$
- $P(\text{Status Pekerjaan} = \text{"B"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{26}{37} = 0.7027$
- $P(\text{Status Pekerjaan} = \text{"TB"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{21}{54} = 0.39$
- $P(\text{Status Pekerjaan} = \text{"TB"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{33}{59} = 0.56$
- $P(\text{Status Pekerjaan} = \text{"TB"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{11}{37} = 0.297$
- $P(\text{Waktu Kuliah} = \text{"P"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{26}{54} = 0.481$
- $P(\text{Waktu Kuliah} = \text{"P"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{41}{59} = 0.69$
- $P(\text{Waktu Kuliah} = \text{"P"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{13}{37} = 0.351$
- $P(\text{Waktu Kuliah} = \text{"M"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{28}{54} = 0.518$
- $P(\text{Waktu Kuliah} = \text{"M"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{18}{59} = 0.305$
- $P(\text{Waktu Kuliah} = \text{"M"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{24}{37} = 0.648$
- $P(\text{Status Keaktifan} = \text{"P"} | \text{Kelulusan} = \text{"Lulus Cepat"}) = \frac{1}{54} = 0.018$
- $P(\text{Status Keaktifan} = \text{"P"} | \text{Kelulusan} = \text{"Tepat Waktu"}) = \frac{1}{59} = 0.016$
- $P(\text{Status Keaktifan} = \text{"P"} | \text{Kelulusan} = \text{"Terlambat"}) = \frac{12}{37} = 0.324$

3. Kalikan semua hasil variabel berdasarkan data uji seperti dibawah ini:

Dari 150 data latih yang digunakan, diambil 30 data, secara acak sebagai data uji.

TABLE IV  
 TABLE DATA LATIH

Nama	I P K	SKS	Status Sipil	Status Pekerjaan	Waktu Kuli ah	Status Keaktif an	Kelulusan
Data Uji 1	M	147	BM	B	P	T	Tepat Waktu
Data Uji 2	C	147	BM	TB	P	T	Tepat Waktu

Data Uji 3	B	147	BM	TB	P	T	Tepat Waktu
Data Uji 4	B	149	BM	TB	P	T	Tepat Waktu
Data Uji 5	B	147	BM	B	P	T	Tepat Waktu
Data Uji 6	B	147	BM	B	P	T	Tepat Waktu
Data Uji 7	B	147	BM	B	M	T	Tepat Waktu
Data Uji h 8	B	147	BM	TB	P	T	Tepat Waktu
Data Uji 9	B	147	BM	TB	P	T	Tepat Waktu
Data Uji 10	C	147	BM	TB	P	T	Tepat Waktu
Data Uji 11	C	147	BM	TB	P	T	Tepat Waktu
Data Uji 12	B	147	BM	TB	P	T	Lulus Cepat
Data Uji 13	M	147	BM	B	P	T	Lulus Cepat
Data Uji 14	M	147	BM	B	M	T	Lulus Cepat
Data Uji 15	M	147	BM	TB	P	T	Lulus Cepat
Data Uji 16	B	147	BM	B	M	T	Lulus Cepat
Data Uji 17	C	144	BM	B	M	T	Terlambat
Data Uji 18	B	148	M	TB	P	T	Tepat Waktu
Data Uji 19	B	149	BM	TB	P	T	Tepat Waktu
Data Uji 20	B	149	BM	TB	P	T	Tepat Waktu
Data Uji 21	B	149	BM	TB	P	T	Tepat Waktu
Data Uji 22	B	147	BM	TB	P	T	Tepat Waktu
Data Uji 23	B	148	BM	TB	P	T	Tepat Waktu
Data Uji 24	B	149	BM	TB	P	T	Tepat Waktu
Data Uji 25	B	146	BM	B	M	T	Tepat Waktu
Data Uji 26	B	148	BM	B	M	T	Tepat Waktu
Data Uji 27	C	150	BM	B	M	T	Terlambat
Data Uji 28	C	150	BM	TB	P	T	Terlambat
Data Uji 29	C	146	BM	B	M	T	Terlambat
Data Uji 30	B	148	M	TB	P	T	Tepat Waktu

Lakukan perhitungan hasil variabel berdasarkan Data Uji 1.

a. Untuk semua atribut Kelulusan = "Lulus Cepat"

$$P(X | \text{Kelulusan} = \text{"Lulus Cepat"}) = 0.25926 \times 0.35185 \times 0.88889 \times 0.6111 \times 0.48148 \times 0.98148 = 0.02341$$

Perkalian prior probability dengan semua atribut Kelulusan = "Lulus Cepat"

$$P(\text{Kelulusan} = \text{"Lulus Cepat"}) \times P(X | \text{Kelulusan} = \text{"Lulus Cepat"}) = 0.36 \times 0.02341 = 0.0084276$$

b. Untuk semua atribut Kelulusan = "Tepat Waktu"

$$P(X | \text{Kelulusan} = \text{"Tepat Waktu"}) = 0.0678 \times 0.37288 \times 0.9322 \times 0.44068 \times 0.6949 \times 0.98305$$

= 0.007084  
 Perkalian prior probability dengan semua atribut Kelulusan = "Tepat Waktu"

$$P(\text{Kelulusan} = \text{"Tepat Waktu"}) \times P(X | \text{Kelulusan} = \text{"Tepat Waktu"}) = 0.393 \times 0.007084 = 0.002786$$

c. Untuk semua atribut Kelulusan = "Terlambat"

$$P(X | \text{Kelulusan} = \text{"Terlambat"}) = 0 \times 0.21622 \times 0.75676 \times 0.7027 \times 0.35135 \times 0.67568 = 0$$

Perkalian prior probability dengan semua atribut Kelulusan = "Terlambat"

$$P(\text{Kelulusan} = \text{"Terlambat"}) \times P(X | \text{Kelulusan} = \text{"Terlambat"}) = 0.2467 \times 0 = 0$$

4. Bandingkan Hasil Kelas

$$P(\text{Kelulusan} = \text{"Lulus Cepat"}) \times P(X | \text{Kelulusan} = \text{"Lulus Cepat"}) > P(\text{Kelulusan} = \text{"Tepat Waktu"}) \times P(X | \text{Kelulusan} = \text{"Tepat Waktu"})$$

$$P(\text{Kelulusan} = \text{"Lulus Cepat"}) \times P(X | \text{Kelulusan} = \text{"Lulus Cepat"}) > P(\text{Kelulusan} = \text{"Terlambat"}) \times P(X | \text{Kelulusan} = \text{"Terlambat"})$$

Untuk Data Uji 1: Kelulusan = "Lulus Cepat" (Perhitungan antara perkalian Kelulusan "Lulus Cepat" dengan Kelulusan "Tepat Waktu" dengan Kelulusan "Terlambat" menunjukkan bahwa nilai Kelulusan = "Lulus Cepat" lebih besar dibandingkan kelas Kelulusan = "Tepat Waktu" dan nilai Kelulusan = "Lulus Cepat" lebih besar dibandingkan kelas Kelulusan = "Terlambat").

Lakukan perhitungan tersebut untuk data uji lain sehingga dapat diperoleh hasil klasifikasi kelulusan mahasiswa berdasarkan metode Naïve Bayes Classifier seperti Tabel V.

TABLE V  
 TABEL HASIL KLASIFIKASI KELULUSAN NAIVEBAYES CLASSIFIER

Nama	I P K	SKS	Statu s Sipil	Sta tus Pe ker jaa n	Wakt u Kuli ah	Statu s Keak tifan	Kelu lusan	Kelulusa n Berdasar kan Metode
Data Uji 1	M	147	BM	B	P	T	Tepat Waktu	Lulus Cepat
Data Uji 2	C	147	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 3	B	147	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 4	B	149	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 5	B	147	BM	B	P	T	Tepat Waktu	Tepat Waktu

Data Uji	B	147	BM	B	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 6	B	147	BM	B	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 7	B	147	BM	B	M	T	Waktu Tepat Waktu	Lulus Cepat
Data Uji 8	B	147	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 9	C	147	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 10	C	147	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 19	B	149	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 20	B	149	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 21	B	149	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 22	B	147	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 23	B	148	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 24	B	149	BM	TB	P	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 25	B	146	BM	B	M	T	Waktu Tepat Waktu	Lulus Cepat
Data Uji 26	B	148	BM	B	M	T	Waktu Tepat Waktu	Tepat Waktu
Data Uji 27	C	150	BM	B	M	T	Terlambat	Terlambat
Data Uji 28	C	150	BM	TB	P	T	Terlambat	Terlambat
Data Uji 29	C	146	BM	B	M	T	Terlambat	Terlambat
Data Uji 30	B	148	M	TB	P	T	Waktu Tepat Waktu	Tepat Waktu

Dari data latih yang ada pada Tabel 3.3 akan dilakukan proses klasifikasi menggunakan Algoritma C4.5. Langkah-langkah yang dilakukan dalam proses klasifikasi tersebut, antara lain:

1. Pilih atribut sebagai akar

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain, sebelumnya perlu dilakukan perhitungan jumlah kasus,

jumlah kasus untuk keputusan Lulus Cepat, jumlah kasus untuk keputusan Tepat Waktu, jumlah kasus untuk keputusan Terlambat, dan entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut IPK, SKS, Status Sipil, Status Pekerjaan, Waktu Kuliah dan Status Keaktifan (Pernah cuti atau tidaknya mahasiswa). Setelah itu lakukan penghitungan Gain untuk masing- masing atribut. Hasil perhitungan ditunjukkan oleh Tabel VI

TABLE VI  
 TABEL HASIL KLASIFIKASI KELULUSAN NAIVEBAYES CLASSIFIER

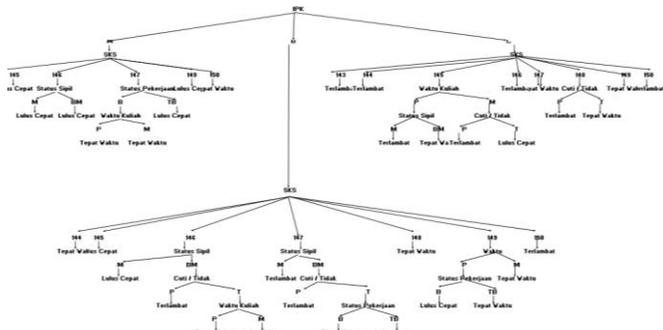
Kriteria	Subkriteria	Jumlah Kasus	Lulus Cepat	Tepat Waktu	Terlambat	Entropy	Gain	
IPK	M	18	14	4	0	0	0.29769	
	B	100	38	44	18	1.49691	0.29769	
	C	32	2	11	19	1.22611	0.27056	
SKS	143	1	0	0	1	0	0.27056	
	144	2	0	1	1	0		
	145	24	15	4	5	1.32609		
	146	40	16	11	13	1.57871		
	147	49	19	22	8	1.47556		
	148	12	0	9	3	0		
Stat	149	14	3	10	1	1.09491		
	150	8	1	2	5	1.29879		
	M	19	6	4	9	1.5090	0.0294	
Sipi	B	131	8	55	28	3	2	
	M					1.5305		
Stat	B	85	33	26	26	1.5788	0.0311	
	M	65	21	33	11	8	9	
Pek	erj					1.4568		
	an					7		
Wa	ktu	P	80	26	41	13	1.4472	0.0525
	Kul	M	70	28	18	24	2	6
iah						1.5702		
						6		
Stat	P	14	1	1	12	0.7345	0.1266	
	Kea	T	136	53	58	25	3	4
ktif						1,5022		
	an							
Tot						1.5572		
	al		150	54	59	37		

2. Dari hasil pada Tabel 3.6 dapat diketahui bahwa atribut dengan Gain tertinggi adalah IPK yaitu sebesar 0.29769. Dengan demikian IPK dapat menjadi node akar untuk perhitungan selanjutnya.

3. Ulangi proses untuk node selanjutnya hingga dapat dibetuk pohon keputusan seperti Gambar 3.1.

4. Dari pohon keputusan, diperoleh rule seperti pada Tabel VI

5. Dari rule yang telah diperoleh, lakukan klasifikasi terhadap data uji yang ada pada Tabel IV. Sehingga diperoleh hasil klasifikasi kelulusan mahasiswa berdasarkan Algoritma C4.5 seperti Tabel VII



Gambar 3.1 Pohon Keputusan

TABLE VI  
 TABEL RULE ALGORITMA C4.5

Kondisi	Hasil
If (IPK = M) and (SKS = 145)	Lulus Cepat
If (IPK = M) and (SKS = 146) and (Status Sipil = M)	Lulus Cepat
If (IPK = M) and (SKS = 146) and (Status Sipil = BM)	Lulus Cepat
If (IPK = M) and (SKS = 147) and (Status Pekerjaan = B) and (Waktu Kuliah = P)	Tepat Waktu
If (IPK = M) and (SKS = 147) and (Status Pekerjaan = B) and (Waktu Kuliah = M)	Tepat Waktu
If (IPK = M) and (SKS = 147) and (Status Pekerjaan = TB)	Lulus Cepat
If (IPK = M) and (SKS = 149)	Lulus Cepat
If (IPK = M) and (SKS = 150)	Tepat Waktu
If (IPK = B) and (SKS = 144)	Tepat Waktu
If (IPK = B) and (SKS = 145)	Lulus Cepat
If (IPK = B) and (SKS = 146) and (Status Sipil = M)	Lulus Cepat
.....	.....
If (IPK = B) and (SKS = 149) and (Waktu Kuliah = P) and (Pekerjaan = TB)	Tepat Waktu
If (IPK = B) and (SKS = 149) and (Waktu Kuliah = M)	Tepat Waktu
If (IPK = B) and (SKS = 150)	Terlambat
If (IPK = C) and (SKS = 143)	Terlambat
If (IPK = C) and (SKS = 144)	Terlambat
If (IPK = C) and (SKS = 145) and (Waktu Kuliah = P) and (Status Sipil = M)	Terlambat
If (IPK = C) and (SKS = 145) and (Waktu Kuliah = P) and (Status Sipil = BM)	Tepat Waktu
If (IPK = C) and (SKS = 145) and (Waktu Kuliah = M) and (Status Keaktifan = P)	Terlambat
If (IPK = C) and (SKS = 145) and (Waktu Kuliah = M) and (Status Keaktifan = T)	Lulus Cepat
If (IPK = C) and (SKS = 146)	Terlambat
If (IPK = C) and (SKS = 147)	Tepat Waktu

TABLE VII  
 TABEL HASIL KLASIFIKASI KELULUSAN NAIVEBAYES CLASSIFIER

Nama	I P K	Statu s Sipil	Statu s Pekerjaan	Wakt u Kuli ah	Sta tus Ke akt ifa n	Kelulusan	Kelulusa n Berdasar kan Metode
Data Uji 1	M	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 2	C	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 3	B	BM	TB	P	T	Tepat Waktu	Lulus cepat
Data Uji 4	B	BM	B	P	T	Tepat Waktu	Tepat Waktu
Data Uji 5	B	BM	B	P	T	Tepat Waktu	Tepat Waktu
Data Uji 6	B	BM	B	P	T	Tepat Waktu	Tepat Waktu
Data Uji 7	B	BM	TB	P	T	Tepat Waktu	Tepat Waktu

Data Uji 8	B	BM	TB	M	T	Tepat Waktu	Tepat Waktu
Data Uji 9	B	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 10	C	BM	TB	P	T	Tepat Waktu	Tepat Waktu
Data Uji 11	C	BM	TB	P	T	Tepat Waktu	Lulus Cepat
.....	..	...	..	..	..	.....	..
Data Uji 25	B	146	BM	M	T	Tepat Waktu	Tepat Waktu
Data Uji 26	B	148	BM	M	T	Tepat Waktu	Tepat Waktu
Data Uji 27	C	150	BM	M	T	Tepat Waktu	Terlambat
Data Uji 28	C	150	BM	P	T	Terlambat	Terlambat
Data Uji 29	C	146	BM	M	T	Terlambat	Terlambat
Data Uji 30	B	148	M	P	T	Tepat Waktu	Tepat Waktu

Salah satu aspek yang menjadi parameter kehandalan dari suatu algoritma klasifikasi adalah tingkat akurasi. Sebuah sistem dalam melakukan klasifikasi diharapkan dapat mengklasifikasi semua data uji dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% akurat. Berdasarkan perbedaan teknik antara metode Naïve Bayes Classifier dengan Algoritma C4.5, maka akan dilakukan proses mengenai perbandingan tingkat akurasi dari kedua algoritma klasifikasi tersebut. Tujuan yang diinginkan dari perbandingan akurasi ini adalah untuk mengetahui tingkat akurasi dari metode Naïve Bayes Classifier dan Algoritma C4.5 serta membandingkan kehandalan antara kedua algoritma tersebut yang diukur berdasarkan tingkat akurasinya. Untuk menghitung tingkat akurasi digunakan rumus seperti dibawah ini.

$$Akurasi = \frac{\text{Jumlah data yang diprediksi dengan benar}}{\text{Jumlah seluruh data yang diprediksi}} \times 100 \%$$

Untuk menghitung jumlah prediksi salah dalam suatu klasifikasi digunakan rumus seperti dibawah ini.

$$Error = \frac{\text{Jumlah data yang diprediksi salah}}{\text{Jumlah seluruh data yang diprediksi}} \times 100\%$$

Pada Tabel V telah diperoleh hasil klasifikasi kelulusan mahasiswa berdasarkan metode Naïve Bayes Classifier dan pada Tabel VIII telah diperoleh hasil klasifikasi kelulusan mahasiswa berdasarkan Algoritma C4.5. Dari hasil tersebut akan dilakukan proses perbandingan akurasi berdasarkan kelulusan mahasiswa pada data Latih dan kelulusan mahasiswa berdasarkan hasil klasifikasi metode. Akurasi hasil klasifikasi metode Naïve Bayes Classifier dapat dilihat pada Tabel VIII. Dari Tabel diketahui  
 Jumlah data yang diprediksi dengan Benar = 25 Data  
 Jumlah data yang diprediksi Salah = 5 Data

Selanjutnya, lakukan perhitungan tingkat akurasi berdasarkan rumus.

$$Akurasi = \frac{25 \text{ Data}}{30 \text{ Data}} \times 100 \%$$

Akurasi = 83.33%

TABLE VIII  
 TABEL HASIL KLASIFIKASI KELULUSAN NAIVEBAYES CLASSIFIER

Nama	Kelulusan	Kelulusan Berdasarkan Metode	Akurasi
Data Uji 1	Tepat Waktu	Lulus Cepat	Salah
Data Uji 2	Tepat Waktu	Tepat Waktu	Benar
Data Uji 3	Tepat Waktu	Tepat Waktu	Benar
Data Uji 4	Tepat Waktu	Tepat Waktu	Benar
Data Uji 5	Tepat Waktu	Tepat Waktu	Benar
Data Uji 6	Tepat Waktu	Tepat Waktu	Benar
Data Uji 7	Tepat Waktu	Lulus Cepat	Salah
Data Uji 8	Tepat Waktu	Tepat Waktu	Benar
.....	....	..	....
Data Uji 25	Tepat Waktu	Lulus Cepat	Salah
Data Uji 26	Tepat Waktu	Tepat Waktu	Benar
Data Uji 27	Terlambat	Terlambat	Benar
Data Uji 28	Terlambat	Terlambat	Benar
Data Uji 29	Terlambat	Terlambat	Benar
Data Uji 30	Tepat Waktu	Tepat Waktu	Benar

Pada tabel VIII telah diketahui hasil akurasi klasifikasi kelulusan mahasiswa menggunakan metode Naïve Bayes Classifier. Untuk tingkat akurasi hasil klasifikasi Algoritma C4.5 dapat dilihat pada Tabel IX

TABLE IX  
 TABEL HASIL KLASIFIKASI KELULUSAN NAIVEBAYES CLASSIFIER

Nama	Kelulusan	Kelulusan Berdasarkan Metode	Akurasi
Data Uji 1	Tepat Waktu	Tepat Waktu	Benar
Data Uji 2	Tepat Waktu	Tepat Waktu	Benar
Data Uji 3	Tepat Waktu	Lulus Cepat	Salah
Data Uji 4	Tepat Waktu	Tepat Waktu	Benar
Data Uji 5	Tepat Waktu	Tepat Waktu	Benar
Data Uji 6	Tepat Waktu	Tepat Waktu	Salah
Data Uji 7	Tepat Waktu	Lulus Cepat	Benar
Data Uji 8	Tepat Waktu	Tepat Waktu	Salah
.....	....	..	....
Data Uji 25	Tepat Waktu	Lulus Cepat	Salah
Data Uji 26	Tepat Waktu	Tepat Waktu	Benar
Data Uji 27	Terlambat	Terlambat	Benar
Data Uji 28	Terlambat	Terlambat	Benar
Data Uji 29	Terlambat	Terlambat	Benar
Data Uji 30	Tepat Waktu	Tepat Waktu	Benar

Dari Tabel 3.9 diketahui

Jumlah data yang diprediksi dengan Benar = 23 Data

Jumlah data yang diprediksi Salah = 7 Data

Selanjutnya, lakukan perhitungan tingkat akurasi berdasarkan rumus.

$$Akurasi = \frac{23 \text{ Data}}{30 \text{ Data}} \times 100 \%$$

Akurasi = 76.67%

Setelah dilakukan pengujian terhadap 30 data uji yang sama serta dilakukan uji dengan metode Naïve Bayes Classifier dan Algoritma C4.5, didapatkan hasil bahwa tingkat akurasi Naïve Bayes Classifier lebih besar dibandingkan dengan tingkat akurasi Algoritma C4.5. Tingkat akurasi metode Naïve Bayes Classifier adalah sebesar 83.33% sedangkan tingkat akurasi Algoritma C4.5 adalah sebesar 76.67%.

Data uji yang digunakan pada pengujian ini adalah data yang sama yang memiliki jumlah kriteria yang sama. Dapat disimpulkan untuk kasus klasifikasi kelulusan mahasiswa metode dengan tingkat akurasi yang lebih baik adalah Naïve Bayes Classifier.

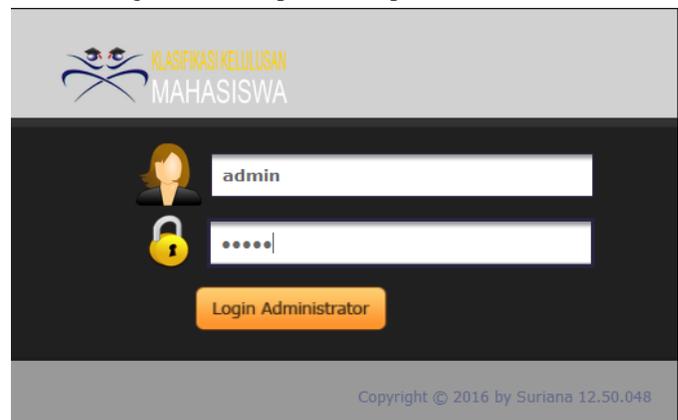
B. Uji Coba

Form Menu Utama Pengunjung dapat dilihat pada Gambar 3.1



Gambar 3.1 Form Utama Pengunjung

Form Login Admin dapat dilihat pada Gambar 3.2



Gambar 3.2 Form Login Admin

Setelah Admin berhasil melakukan login pada Form Login Admin, maka web akan mengalihkan Admin pada Menu Utama Admin. Form Menu Utama Admin dapat dilihat pada Gambar 3.3



Gambar 3.3 Form Menu Utama Admin

Saat Admin memilih Sub Menu Data Training, maka akan tampil Form Data Training. Admin dapat melakukan entri data latih baru, dengan menekan tombol "Tambah Data Training". Ketika admin tidak mengisi form dengan lengkap, maka akan muncul pesan "Isi Semua Kolom". Ketika admin menekan tombol "Simpan", maka data latih akan tersimpan dan muncul pesan "Data Berhasil Disimpan". Tombol "Edit" untuk mengubah data latih yang telah tersimpan dan akan muncul pesan "Data Berhasil Diubah" saat data berhasil diubah.

Tombol “Hapus” untuk menghapus data latih. Tampilan Form Data Training dapat dilihat pada Gambar 3.4

No	NIM	Nama	IPK	SKS	Status Sipl	Status Pekerjaan	Waktu Kuliah	Pernah Cuti / Tidak	Kelulusan	Aksi
1	07.50.107	Data Training 1	C	148	M	B	M	P	Lulus Cepat	[EDIT] [HAPUS]
2	08.50.048	Data Training 2	B	146	BH	TB	M	T	Terlambat	[EDIT] [HAPUS]
3	08.50.076	Data Training 3	B	147	BH	B	M	T	Tepat Waktu	[EDIT] [HAPUS]
4	08.50.105	Data Training 4	B	147	M	B	P	T	Terlambat	[EDIT] [HAPUS]
5	08.50.114	Data Training 5	C	144	BH	B	M	T	Terlambat	[EDIT] [HAPUS]
6	09.50.003	Data Training 6	C	150	BH	TB	P	T	Terlambat	[EDIT] [HAPUS]

Gambar 3.4 Uji Coba Program Data Training

Saat Admin memilih Sub Menu Data Testing, maka akan tampil Form Data Testing. Admin dapat melakukan entri data uji baru, dengan mengisi form yang tersedia. Ketika admin tidak mengisi form dengan lengkap, maka akan muncul pesan “Isi Semua Kolom”. Ketika admin menekan tombol “Simpan”, maka data uji akan tersimpan dan muncul pesan “Data Berhasil Disimpan”. Tombol “Edit” untuk mengubah data uji yang telah tersimpan dan akan muncul pesan “Data Berhasil Diubah” saat data berhasil diubah. Tombol “Hapus” untuk menghapus data uji. Tampilan Form Data Testing dapat dilihat pada Gambar 3.5

No	NIM	Nama	IPK	SKS	Status Sipl	Status Pekerjaan	Waktu Kuliah	Pernah Cuti / Tidak	Aksi
1	13.50.195	Data Training 1	B	147	BH	B	M	T	[EDIT] [HAPUS]
2	12.50.177	Data Training 2	M	145	M	B	P	T	[EDIT] [HAPUS]
3	12.50.172	Data Training 3	B	150	BH	B	M	T	[EDIT] [HAPUS]
4	12.50.153	Data Training 4	M	147	BH	B	P	T	[EDIT] [HAPUS]
5	12.50.124	Data Training 5	B	150	BH	TB	P	T	[EDIT] [HAPUS]

Gambar 3.5 Uji Coba Program Data Testing

Saat Admin memilih Sub Menu Proses Naive Bayes Classifier, maka akan tampil Form Proses Naive Bayes Classifier. Admin dapat melakukan klasifikasi kelulusan mahasiswa pada form ini. Langkah pertama ialah menekan tombol “Proses” pada Tahap 1. Maka akan muncul jumlah data per kelas Kelulusan. Setelah itu jika ingin melanjutkan ke tahap berikutnya maka tekan tombol “Lanjut Ke Tahap 2”. Setelah melakukan proses pada Tahap 2 maka jika ingin melanjutkan, pilih “Klasifikasi Kelulusan”. Maka akan muncul hasil klasifikasi kelulusan berdasarkan metode Naive Bayes

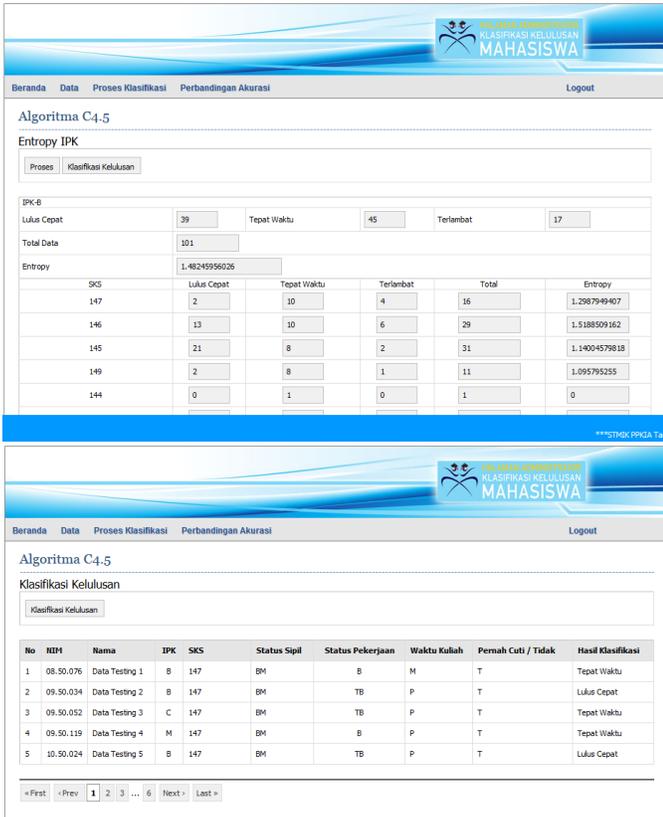
Classifier. Tampilan Form Proses Naive Bayes Classifier dapat dilihat pada Gambar 3.6

No	NIM	Nama	IPK	SKS	Status Sipl	Status Pekerjaan	Waktu Kuliah	Pernah Cuti / Tidak	Hasil Kelulusan
1	08.50.076	Data Testng 1	B	147	BH	B	M	T	Lulus Cepat
2	09.50.034	Data Testng 2	B	147	BH	TB	P	T	Lulus Cepat
3	09.50.052	Data Testng 3	C	147	BH	TB	P	T	Lulus Cepat

Gambar 3.6 Uji Coba Program Proses Naive Bayes Classifier

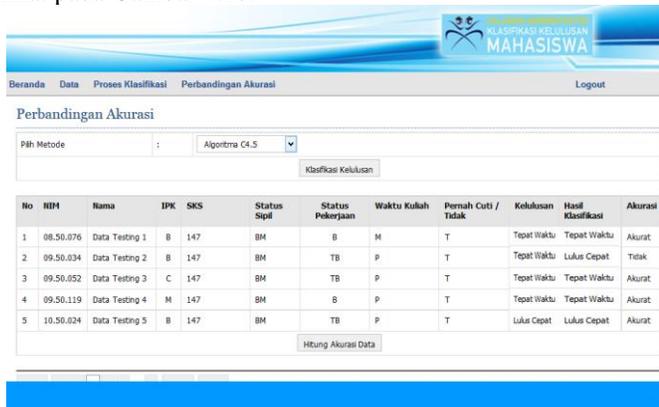
Saat Admin memilih Sub Menu Proses Algoritma C4.5, maka akan tampil Form Proses Algoritma C4.5. Admin dapat melakukan klasifikasi kelulusan mahasiswa pada form ini. Langkah pertama ialah menekan tombol “Proses” pada tahap Entropy. Maka akan muncul jumlah data per kriteria per kelulusan, total data, entropy masing-masing kriteria, dan gain. Setelah mendapat gain terbesar, jika admin ingin melanjutkan proses, maka pilih “Entropy Selanjutnya” dan “Proses”. Jika ingin melakukan klasifikasi kelulusan, pilih “Klasifikasi Kelulusan”. Maka akan muncul hasil klasifikasi kelulusan berdasarkan metode Algoritma C4.5. Tampilan Form Proses Algoritma C4.5 dapat dilihat pada Gambar 3.7.

IPK	Lulus Cepat	Tepat Waktu	Terlambat	Total	Entropy
3.51 - 4.00	11	5	0	16	0
3.01 - 3.50	39	45	17	101	1.48249956026
2.51 - 3.00	3	12	18	33	1.32217934552
0 - 2.50	0	0	0	0	0



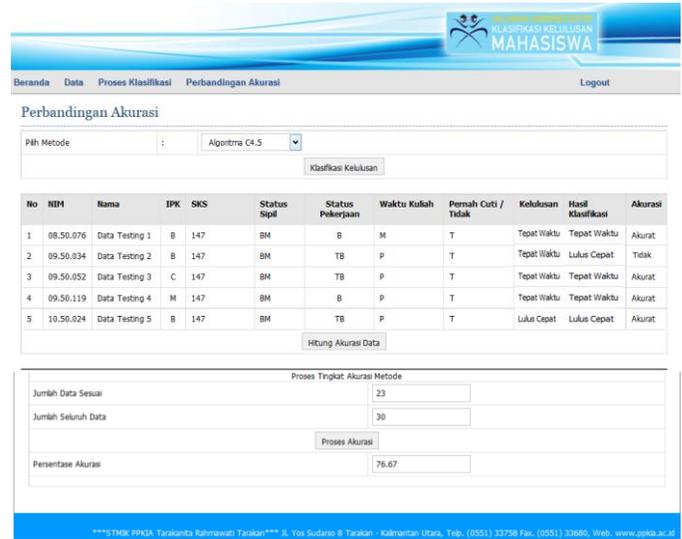
Gambar 3.7 Proses Algoritma C4.5

Saat Admin memilih Sub Menu Perbandingan Akurasi, maka akan tampil Form Perbandingan Akurasi. Admin dapat melakukan uji akurasi terhadap 2 metode yang ada. Admin dapat memilih metode yang ingin diuji, setelah itu pilih “Klasifikasi Kelulusan”. Setelah itu tekan tombol “Hitung Akurasi Data” untuk menghitung jumlah data yang diprediksi dengan benar serta menghitung persentase akurasi data dari metode tersebut. Tampilan Form Perbandingan Akurasi dapat dilihat pada Gambar 4.17.



Gambar 3.8 Uji Coba Program Perbandingan Akurasi

Hal terpenting dari perangkat lunak ini adalah bagaimana sistem melakukan perhitungan klasifikasi kelulusan mahasiswa berdasarkan Metode Naïve Bayes Classifier ataupun Algoritma C4.5.



Gambar 3.9 Hasil Perhitungan Akurasi Metode Algoritma C4.5

Dari hasil implementasi metode Naïve Bayes Classifier dan Algoritma C4.5, serta berdasarkan hasil uji akurasi yang dilakukan sebanyak 5 kali dengan jumlah data yang berbeda pada hasil klasifikasi masing-masing metode, maka diperoleh hasil sebagai berikut:

TABLE 3.  
 TABEL HASIL UJI AKURASI

Metode	Jumlah Data	Akurasi	Jumlah Lulus Cepat	Jumlah Tepat Waktu	Jumlah Terlambat
NBC	30	83.33 %	4	17	4
Algoritma C4.5		76.67 %	3	16	4
NBC	40	60 %	10	12	2
Algoritma C4.5		70 %	14	12	2
NBC	55	61.82 %	14	14	6
Algoritma C4.5		54.55 %	11	14	5
NBC	75	69.33 %	17	23	12
Algoritma C4.5		70.67 %	17	24	12
NBC	100	68 %	21	28	19
Algoritma C4.5		75 %	23	33	19

Berdasarkan Tabel 4.1, diperoleh hasil analisa sebagai berikut:

1. Hasil Analisa Rata-Rata Akurasi Metode Naïve Bayes Classifier  
 $(83.33 \% + 60 \% + 61.82 \% + 69.33 \% + 68 \%) / 5 = 68.5 \%$
2. Hasil Analisa Rata-Rata Akurasi Algoritma C4.5  
 $(76.67 \% + 70 \% + 54.55 \% + 70.67 \% + 75 \%) / 5 = 69.4 \%$
3. Hasil Analisa Kelulusan dengan jumlah mahasiswa terbesar berdasarkan klasifikasi dari metode Naïve Bayes Classifier dan Algoritma C4.5 yaitu, TEPAT WAKTU.
4. Untuk kasus sejenis metode dengan tingkat akurasi yang lebih baik yaitu, Agoritma C4.5.

#### IV. KESIMPULAN

Dari hasil penelitian dan analisis penerapan Metode *Naïve Bayes Classifier* dan Algoritma C4.5 dalam Klasifikasi Kelulusan Mahasiswa pada STMIK PPKIA Tarakanita Rahmawati Tarakan dapat disimpulkan sebagai berikut:

1. Dengan adanya perangkat lunak ini, kelulusan mahasiswa telah dapat diprediksi atau dikelompokkan. Prediksi atau klasifikasi kelulusan tidak lagi berdasarkan penilaian subjektif dosen melainkan berdasarkan pada kriteria atau parameter, yaitu : IPK (Indeks Prestasi Kumulatif), SKS (Satuan Kredit Semester), Status Sipil, Status Pekerjaan, Waktu Kuliah, dan Status Keaktifan (Pernah Cuti atau Tidaknya Mahasiswa);
2. Proses klasifikasi berdasarkan pola yang terbentuk dari data training memungkinkan masih adanya data yang klasifikasinya tidak bernilai sama dengan data asli. Maka dari itu jumlah data training dengan berbagai variasi data sangat membantu dalam proses klasifikasi;
3. Sistem penilaian ini dapat membantu bagian Program Studi atau bagian Akademik dalam memprediksi kelulusan dan tingkat kelulusan Mahasiswa, serta dapat memprediksi secara dini jumlah Wisudawan/Wisudawati;
4. Dari 150 data Mahasiswa yang digunakan sebagai data latih atau data training, dan berdasarkan hasil uji akurasi yang dilakukan, diperoleh hasil sebagai berikut :
  - a. Tingkat Akurasi metode *Naïve Bayes Classifier* : 68.5 %
  - b. Tingkat Akurasi Algoritma C4.5 : 69.4 %
  - c. Tingkat Kelulusan tertinggi berdasarkan hasil klasifikasi ialah Tepat Waktu.

#### REFERENSI

- [1] Badan Pengembangan dan Pembinaan Bahasa, *Kamus Besar Bahasa Indonesia*, (Edisi keempat; Jakarta: Gramedia, 2008)
- [2] Fajar Astuti Hermawati, *Data Mining*, (Cet. I; Yogyakarta: ANDI, 2013) hal. 14
- [3] Prabowo, Rahmadya Trias, *Penerapan Data Mining Dengan Matlab*, (Cet. I; Bandung: Rekayasa Sains, 2013) hal. 5-8
- [4] Badan Pengembangan dan Pembinaan Bahasa, *Kamus Besar Bahasa Indonesia*, (Edisi keempat; Jakarta: Gramedia, 2008)
- [5] Republik Indonesia, *Peraturan Pemerintah RI No. 30 Tahun 1990 Tentang Pendidikan Tinggi*, Bab 1, pasal 1 ayat 6, hal. 1
- [6] STMIK PPKIA Tarakanita Rahmawati, *Prosedur Mutu Evaluasi Kelulusan* (PR/10/WK1), 2012 hal. 2-6
- [7] Oded Z. Maimon, Lior Rokach, *Decomposition methodology for knowledge discovery and data mining : theory and applications*, 2005
- [8] Kusriani, Emha Taufiq, *Algoritma Data Mining*, (Cet. I; Yogyakarta: ANDI, 2009) hal. 189
- [9] Larose, Daniel, *Discovering Knowledge in Data: An Introduction to Data Mining* (John Wiley & Sons, Inc, 2005)
- [10] Berry, Linnoff, *Data Mining Techniques for Marketing, Sales, Customer Relationship Management* (Cet. 2; Wiley Publishing, Inc, 2004)
- [11] Kusriani, Emha Taufiq, *Algoritma Data Mining* (Cet. I; Yogyakarta: ANDI, 2009) hal. 14
- [12] Kusriani, Emha Taufiq, *Algoritma Data Mining* (Cet. I; Yogyakarta: ANDI, 2009) hal. 16

**Penulis I**, Suriana, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2016.

**Penulis II**, Evi Dianti Bintari, memperoleh gelar Magister Komputer (M.Kom), Sekolah Tinggi Teknik Surabaya. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati

**Penulis III**, Deby Kurniawan, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2012. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati