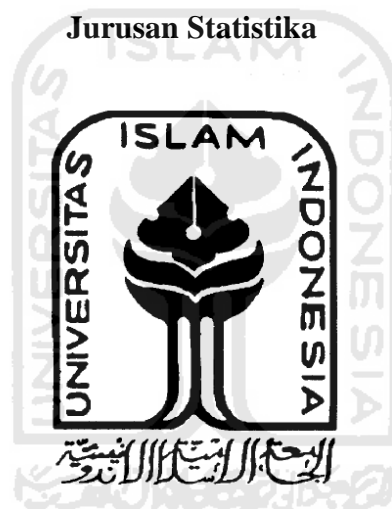


**IMPLEMENTASI ALGORITMA NAÏVE BAYES DAN  
REGRESI LOGISTIK DALAM KLASIFIKASI  
TEPAT STUDI MAHASISWA STATISTIKA  
UNIVERSITAS ISLAM INDONESIA**

(Studi Kasus : Data Akademik Alumni dan Mahasiswa Program Studi Statistika  
UII Angkatan 2012)

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana**



Din Waikabu

12611071

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2016**

**IMPLEMENTATION NAIVE BAYES ALGORITHM AND  
LOGISTIC REGRESSION FOR CLASSIFICATION  
PRECISE STUDIES STUDENT SATATISTICS  
ISLAMIC UNIVERSITY OF INDONESIA**

**ABSTRACT**

*Precise studies is college students graduated status to finish their study in college. normally, to have Bachelor's degree need 4 years, but there people that finish their studies more than 4 years. it will effect the value of accreditation. in this research 6 variable will be used, that is grade point average (GPA), Concentration in High School, Sex, participation in assistance, curriculum, and city of residence, which areclassified by college student precise studies status more than 4 years and less than equal to 4 years. method that wickl be used to classified it is naïve bayes algorithm and will be compared with logistic regression method. based on research results, naïve Bayes algorithm is more superior than logistic regression to classified precise studies college student in statistics UII 2012 with accuracy of 75.86%.*

**Keywords:** *Logistic Regression, Naive Bayes, Precise Studies.*

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 1.1 Grafik Persentase Tepat Studi Prodi Statistika Tahun1995-2011 .....	2
Gambar 4.1 Diagram Alur Penelitian.....	27
Gambar 5.1 Grafik Persentase Tepat Studi .....	28
Gambar 5.2 Grafik Persentase Tepat Studi Dengan Jenis Kelamin .....	29
Gambar 5.3 Grafik Persentase Tepat Studi Dengan Asal Daerah .....	30
Gambar 5.4 Grafik Persentase Tepat Studi Dengan Jurusan Sekolah .....	31
Gambar 5.5 Grafik Persentase Tepat Studi Dengan Asisten .....	31
Gambar 5.6 Grafik Persentase Tepat Studi Dengan IPK .....	32
Gambar 5.6 Grafik Persentase Tepat Studi Dengan Kurikulum .....	33



## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PERSETUJUAN PEMBIMBING</b> .....	ii
<b>HALAMAN PENGESAHAN TUGAS AKHIR</b> .....	iii
<b>KATA PENGANTAR</b> .....	iv
<b>DAFTAR ISI</b> .....	vi
<b>DAFTAR TABEL</b> .....	x
<b>DAFTAR GAMBAR</b> .....	xi
<b>DAFTAR LAMPIRAN</b> .....	xii
<b>PERNYATAAN</b> .....	xiii
<b>INTISARI</b> .....	xiv
<b>ABSTRACT</b> .....	xv
<b>BAB I PENDAHULUAN</b> .....	1
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah .....	5
1.4. Hipotesis .....	5
1.5. Jenis Penelitian dan Metode Analisis .....	7
1.6. Tujuan Penelitian.....	7
1.7. Manfaat Penelitian.....	7
<b>BAB II TINJAUAN PUSTAKA</b> .....	9
2.1. Penelitian Terdahulu .....	9
2.1.1. Algoritma Naïve Bayes.....	9

2.1.2. Regresi Logistik.....	11
<b>BAB III LANDASAN TEORI.....</b>	<b>13</b>
3.1. Regresi Logoistik .....	13
3.1.1. <i>Maximum Likelihood Estimator</i> .....	14
3.1.2. Pengujian Signifikansi Regresi.....	16
3.1.2.1. Uji Rasio <i>Likelihood</i> .....	16
3.1.2.2. Uji <i>wald</i> .....	16
3.1.3. <i>Odds Ratio</i> .....	17
3.2. Algoritma Naïve Bayes .....	17
3.3. Ketepatan Klasifikasi.....	20
3.4. Masa Studi .....	20
3.4.1. Tepat Studi.....	21
3.4.2. Kelulusan .....	21
<b>BAB IV METODOLOGI PENELITIAN .....</b>	<b>22</b>
4.1. Sumber Data Penelitian .....	22
4.2. Variabel Penelitian .....	22
4.2.1. Definisi Variabel Operasional.....	22
4.3. Metode dan Alat .....	24
4.4. Tahapan Penelitian .....	25
4.5. Diagram Alur Penelitian.....	27
<b>BAB V HASIL DAN PEMBAHASAN.....</b>	<b>28</b>
5.1. Analsis Deskriptif .....	28
5.1.1. Profil Lulusan Statistika UII Berdasarkan Tepat Studi.....	28

5.1.2.	Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Jenis Kelamin.....	29
5.1.3.	Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Asal Daerah.....	29
5.1.4.	Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Jurusan Sekolah.....	30
5.1.5.	Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Asisten.....	31
5.1.6.	Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan IPK .....	32
5.1.7.	Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Kurikulum .....	32
5.2.	Pemodelan Data Akademik Alumni Statistika UII dan Mahasiswa Statistika UII Angkatan 2012 Dengan Regresi Logistik.....	33
5.2.1.	Uji Secara Bersama.....	33
5.2.2.	Uji <i>Wald</i> .....	34
5.2.3.	<i>Odds Ratio</i> .....	36
5.2.4.	Prediksi Tepat Studi .....	37
5.2.5.	Ketepatan Klasifikasi .....	37
5.3.	Pemodelan Data Akademik Alumni Statistika UII dan Mahasiswa Statistika UII Angkatan 2012 Dengan Naïve Bayes.....	38
5.3.1.	Pengujian Data <i>Training</i> .....	38
5.3.2.	Pengujian Data <i>Testing</i> .....	42
5.3.3.	Ketepatan Klsifikasi .....	44
5.4.	Perbandingan Ketepatan Klasifikasi.....	45

<b>BAB VI KESIMPULAN DAN SARAN</b> .....	46
6.1. Kesimpulan.....	46
6.2. Saran.....	47
<b>DAFTAR PUSTAKA</b> .....	48
<b>LAMPIRAN</b> .....	50



## DAFTAR LAMPIRAN

Lampiran 1 Data *Training*

Lampiran 2 Data *Testing*

Lampiran 3 Model Naïve Bayes

Lampiran 4 Tabel Prediksi Data *Training*

Lampiran 5 Hasil Prediksi Data *Testing*

Lampiran 6 Regresi Logistik





## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 4.1. Variabel Penelitian Untuk Naïve Bayes .....	23
Tabel 4.2. Variabel Penelitian Untuk Regresi Logistik .....	24
Tabel 5.1. Statistik Uji G .....	34
Tabel 5.2. Statistik Uji Wald .....	34
Tabel 5.3. Estimasi Probabilitas Regresi Logistik .....	35
Tabel 5.4. <i>Odds Ratio</i> .....	36
Tabel 5.5. Hasil Prediksi Tepat Studi .....	37
Tabel 5.6. Ketepatan Klasifikasi .....	38
Tabel 5.7. <i>Data Training</i> .....	38
Tabel 5.8. Probabilitas Tepat Studi .....	39
Tabel 5.9. Probabilitas Tepat Studi Dengan Jenis Kelamin.....	39
Tabel 5.10. Probabilitas Tepat Studi Dengan Asal Daerah.....	40
Tabel 5.11. Probabilitas Tepat Studi Dengan Jurusan Sekolah .....	40
Tabel 5.12. Probabilitas Tepat Studi Dengan Asisten .....	41
Tabel 5.13. Probabilitas Tepat Studi Dengan Kurikulum .....	41
Tabel 5.14. Probabilitas Tepat Studi Dengan IPK .....	42
Tabel 5.16. <i>Data Testing</i> .....	42
Tabel 5.16. Hasil Prediksi Tepat Studi Mahasiswa Statistika UII Angkatan 2012 .....	43
Tabel 5.17. Ketepatan Klasifikasi .....	44
Tabel 5.18. Perbandingan Ketepatan Klasifikasi .....	45

**IMPLEMENTASI ALGORITMA NAÏVE BAYES DAN  
REGRESI LOGISTIK DALAM KLASIFIKASI  
TEPAT STUDI MAHASISWA STATISTIKA  
UNIVERSITAS ISLAM INDONESIA**

OLEH : Din Waikabu

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

**INTISARI**

*Tepat studi merupakan status lulus mahasiswa dalam menyelesaikan studinya di perguruan tinggi. Dalam mencapai gelar sarjana dibutuhkan waktu normal yaitu selama empat tahun, tetapi masih ada mahasiswa yang menyelesaikan studinya melebihi batas normal (lebih dari empat tahun). Hal tersebut akan berdampak pada nilai akreditasi program studi. Dalam penelitian ini digunakan 6 variabel yaitu Indeks Prestasi Kelulusan (IPK), Jurusan sekolah, Jenis kelamin, keikutsertaan menjadi asisten, kurikulum, dan asal daerah, yang diklasifikasikan berdasarkan status tepat studi mahasiswa lebih dari empat tahun dan kurang dari sama dengan empat tahun. Metode yang digunakan untuk klasifikasi tepat studi mahasiswa adalah algoritma naïve bayes dan akan dibandingkan dengan metode regresi logistik. Berdasarkan hasil penelitian didapatkan algoritma naïve bayes lebih unggul dari regresi logistik dalam pengklasifikasian tepat studi mahasiswa statistika UII angkatan 2012 dengan tingkat akurasi sebesar 75.86%.*

**Kata Kunci:** *RegresiLogistik, Naïve Bayes, Tepat Studi. Klasifikasi*

## KATA PENGANTAR



Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya berupa keimanan, kekuatan, kesabaran, kelancaran serta keselamatan sehingga penyusunan tugas akhir ini dapat terselesaikan. Shalawat serta salam tercurah kepada junjungan kita Nabi Muhammad SAW beserta keluarga dan para pengikut-pengikutnya sampai akhir zaman.

Tugas akhir ini tersusun sebagai hasil proses pembelajaran yang telah penulis dapatkan selama melakukan proses pembelajaran di Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia. Tugas akhir ini berisi tentang “Implementasi Algoritma Naïve Bayes Dan Regresi Logistik Dalam Klasifikasi Lama Studi Mahasiswa Statistika Univesitas Islam Indonesia”. Selama proses menyusun tugas akhir ini, penulis telah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis bermaksud menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Kedua orang tua bapak Ismail Waikabu, ibu Salima, serta keluarga yang sangat penulis cintai, yang turut memberi semangat dan sumber motivasi bagi penulis dalam penyusunan tugas ini.
2. Bapak Drs. Allwar, M,Sc, Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. .
3. Bapak Dr. Raden Bagus Fajriya Hakim, M.Si., selaku Ketua Jurusan Statistika, Universitas Islam Indonesia
4. Ibu Ayundyah Kesumawati, S.Si, M.Si., selaku dosen pembimbing yang selalu mengingatkan serta membimbing penulis untuk menyelesaikan tugas akhir ini.

5. Keluarga besar Statistika FMIPA UII.
6. Galih Ramadona, M Ulinuha, Asep A. Yakin, Septu Kornianto, Syaisar, Ibrahim, dan seluruh teman-teman saya yang selalu memberikan semangat, ejekan, dan dukungan dalam menyelesaikan tugas akhir ini.
7. Semua pihak yang tidak dapat penulis sebutkan satu per satu, terima kasih atas segala bantuannya.

Penulis menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna, oleh karena itu segala kritik dan saran yang sifatnya membangun selalu penulis harapkan. Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan umumnya. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal ‘alamiin

Yogyakarta, 5 Oktober 2016

Penulis

**HALAMAN PERSETUJUAN PEMBIMBING  
TUGAS AKHIR**

Judul : Implementasi Algoritma Naïve Bayes dan Regresi Logistik dalam kalsifikasi Tepat Studi Mahasiswa Statistika UII ( Studi Kasus Data Akademik Alumni Dan Mahasiswa Prodi Statistika UII Angkatan 2012)

Nama Mahasiswa : Din Waikabu

Nomor Induk Mahasiswa : 12611071

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK  
DIUJIKAN**

Yogyakarta, 3 Oktober 2016

Mengetahui

Dosen Pembimbing



(Ayundyah Kesumawati, S.Si, M.Si. )

**HALAMAN PENGESAHAN**  
**TUGAS AKHIR**

Implementasi Algoritma Naïve Bayes dan Regresi Logistik dalam Klasifikasi  
Tepat Studi Mahasiswa Statistika UII

(Studi Kasus Data Akademik Alumni Dan Mahasiswa Prodi Statistika UII  
Angkatan 2012)

Nama Mahasiswa : Din Waikabu

Nomor Mahasiswa : 12611071

TUGAS AKHIR INI TELAH DIAJUKAN

PADA TANGGAL 20 JULI 2016

Nama Penguji :

Tanda Tangan

1. Ir. Ali Parkhan, M.T

2. Dr. Jaka Nugraha, S.Si, M.Si.

3. Ayundyah Kesumawati, S.Si, M.Si



Mengetahui

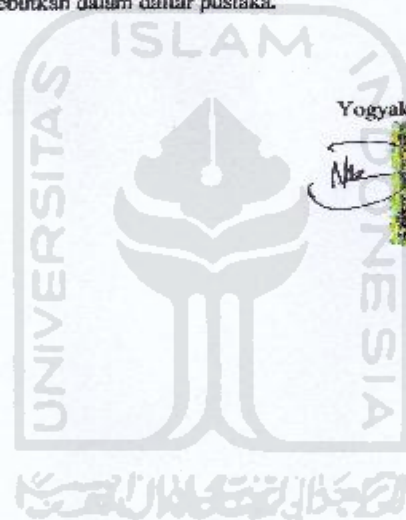
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



  
Drs. Allwar. M.Sc., Ph.D

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang di acu dalam naskah ini dan disebutkan dalam daftar pustaka.



Yogyakarta, 1 Oktober 2016



Din Waikahu

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

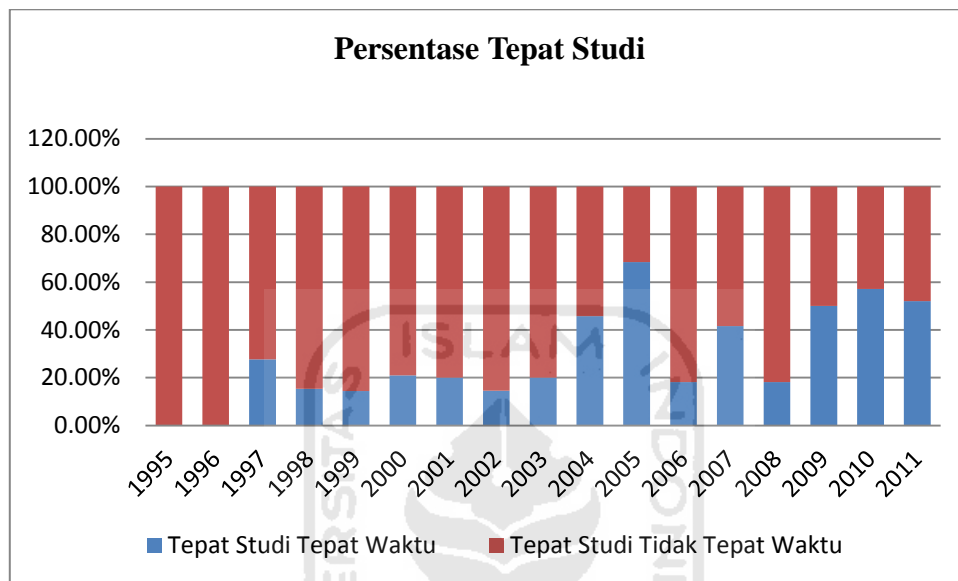
Pendidikan merupakan aspek utama dalam meningkatkan kualitas hidup manusia, dengan pendidikan manusia dapat mengetahui banyak hal dan mengembangkan diri. Pendidikan diadakan untuk menyiapkan manusia demi menunjang kontribusinya dimasa yang akan datang. Berdasarkan Undang-Undang Republik Indonesia no. 20 tahun 2003 tentang sistem pendidikan. Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual, keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara. Pendidikan merupakan sarana untuk mendapatkan suatu ilmu atau pengalaman yang berguna untuk mengembangkan otak bagi orang yang menempuhnya. Dengan demikian pendidikan merupakan sarana terbaik untuk menciptakan generasi baru Indonesia yang tangguh dan hebat dalam hal Intelektual.

Pendidikan merupakan suatu aspek yang sangat penting, dan dapat dilaksanakan oleh siapa saja untuk mencerdaskan dan menciptakan generasi bangsa yang handal dan tangguh, salah satunya adalah perguruan tinggi. Perguruan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi. Pendidikan tinggi meliputi program diploma, sarjana, program profesi, program master, dan doktoral. Perguruan tinggi di Indonesia dibagi menjadi dua yaitu perguruan tinggi negeri (PTN) dan perguruan tinggi swasta (PTS) (PP No 66, 2010).

Universitas Islam Indonesia (UII) merupakan salah satu PTS yang ada di Indonesia. UII memiliki 8 fakultas, dan disetiap fakultas terdapat beberapa jurusan. Jurusan statistika merupakan salah satu jurusan yang berada di Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA).



Setiap tahunnya UII menyelenggarakan upacara wisuda, dan masih ada mahasiswa prodi statistika yang lulus tidak tepat waktu. Persentase tepat studi mahasiswa prodi statistika terlihat seperti gambar 1.1 sebagai berikut:



Gamabar1.1 :Grafik Persentase Tepat Studi Prodi Statistika UII  
Tahun1995-2011

Dari Gambar 1.1, maka dapat dikatakan masih ada mahasiswa prodi statistika dalam menyelesaikan studinya lebih dari waktu normal yaitu 4 tahun. Keadaan ini akan mempengaruhi kualitas prodi statistika, karena sebuah perguruan tinggi khususnya program studi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN PT). Kualitas tersebut diukur berdasarkan 7 standar utama, salah satu-nya adalah mahasiswa dan lulusan. Khusus mengenai evaluasi standar mahasiswa dan lulusan, komponen yang menjadi penilaian adalah indeks prestasi kumulatif dan lama studi (BAN. PT, 2011), berdasarkan matriks penilaian instrument akreditasi program studi BAN PT (2008) bahwa persentase mahasiswa yang lulus tepat waktu merupakan salah satu elemen penilaian akreditasi.

Ada banyak faktor yang dapat menyebabkan seorang mahasiswa lulus tepat waktu. Menurut Diah dan Yuciana (2013) bertambahnya IPK maka peluang seorang mahasiswa untuk lulus lebih dari 4 tahun lebih kecil daripada lulus kurang dari 4 tahun. Tepat studi atau dalam hal ini ketepatan masa studi setiap mahasiswa dapat diprediksi berdasarkan faktor-faktor yang berkaitan dengan latar belakang sekolah sebelumnya dan data akademik serta pribadi saat berada di perguruan tinggi seperti jenis kelamin jurusan sekolah, asal daerah, dan asisten (Jananto, 2013). Hastuti (2012) dalam penelitiannya menyatakan bahwa domisili saat menempuh studi mempengaruhi lama studi. Penelitian yang dilakukan oleh Deker (2009) menyebutkan bahwa kurikulum yang sulit dianggap sebagai salah satu penyebab tingginya jumlah mahasiswa *drop out*.

Hal tersebut menjadi sangat penting bagi manajemen prodi mengingat persentasi mahasiswa lulus tepat waktu adalah salah satu elemen penilaian akreditasi yang ditetapkan oleh BAN-PT. Manajemen memerlukan tindakan untuk mengetahui faktor-faktor penyebab mahasiswa lulus tidak tepat waktu. Oleh karena itu, peneliti ingin meneliti faktor-faktor yang mempengaruhi tepat studi mahasiswa serta ingin mengklasifikasikan kelulusan mahasiswa kedalam dua kategori yaitu tepat waktu untuk mahasiswa yang menempuh pendidikan S1 kurang dari sama dengan 4 tahun (8 semester) dan tidak tepat waktu untuk mahasiswa yang menempuh pendidikan lebih dari 4 tahun (8 semester). Metode klasifikasi yang digunakan dalam metode penelitian ini adalah metode Naïve Bayes dan akan dibandingkan dengan metode Regresi Logistik.

Kusrini dalam Jananto (2013), Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Naïve Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. Naïve Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan kedalam database dengan data yang besar. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain,

diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Ridwandkk, 2013). Shadiq dalam Kartika (2016) teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.

Regresi logistik adalah salah satu model untuk menduga hubungan antara peubah respon kategori dengan satu atau lebih peubah prediktor yang kontinyu atau pun kategori. Peubah respon yang terdiri dari dua kategori yaitu “ya (sukses)” dan “tidak (gagal)”, dandinotasikan  $1=$ ”sukses” dan  $0=$ ”gagal”, maka akan mengikuti sebaran Bernoulli (Agresti, 2002).

Berdasarkan pertimbangan latar belakang diatas maka penulis ingin menerapkan algoritma Naïve Bayes dan metode regresi logistic biner pada data akademik alumni dan mahasiswa angkatan 2012 statistika UII. Oleh karena itu dalam penyusunan tugas akhir ini penulis mengambil judul “Implementasi Algoritma Naïve Bayes dan Regresi Logistik dalam Klasifikasi Tepat Studi Mahasiswa Statistika UII (Studi Kasus Data Akademik Alumni Dan Mahasiswa Angkatan2012 Prodi Statistika)”.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang, maka permasalahan yang dapat diidentifikasi penulis dalam penelitian ini adalah :

1. Bagaimanakah model dan ketepatan klasifikasi tepat studi mahasiswa statistika UII angkatan 2012 dengan menggunakan algoritma naïve bayes dan regresi logistik.
2. Bagaimanakah perbandingan ketepatan klasifikasi tepat studi mahasiswa statistika UII angkatan 2012 dengan menggunakan algoritma Naïve Bayes dan Regersi Logistik.

### 1.3. Batasan Masalah

Agar pembahasan dalam penelitian tidak terlalu meluas, maka dalam penelitian ini diberikan batasan masalah sebagai berikut :

1. Data yang digunakan adalah data akademik alumni statistika UII periode 1995-2011, dan mahasiswa statistika UII angkatan 2012.
2. Metode yang digunakan adalah algoritma naïve bayes dan regresi logistik.

### 1.4. Hipotesis

Berikut ini adalah hipotesis untuk setiap variabel independen yang digunakan dalam penelitian:

#### 1. Jenis Kelamin

Terdapat perbedaan antara laki-laki dan wanita yaitu tingkat ketelitian dan kesabaran laki-laki dan wanita berbeda. Seorang wanita akan cenderung lulus tepat waktu karena wanita lebih teliti dan sabar dalam menjalankan perkuliahan.

#### 2. Asal Daerah

Hastuti (2012) domisili saat menempuh studi mempengaruhi tepat studi. Seorang yang berdomisili di daerah dimana kampus tersebut berada akan lulus tepat waktu.

#### 3. Asisten

Mahasiswa yang menjadi asisten akan lulus tepat waktu, kondisi ini disebabkan karena menjadi seorang asisten merupakan mahasiswa yang mempunyai kemampuan di atas rata-rata. Untuk menjadi seorang asisten harus memenuhi syarat seperti IPK lebih dari 3 dan nilai mata kuliah yang akan di asisten harus A.

#### 4. Jurusan Sekolah

Seorang mahasiswa yang semasa SMA jurusan IPA akan lebih mudah dalam memahami mata kuliah yang diajarkan di perkuliahan, apalagi jurusan statistika merupakan bagian dari ilmu eksak sehingga akan mempengaruhi pencapaian dalam perkuliahan seperti IPK dan akan mempengaruhi tepat studi.

## 5. Kurikulum

Kurikulum berhubungan dengan sistem perkuliahan, kurikulum 2012 merupakan perbaikan dari kurikulum yang sebelumnya sehingga akan lebih baik dan berpengaruh terhadap tepat studi.

## 6. Indeks Prestasi Kumulatif (IPK)

Menurut Diah dan Yuciana (2013) bertambahnya IPK maka peluang seorang mahasiswa untuk lulus lebih dari 4 tahun lebih kecil daripada lulus kurang dari 4 tahun. IPK berhubungan dengan jumlah SKS yang diambil di semester depan sehingga akan berhubungan dengan tepat studi.

### **1.5. Jenis Penelitian dan Metode Analisis**

Tugas akhir ini tergolong dalam kategori aplikasi. Metode analisis yang digunakan adalah algoritma naïve bayes dan regresi logistik. Harapan dari peneliti untuk tugas akhir ini yaitu dapat memprediksi kelulusan mahasiswa statistika UII angkatan 2012.

### **1.6. Tujuan Penelitian**

Terdapat beberapa tujuan yang ingin dicapai dalam penelitian ini, diantaranya yaitu :

1. Mengetahui model ketepatan klasifikasi tepat studi mahasiswa statistika UII angkatan 2012 dengan menggunakan algoritma naïve bayes dan regresi logistik.
2. Mengetahui perbandingan ketepatan klasifikasi antara algoritma naïve bayes dan regresi logistik dalam mengklasifikasi tepat studi mahasiswa statistika UII angkatan 2012.

### 1.7. Manfaat Penelitian

Dengan adanya penelitian ini diharapkan dapat memberikan beberapa manfaat di antaranya yaitu :

1. Memberikan informasi kepada Prodi Statistika UII tentang model algoritma Naïve Bayes dan Regresi Logistik untuk pengklasifikasian tepat studi mahasiswa statistika UII, dan faktor-faktor yang mempengaruhi tepat studi.
2. Penelitian ini dijadikan bahan pertimbangan bagi penelitian selanjutnya khususnya penelitian terkait tepat studi mahasiswa dengan algoritma Naïve Bayes dan Regresi Logistik.
3. Sebagai pengembangan dan pengaplikasian ilmu statistika, khususnya tentang algoritma Naïve Bayes dan Regresi Logistik



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Penelitian Terdahulu**

Sampai saat ini penelitian mengenai klasifikasi tepat studi mahasiswa dengan menggunakan algoritma naïve bayes dan regresi logistik telah banyak dilakukan, diantaranya yaitu :

##### **2.1.1. Algoritma Naïve Bayes**

Murtopo (2016) melakukan penelitian tentang kelulusan tepat waktu mahasiswa STMIK YMI Tegal menggunakan algoritma naïve bayes. Hasil dari penelitian ini menunjukan dimana sebelum didapatkan nilai akurasi dilakukan pengujian dengan memanfaatkan ROC *Curva* dan k-fold *cross validation*, pengujian dilakukan sebanyak 10 *fold*. Dari hasil pengujian didapat nilai akurasi rata-rata sebesar 91,29%, sedangkan nilai akurasi tertinggi dari hasil pengujian 10-fold *cross validation* sebesar 94,34%.

Jananto (2013) melakukan penelitian tentang lama studi mahasiswa dengan menggunakan algoritma naïve bayes. Faktor-faktor yang digunakan adalah nilai-nilai matakuliah yang telah ditempuh hingga semester IV (masa evaluasi tahap I) atau IPK sampai dengan semester IV dan faktor data induk mahasiswa yang berupa jenis kelamin, kota lahir, tipe sekolah, kota sekolah akan menjadi variabel prediktor. Hasil penelitiannya yaitu diperoleh tingkat kesalahan prediksi berkisar 20% sampai dengan 50% dengan data training dan testing yang diambil secara random. Namun rata-rata tingkat kesalahan berkisar 20 % hingga 34%. Tinggi rendahnya tingkat kesalahan dapat disebabkan oleh jumlah *record* data dan tingkat konsistensi dari data training yang digunakan.

Mauriza dan Nugroho (2014) melakukan penelitian tentang implementasi data mining untuk memprediksi kelulusan mahasiswa fakultas komunikasi dan informatika UMS menggunakan metode naïve bayes. Berdasarkan hasil prediksi

dari jumlah 342 data sampel yang diuji hanya 86 mahasiswa yang akan lulus tepat waktu atau hanya sekitar 25,15 %, sedangkan mahasiswa yang akan lulus terlambat berjumlah 256 mahasiswa mencapai 74,85% dari jumlah data sampel.

Meinanda, dkk (2009) dalam penelitiannya tentang prediksi masa studi sarjana dengan *artificial neural network* menyebutkan bahwa masa studi seorang sarjana dipengaruhi oleh IPK, jumlah matakuliah yang diambil, jumlah matakuliah yang mengulang, jumlah matakuliah tertentu. Metode penelitian yang digunakan merupakan adopsi dari CRISP-DM atau *Cross-Industry Standard Process for Data Mining*, dimana di dalamnya terdapat enam fase yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation* dan *Deployment*. Setelah memahami masalahnya kemudian melakukan processing data dengan membuat *crossstabulation*, koreksi data yang *misclassification*, *missing value* maupun *outlier*.

Ridwan, dkk (2013) melakukan penelitian mengenai penerapan data mining untuk evaluasi kinerja akademik mahasiswa menggunakan algoritma naive bayes *classifier*. Hasil Penelitian yang dilakukan difokuskan untuk mengevaluasi kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang dapat lulus tepat waktu atau tidak. Kemudian dari klasifikasi tersebut, sistem akan memberikan rekomendasi solusi untuk memandu mahasiswa lulus dalam waktu yang paling tepat dengan nilai optimal berdasarkan histori nilai yang telah ditempuh mahasiswa. Input dari sistem ini adalah data induk mahasiswa dan data akademik mahasiswa. Sampel mahasiswa angkatan 2005-2009 yang sudah dinyatakan lulus akan digunakan sebagai data training dan testing. Sedangkan data mahasiswa angkatan 2010-2011 dan belum lulus akan digunakan sebagai data target. Data input akan diproses menggunakan teknik data mining algoritma Naïve Bayes *Classifier* (NBC) untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kelulusan mahasiswa. Output dari sistem ini berupa klasifikasi kinerja akademik mahasiswa yang diprediksi kelulusannya dan memberikan rekomendasi untuk proses kelulusan tepat waktu atau lulus dalam waktu yang paling tepat dengan nilai optimal. Hasil pengujian menunjukkan bahwa faktor yang paling berpengaruh dalam penentuan



klasifikasi kinerja akademik mahasiswa yaitu indeks prestasi kumulatif (IPK), indeks prestasi (IP) semester 1, IP semester 4, dan jenis kelamin. Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi. Pengujian pada data mahasiswa angkatan 2005-2009, algoritma NBC menghasilkan nilai precision, recall, dan *accuracy* masing-masing 83%, 50%, dan 70%.

Sari, dkk (2016) melakukan penelitian terkait pembangunan aplikasi klasifikasi mahasiswa baru untuk prediksi hasil studi menggunakan naïve bayes *classifier*. Pada penelitian yang dilakukan, diambil beberapa contoh data mahasiswa pada suatu universitas yaitu, data pada tahun 2010, 2011 dan 2012. Data rata-rata lama studi pada tahun 2010, 2011 dan 2012 masih lebih besar dari empat tahun dan persentase mahasiswa yang menempuh lama studi lebih dari tujuh tahun masih cukup tinggi. Sedangkan dilihat dari rata-rata IPK mahasiswa yang lulus pada tahun 2010, 2011, dan 2012 sudah cukup baik, namun jika ditinjau dari rata-rata IPK seluruh mahasiswa pada ketiga tahun tersebut rata-rata IPK mahasiswa masih perlu ditingkatkan. Dari permasalahan tersebut, maka dibangun aplikasi untuk memprediksi IPK, rata-rata SKS dan persentase kehadiran mahasiswa baru. Aplikasi ini dibangun supaya mahasiswa yang diprediksi mendapatkan IPK atau rata-rata SKS dibawah rata-rata dapat diberikan bimbingan studi untuk meningkatkan IPK atau rata-rata SKS mahasiswa tersebut. Sedangkan mahasiswa yang diprediksi menempuh lama studi lebih dari empat tahun dapat diupayakan untuk dapat menempuh studi kurang dari atau sama dengan empat tahun. Hasil prediksi presentase kehadiran mahasiswa digunakan untuk mengamati keaktifan mahasiswa dalam mengikuti kegiatan belajar mengajar.

### **2.1.2. Regresi Logistik**

Diah dan Yuciana (2013) melakukan penelitian tentang lama studi mahasiswa Universitas Diponegoro menggunakan regresi logistik. Faktor-faktor yang mempengaruhi lama studi mahasiswa FSM UNDIP adalah indeks prestasi kumulatif

(IPK), jurusan dan jalur masuk universitas. hasil penelitian yang diperoleh yaitu IPK, dan jurusan mempengaruhi lama studi dengan ketepatan klasifikasi 80,8%.

Margasari (2014) melakukan penelitian tentang klasifikasi profil mahasiswa FMIPA Universitas Brawijaya, metode yang digunakan adalah regresi logistik, dan metode CART. Hasil penelitian menunjukkan akurasi klasifikasi, metode CART mampu mengklasifikasi data profil mahasiswa FMIPA Universitas Brawijaya sebesar 94.2% sedangkan analisis regresi logistik sebesar 86.7%.

Kundriasworo (2013) melakukan penelitian tentang masa studi mahasiswa Universitas Islam Negeri Sunan Kalijaga, dengan menggunakan metode regresi logistik, dan *Chi-Square Automatic Iteraction Detection* (CHAID). Hasil penelitian menunjukan metode CHAID menyebutkan segmen yang memiliki ketepatan waktu masa studi adalah segmen dimana yang alumninya memiliki  $IPK \geq 3,50$  dan berasal dari fakultas keguruan dan tarbiyah, serta fakultas ushuludian dengan nilai sebesar 93,6%, kemudian segmen yang memiliki ketidak tepatan waktu masa studi terbesar adalah segmen dimana yang alumninya memiliki nilai  $IPK \leq 3,00$  dengan nilai sebesar 98,2%. Ketepatan klasifikasi metode Regresi Logistik 74,7%, sedangkan metode CHAID ketepatan klasifikasinya sebesar 77,4%.

## BAB III

### LANDASAN TEORI

#### 3.1. Regresi Logistik

Regresi logistik adalah regresi yang menggunakan dua nilai yang berbeda untuk menyatakan variabel responnya ( $Y$ ), biasanya digunakan nilai 0 (gagal) dan 1 (sukses). Fungsi distribusi yang digunakan adalah distribusi logistic dengan notasi  $\pi(X)$  untuk menyatakan bersyarat dari  $Y$  jika diberikan vektor kovariate  $X = (x_1, x_2, \dots, x_p)^T$ . Model regresi logistiknya adalah (Nugraha, 2013) :

$$\pi(X) = \frac{\exp(X^T \beta)}{1 + \exp(X^T \beta)} \quad \text{dengan } X^T \beta = \beta_0 + x_1 \beta_1 + \dots + x_p \beta_p \quad (3.1)$$

$\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  adalah vektor parameter. Didefinisikan suatu transformasi *logit*  $\pi(X)$  yaitu (Nugraha, 2013) :

$$g(X) = \log\left(\frac{\pi(X)}{1 - \pi(X)}\right) = X^T \beta \quad (3.2)$$

Sehingga  $g(X)$  linear dalam parameter  $\beta$ .

##### 3.1.1. *Maximum Likelihood Estimator* (MLE)

Misalkan suatu sampel terdiri dari  $n$  observasi dari pasangan  $(X_i, y_i)$ ,  $i = 1, \dots, n$ . model regresi logistic (Nugraha, 2013) :

$$\pi(X_i) = \frac{\exp(X_i^T \beta)}{1 + \exp(X_i^T \beta)} \quad (3.3)$$

Untuk menentukan model regresi, harga  $\beta$  ditaksir lebih dahulu dengan menggunakan metode kemungkinan *maximum* (Nugraha, 2013). Fungsi log likelihood sebagai berikut :

$$\log L(\beta) = \sum_{i=1}^n \{y_i \log(\pi_i) + (n_i - y_i) \log(1 - \pi_i)\} \quad (3.4)$$

Dari fungsi log *likelihood* ini dicari dervatif pertama dan dervatif kedua. Penaksir parameter  $\beta$  merupakan nilai  $\beta$  yang memaksimumkan fungsi log *likelihood* pada data sampel (X,Y) (Nugraha, 2013). nilai maksimum *likelihood* didapat dengan syarat :

$$\frac{\partial \log L(\beta)}{\partial \beta} = 0 \quad (3.5)$$

dan

$$H(\beta) = \frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^T} = 0 \quad (3.6)$$

Matrik  $H(\beta)$  disebut matrik Hessain yang merupakan matrik definet negatif.

Berdasarkan kondisi derivatif pertama, penaksir parameter  $\beta$  menggunakan metode kemungkinan adalah penyelesaian dari persamaan berikut (Nugraha, 2013) :

$$\sum_{i=1}^n [y_i - \pi(X_i)] = 0 \quad (3.7)$$

dan

$$\sum_{j=0}^p \sum_{i=1}^n x_{ij} [y_i - \pi(X_i)] = 0 \quad (3.8)$$

Untuk  $j = 0, 1, \dots, p$ .

Derivatif kedua fungsi log *likelihood* terdapat semua parameternya disebut matrik Hessain (H) yang mempunyai elemen (Nugraha, 2013) :

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (3.9)$$

dan

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_i \partial \beta_u} = - \sum_{i=1}^n x_{ij} x_{iu} \pi_i (1 - \pi_i) \quad (3.10)$$

Misalkan

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

Dan matriks V adalah

$$V = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix} \quad (3.11)$$

$$\hat{\pi}_i = \pi_i(\hat{\beta}) \quad (3.12)$$

$$I(\hat{\beta}) = X^T V X = -H(\hat{\beta}) \quad (3.12)$$

$I(\hat{\beta})$  disebut matriks informasi atau bisa disebut informasi Fisher dan

$$\text{Var}(\hat{\beta}) = [I(\hat{\beta})]^{-1} \quad (3.13)$$

$\text{Var}(\hat{\beta}_j)$  adalah elemen diagonal ke-j (baris ke-j dan kolom ke-j) dari  $\text{Var}(\hat{\beta})$ . Nilai penaksir  $\beta$  dengan menggunakan metode Newton-Rapson pada langkah ke-t sebagai berikut (Nugraha, 2013) :

$$\beta^{(t+1)} = \beta^{(t)} + [-H(\beta^{(t)})]^{-1} [y - \pi^{(t)}] \quad (3.14)$$

dengan

$$\pi_i^t = \frac{\exp(X_i^T \beta^{(t)})}{1 + \exp(X_i^T \beta^{(t)})}, i = 1, \dots, n \quad (3.15)$$

### 3.1.2. Pengujian Signifikansi Parameter Regresi

Untuk menguji signifikansi parameter-parameter yang telah diestimasi dapat digunakan dua uji, yaitu uji rasio *likelihood* dan uji Wald (Hosmer dan Lemeshow, 2000).

#### 3.1.2.1. Uji Rasio *Likelihood*

Untuk mengetahui apakah variabel prediktor mempengaruhi variabel respon secara bersama-sama maka dapat dilakukan uji Rasio *Likelihood*. Pengujian secara simultan dilandaskan pada hipotesis (Hosmer dan Lemeshow, 2000) :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{salah satu dari } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$$

Statistik uji:

$$G = -2 \ln \left[ \frac{l_0}{l_1} \right] \quad (3.16)$$

Kriteria uji : tolak  $H_0$  jika nilai  $G > \chi^2_{(\alpha, p)}$

#### 3.1.2.2. Uji Wald

Uji wald bertujuan untuk mengetahui peranan masing-masing variabel prediktor terhadap variabel respon. Pengujian pada masing-masing variabel didasari oleh hipotesis (Hosmer dan Lemeshow, 2000) :

$$H_0 : \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1 : \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

Statistik Uji :

$$W = \left[ \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \right] \quad (3.17)$$

Kriteri Uji : tolak  $H_0$  jika  $W > \chi^2_{(\alpha, 1)}$

### 3.1.3. Odds Ratio

*Odds ratio* adalah salah satu dari berbagai statistic yang digunakan untuk menilai resiko kejadian tertentu (seperti penyakit) jika suatu faktor tertentu ada. Odds ratio digunakan sebagai statistic deskriptif, dan memainkan peran penting dalam regresi logistik. Odds ratio (OR) merupakan rasio dari dua odds (Nugraha, 2013) :

$$OR = \frac{\pi_1(1 - \pi_2)}{\pi_2(1 - \pi_1)} \quad (3.18)$$

Jika ukuran sampelnya sangat besar maka distribusi dari OR menjadi sangat tidak simetris. Oleh karena itu digunakan transformasi natural logaritma ( $\log$ ) terhadap *odds ratio*, sehingga  $OR = 1$  (independen) adalah ekuivalen dengan  $\log(OR) = 0$ . Menggunakan pendekatan distribusi normal, mean dan deviasi standar untuk  $\log(OR)$  adalah (Nugraha, 2013) :

$$\log(\hat{OR}) = \log(n_{11}) + \log(n_{22}) - \log(n_{12}) - \log(n_{21}) \quad (3.19)$$

$$SE = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \quad (3.20)$$

Nilai SE semakin kecil jika frekuensi masing-masing sel semakin besar. Karena distribusi sampling untuk  $\log(\hat{OR})$  lebih mendekati distribusi normal dibandingkan dengan statistic  $\hat{OR}$ , maka lebih baik menggunakan statistik  $\log(\hat{OR})$  untuk inferensi maupun mendapat estimasi interval. Interval konfidensi  $\log(\hat{OR})$  untuk sampel besar adalah (Nugraha, 2013) :

$$\log(\hat{OR}) \pm z_{\alpha/2}(SE) \quad (3.21)$$

## 3.2. Algoritma Naïve Bayes

Kusrini dalam Jananto (2013), Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naïve Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naïve Bayes terbukti memiliki

akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Ridwan dkk, 2013).

Keuntungan penggunaan Naïve Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naïve Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari, Parveen, 2012).

Teorema Bayes memiliki bentuk umum sebagai berikut (Jananto, 2013) :

$$P(A|B) = \frac{P(X|H).P(H)}{P(X)} \quad (3.22)$$

Keterangan :

- $X$  : Data dengan *class* yang belum diketahui
- $H$  : Hipotesis data  $X$  merupakan suatu *class* spesifik
- $P(H|X)$  : Probabilitas hipotesis  $H$  berdasarkan kondisi  $x$
- $P(H)$  : Probabilitas hipotesis  $H$  (prior prob.)
- $P(X|H)$  : Probabilitas  $X$  berdasarkan kondisi tersebut
- $P(X)$  : Probabilitas dari  $X$

Saleh (2015) untuk menjelaskan metode Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naïve Bayes di atas disesuaikan sebagai berikut:



$$P(C | F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n | C)}{P(F_1 \dots F_n)} \quad (3.23)$$

Di mana Variabel  $C$  merepresentasikan kelas, sementara variabel  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas  $C$  (*Posterior*) adalah peluang munculnya kelas  $C$  (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas  $C$  (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut (Saleh, 2015) :

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}} \quad (3.24)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan  $(C|F_1, \dots, F_n)$  menggunakan aturan perkalian sebagai berikut (Saleh, 2015) :

$$\begin{aligned} P(C | F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n | C) \\ &= P(C)P(F_1 | C)P(F_2, \dots, F_n | C, F_1) \\ &= P(C)P(F_1 | C)P(F_2 | C, F_1)P(F_3, \dots, F_n | C, F_1, F_2) \\ &= P(C)P(F_1 | C)P(F_2 | C, F_1)P(F_3 | C, F_1, F_2)P(F_4, \dots, F_n | C, F_1, F_2, F_3) \\ &= P(C)P(F_1 | C)P(F_2 | C, F_1)P(F_3 | C, F_1, F_2) \dots P(F_n | C, F_1, F_2, F_3, \dots, F_{n-1}) \quad (3.25) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat

tinggi (*naif*), bahwa masing-masing petunjuk ( $F_1, F_2, \dots, F_n$ ) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut (Saleh, 2015) :

$$P(F_i | F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (3.26)$$

Untuk  $i \neq j$ , sehingga :

$$P(F_i | C, F_j) = P(F_i | C) \quad (3.27)$$

Dari persamaan (24) dapat disimpulkan bahwa asumsi independensi tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya penjabaran  $P(C | F_1, \dots, F_n)$  dapat disederhanakan menjadi :

$$\begin{aligned} P(C | F_1, \dots, F_n) &= P(C)P(F_1 | C)P(F_2 | C)P(F_3 | C) \dots \\ &= P(C) \prod_{i=1}^n P(F_i | C) \end{aligned} \quad (3.28)$$

Persamaan (25) merupakan model dari teorema Naïve Bayes yang selanjutnya digunakan dalam proses klasifikasi.

### 3.3. Ketepatan Klasifikasi

Ketepatan Klasifikasi dilakukan untuk melihat hasil yang didapatkan dari klasifikasi. Terdapat beberapa cara untuk mengukur ketepatan klasifikasi, beberapa cara yang sering digunakan adalah dengan menghitung akurasi. Akurasi merupakan persentase dari total dokumen yang teridentifikasi secara tepat dalam proses klasifikasi (Achmad dan Brodjol, 2016). Persamaan untuk mencari ketepatan klasifikasi adalah sebagai berikut :

$$Akurasi = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah data}} * 100\% \quad (3.29)$$

### 3.4. Masa Studi

Masa studi merupakan waktu untuk menyelesaikan studi, masa studi program strata satu adalah 14 semester. Mahasiswa statistika dapat menyelesaikan studinya  $\leq 4$  tahun atau  $> 4$  tahun. Seorang mahasiswa statistika dinyatakan berhasil menyelesaikan studinya atau lulus dan berhak memakai gelar sarjana apabila telah menyelesaikan minimal 144 sks dengan indeks prestasi kumulatif 2.25 (Panduan Akademik Statistika UII, 2012-2013).

#### 3.4.1. Tepat Studi

Tepat studi adalah status lulus mahasiswa dalam menyelesaikan studi dari perguruan tinggi, yang kemudian akan di kategorikan ke dalam dua kategorik yaitu tepat waktu dan tidak tepat waktu. Mahasiswa dikatakan lulus tepat waktu jika masa studi dapat ditempuh kurang sama dengan 4 tahun ( $\leq 4$  tahun). Mahasiswa dikatakan lulus tidak tepat waktu jika masa studi ditempuh lebih dari 4 tahun ( $> 4$  tahun) (Dikiti dan BSNP, 2013).

#### 3.4.2. Kelulusan

Mahasiswa dapat dinyatakan selesai studi atau lulus dari prodi Statistika, apabila memenuhi persyaratan kelulusan, dan diputuskan dalam rapat yudisium akhir studi. Persyaratan lulus prodi Statistika adalah sebagai berikut :

1. Telah dinyatakan lulus tutup teori.
2. Telah menempuh BTAQ dan praktek ibadah, dengan nilai minimum C, dibuktikan dengan sertifikat.
3. Telah menempuh kolokium dengan nilai minimum C.
4. Telah menempuh ujian pendadaran Tugas Akhir dan dinyatakan lulus dengan nilai minimum B/C.
5. Indeks prestasi kumulatif (IPK), minimal 2,25 (dua koma dua lima)
6. Telah menempuh ujian CEPT dan memperoleh skor minimal 422 yang diselenggarakan oleh Cillacs UII.

7. Memenuhi persyaratan administratif yudisium akhir studi lainnya sebagai mana ditentukan oleh fakultas.



## **BAB IV**

### **METODOLOGI PENELITIAN**

#### **4.1. Sumber Data Penelitian**

Penelitian ini dilakukan di prodi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Data yang digunakan adalah data sekunder yaitu data akademik alumni dan mahasiswa aktif angkatan 2012 yang diakses dari bagian akademik dan laboratorium komputasi statistika.

#### **4.2. Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini adalah tepat studi, jenis kelamin, asal daerah, jurusan sekolah, asisten, kurikulum, dan IPK.

##### **4.2.1. Definsi Variabel Operasional**

1. Tepat studi adalah status kelulusan, tepat studi dikategorikan menjadi dua yaitu  $\leq 4$  tahun (Tepat Waktu), dan  $> 4$  tahun (Tidak Tepat Waktu)
2. Jenis kelamin dibagi menjadi dua kategorik yaitu Laki-Laki dan Wanita
3. Asal Daerah adalah kampung halaman mahasiswa statistika UII berasal, asal daerah dibagi menjadi 2 kategorik yaitu dari Yogyakarta dan luar Yogyakarta.
4. Jurusan Sekolah adalah jurusan yang diambil semasa SMA, jurusan sekolah dibagi menjadi tiga kategorik yaitu IPA, IPS, dan Lain-lain
5. Asisten adalah pernah menjadi tutor di laboratorium komputasi statistika UII. Asisten dibagi menjadi dua kategorik yaitu YA dan Tidak
6. IPK merupakan ukuran pencapaian belajar mahasiswa selama kuliah, IPK dibagi menjadi 4 kategorik yaitu cukup, memuaskan, sangat memuaskan, dan dengan pujian.
7. Kurikulum adalah sesuatu yang direncanakan untuk dipelajari oleh mahasiswa yang memuat rencana untuk peserta didik. Kurikulum dibagi menjadi 5 kategorik yaitu kurikulum 1995, kurikulum 1999, kurikulum 2003, kurikulum 2007, dan kurikulum 2012.

Berdasarkan penjelasan diatas maka dapat disimpulkan untuk variabel penelitian sebagai berikut :

Tabel 4.1. Variabel Penelitian Untuk Naïve Bayes

Variabel	Keterangan	Kategori	Keterangan
VariabelDependen	TepatStudi	TepatWaktu	$\leq 4$ tahun
		TidakTepatWaktu	$>4$ Tahun
VariabelIndependen	JenisKelamin	Laki-laki	
		Wanita	
	Asal Daerah	Yogyakarta	
		LuarYogyakarta	
	JurusanSekolah	IPA	
		IPS	
		Lain-Lain	
	Asistensi	Ya	
		Tidak	
	IPK	Cukup	2.00 - 2.75
		Memuaskan	2.76 - 2.99
		SangatMemuaskan	3.00 - 3.50
		DenganPujian	3.51- 4.00
	Kurikulum	Kurikulum 1995	1995-1997
Kurikulum1999		1998-2001	
Kurikulum 2003		2002-2005	
Kurikulum 2007		2006-2010	
Kurikulum 2012		2011- sekarang	

Tabel 4.2. Variabel Penelitian Untuk Regresi Logistik

Variabel	Keterangan	Kategori	Variabel <i>Dummy</i>			
Dependen	Tepat Studi	Tepat Waktu : 1				
		Tidak Tepat Waktu : 0				
Independen	Jenis Kelamin	Laki-Laki : 1				
		Wanita : 0				
	Asal Daerah	Luar Yogyakarta : 1				
		Yogyakarta : 0				
			JS1	JS2		
	Jurusan Sekolah	IPA : 2	1	0		
		IPS : 1	0	1		
		Lain-Lain : 0	0	0		
	Asisten	Ya : 1				
		Tidak : 0				
			IPK1	IPK2	IPK3	
	IPK	Cukup : 0	0	0	0	
		Memuaskan : 1	1	0	0	
		Sangat Memuaskan: 2	0	1	0	
		Dengan Pujian : 3	0	0	1	
			KR1	KR2	KR3	KR4
	Kurikulum	Kurikulum 1995	0	0	0	0
Kurikulum 1999		1	0	0	0	
Kurikulum 2003		0	1	0	0	
Kurikulum 2007		0	0	1	0	
Kurikulum 2012		0	0	0	1	

### 4.3. Metodologi dan Alat

Dalam penelitian ini akan dilakukan klasifikasi tepat studi mahasiswa statistika UII angkatan 2012 dengan menggunakan data profil mahasiswa yang telah menjadi alumni untuk digali informasi dan polanya menggunakan 2 metode klasifikasi yaitu Algoritma Naïve Bayes dan Regresi Logistik. Jumlah data alumni yang digunakan yaitu sebanyak 495 alumni, dan untuk mahasiswa statistika angkatan 2012 digunakan

sebanyak 116 mahasiswa. Setelah didapatkan hasil dari klasifikasinya maka akan dilakukan pemilihan metode klasifikasi terbaik. Alat dalam penelitian ini adalah software R. 3.2.3., dan SPSS 20.

#### 4.4. Tahapan Penelitian

Dalam menyelesaikan penelitian ini adapun langkah-langkah yang harus dilakukan sebagai berikut:

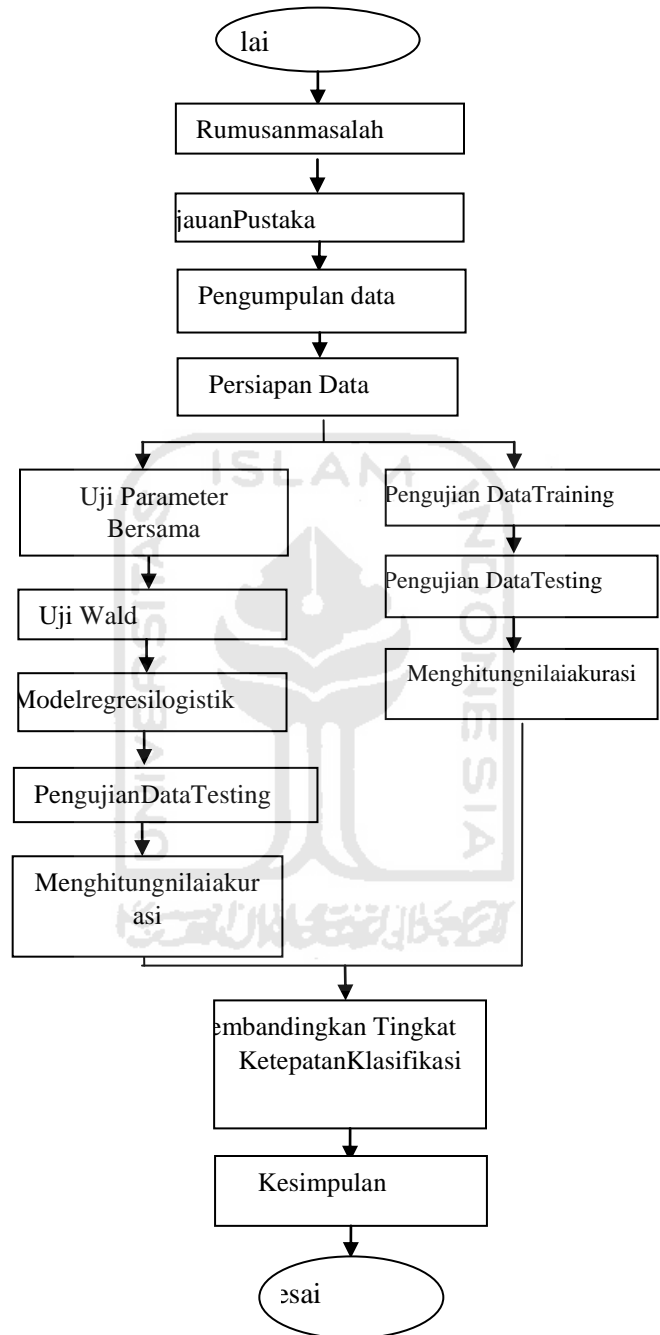
1. Melakukan perumusan masalah.
2. Melakukan kajian pustaka terkait permasalahan.
3. Mengumpulkan data variabel dependen yaitu tepat studi dan variabel independen yaitu : jenis kelamin, asal daerah, jurusan sekolah, asisten, kurikulum, dan indeks prestasi kumulatif.
4. Mempersiapkan data sesuai dengan kebutuhan algoritma Naïve Bayes dan Regresi Logistik, kemudian membagi data menjadi data *training* dan data *testing*.
5. Klasifikasi tepat studi menggunakan Regresi Logistik dengan tahapan :
  - a. Membuat model Regresi Logistik antara variabel independen dan variabel dependen.
  - b. Uji Simultan
  - c. Uji Wald
  - d. Memprediksi tepat studi mahasiswa Statistika UII angkatan 2012.
  - e. Menghitung ketepatan klasifikasi model Regresi Logistik.
6. Klasifikasi tepat studi menggunakan algoritma Naïve Bayes dengan tahapan sebagai berikut :
  - a. Implementasi algoritma Naïve Bayes pada data *training* sehingga mendapatkan model. Setelah itu model yang telah didapat digunakan untuk data *testing*.
  - b. Memprediksi tepat studi mahasiswa jurusan statistika UII angkatan 2012.
  - c. Menghitung ketepatan klasifikasi.



7. Membandingkan tingkat ketepatan klasifikasi dari algoritma Naïve Bayes dan RegresiLogistik.
8. Kesimpulan.
9. Selesai.



#### 4.5. Diagram Alur Penelitian



Gambar 4.1. Diagram Alur Penelitian

## BAB V

### HASIL DAN PEMBAHASAN

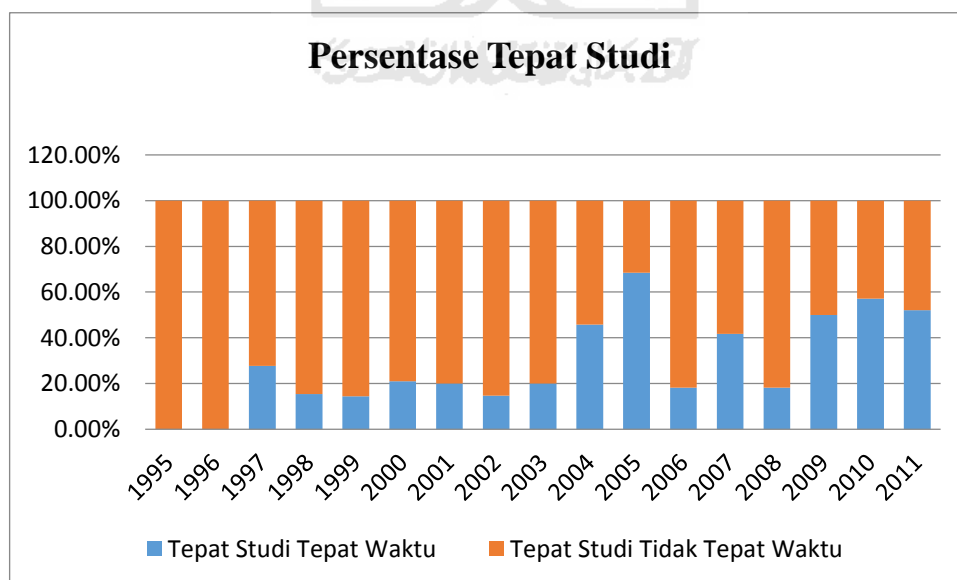
Dalam bab ini akan dijelaskan tentang hasil dari penelitian yang dilakukan, diantaranya seperti analisis deskriptif, analisis regresi logistik, serta algoritma naïve bayes.

#### 5.1. Analisis Deskriptif

Analisis deskriptif bertujuan untuk memberikan gambaran umum terkait profil tepat studi mahasiswa prodi statistika UII angkatan 1995-2011.

##### 5.1.1. Profil Lulusan Statistika UII Berdasarkan Tepat Studi

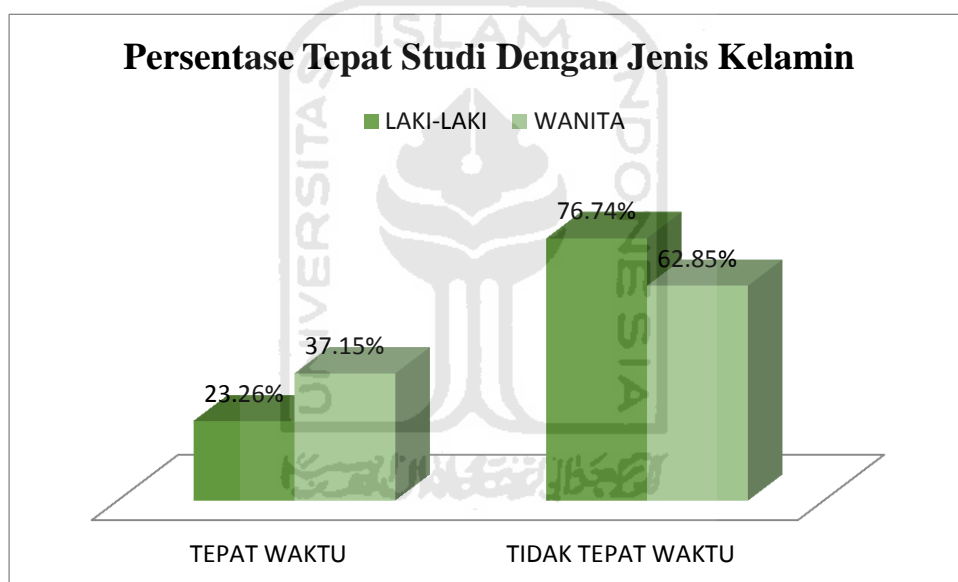
Berdasarkan Gambar 5.1 maka dapat dilihat bahwa tepat studi mahasiswa statistika UII periode 1995-1996 yaitu 100% tidak tepat waktu, selama periode tersebut mahasiswa statistika UII menjalani masa studinya lebih dari 4 tahun. Setelah periode 1995-1996 barulah terdapat mahasiswa yang masa studinya tepat waktu yaitu  $\leq 4$ . Dari gambar 5.1 juga dapat disimpulkan bahwa jumlah mahasiswa yang lulus tidak tepat waktu masih lebih banyak dari mahasiswa yang lulus tepat waktu.



Gambar 5.1 :Grafik Persentase Tepat Studi

### 5.1.2. Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Jenis Kelamin

Berdasarkan Gambar 5.2 dapat dilihat bahwa persentase tepat studi untuk *class* tepat waktu dengan jenis kelamin laki-laki sebesar 23.26%, jenis kelamin wanita sebesar 37.15%. Tepat studi untuk *class* tidak tepat waktu dengan jenis kelamin laki-laki sebesar 76.74%, dan wanita sebesar 62.85%. Dengan demikian dapat dikatakan persentase tertinggi untuk tepat studi tepat waktu yaitu jenis kelamin wanita, dengan persentase lulusnya sebesar 37.15%, dan untuk persentase tepat studi tidak tepat waktu yang tertinggi yaitu dengan jenis kelamin laki-laki, persentase lulusnya sebesar 76.74%.

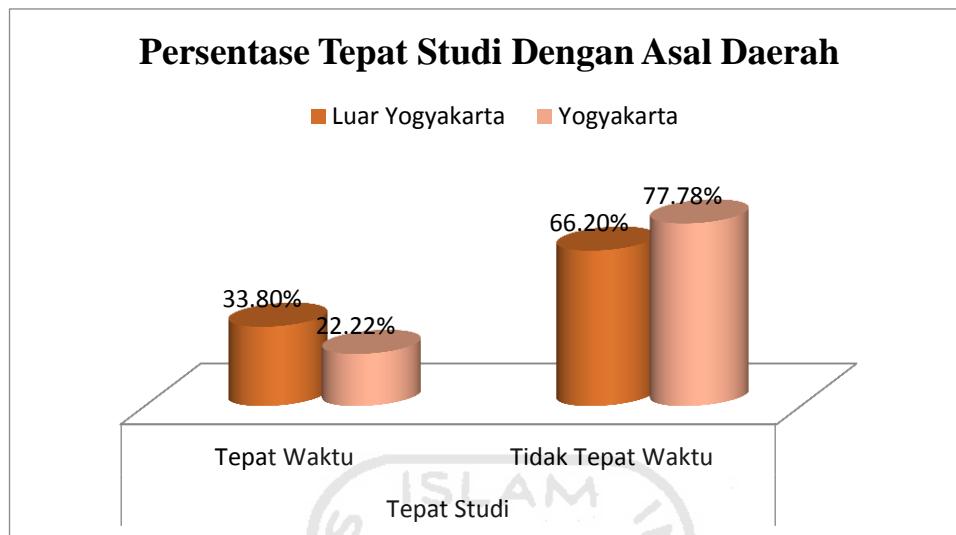


Gambar 5.2 :Grafik Persentase Tepat Studi Dengan Jenis Kelamin

### 5.1.3. Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Asal Daerah

Gambar 5.3 menunjukan persentase tepat studi dan asal daerah. Dari gambar 5.3 dapat dilihat bahwa persentase tertinggi tepat studi di *class* tepat waktu yaitu dengan kriteria asal daerah luar Yogyakarta sebesar 33.80%, sedangkan untuk tepat studi di *class* tidak tepat waktu yaitu dengan asal daerah Yogyakarta sebesar 77.78%. Persentase tepat studi untuk *class* tepat waktu dengan asal daerah Yogyakarta sebesar 22.22%, dan asal daerah luar

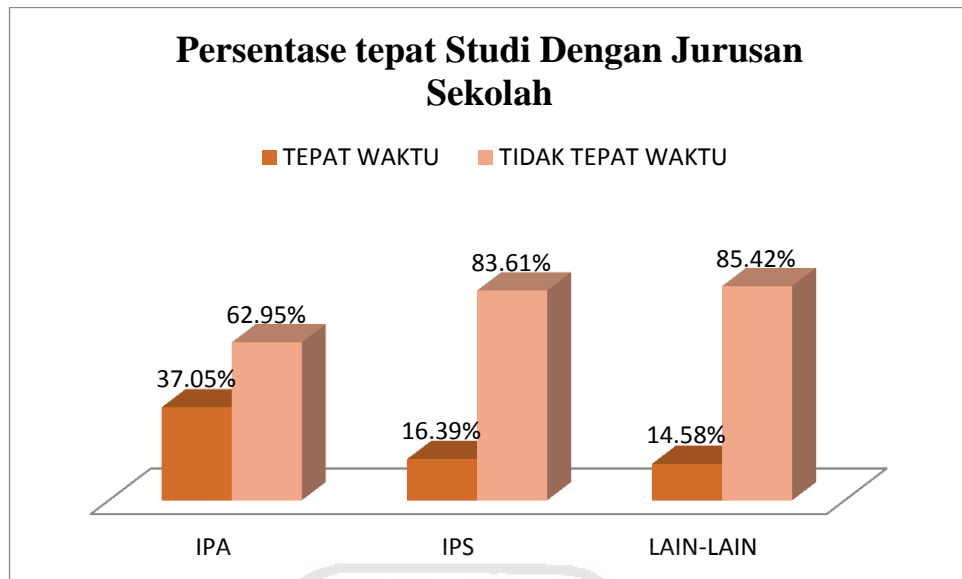
Yogyakarta sebesar 33.80%. Persentase tepat studi tidak tepat waktu untuk daerah Yogyakarta sebesar 77.78%, dan luar Yogyakarta sebesar 66.20%.



Gambar 5.3 : Grafik Persentase Tepat Studi Dengan Asal Daerah

#### 5.1.4. Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Jurusan Sekolah

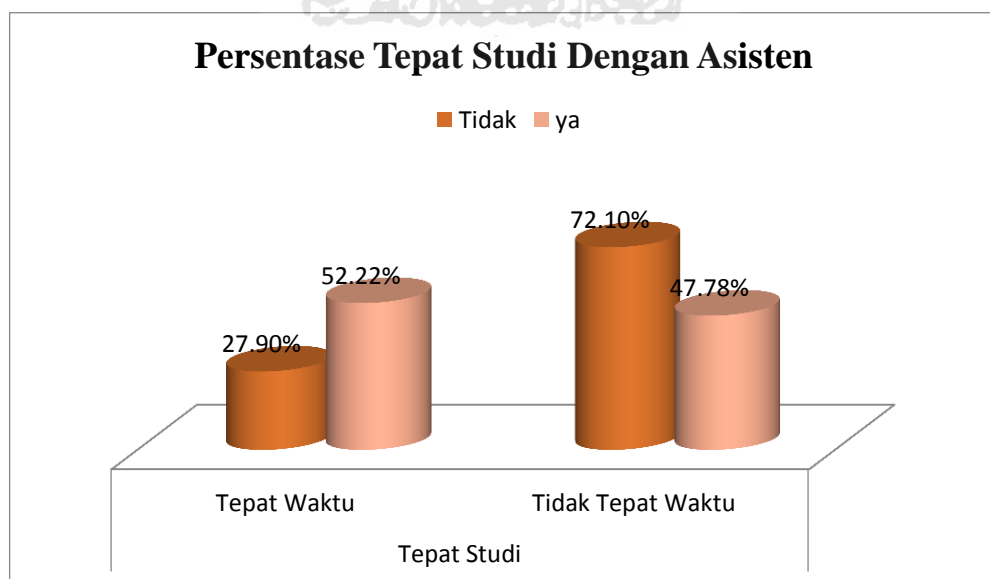
Gambar 5.4 menunjukkan persentase jurusan sekolah terhadap tepat studi. Persentase tepat studi untuk *class* tepat waktu dengan Jurusan IPA sebesar 37.05%, jurusan IPS sebesar 16.39%, dan jurusan lainnya sebesar 14.58%. Persentase tepat studi untuk *class* tidak tepat waktu dengan jurusan IPA sebesar 62.59%, Jurusan IPS sebesar 83.61%, dan jurusan lainnya 85.42%. Berdasarkan gambar diatas dapat dilihat bahwa persentase tertinggi tepat studi pada *class* tepat waktu yaitu dengan jurusan IPA , dan yang terendah yaitu jurusan selain IPA dan IPS.



Gambar 5.4 : Grafik Persentase Tepat Studi Dengan Jurusan Sekolah

#### 5.1.5. Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Asisten

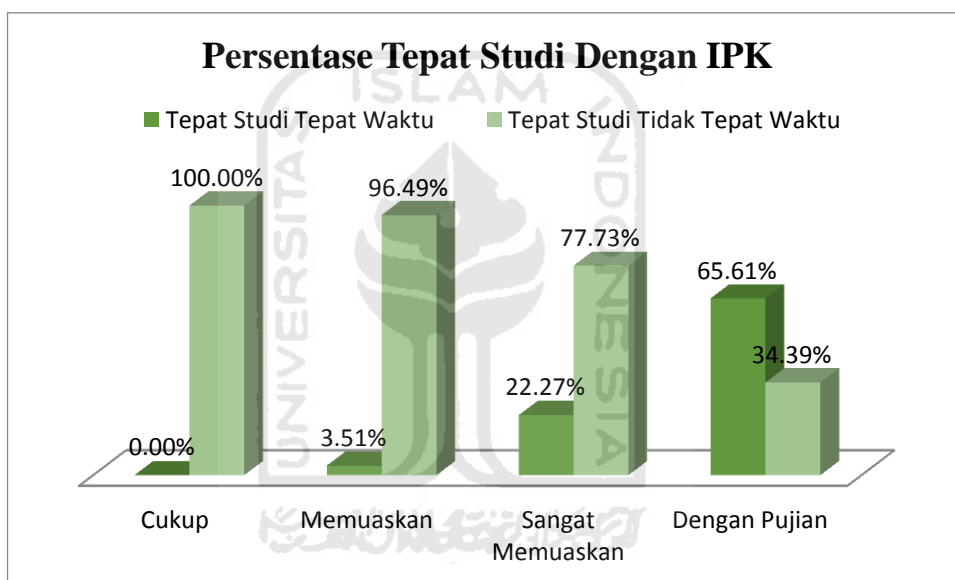
Berdasarkan Gambar 5.5 dapat dilihat bahwa persentase lulusan yang tertinggi dengan kriteria tidak tepat waktu yaitu tidak menjadi asisten dengan nilai sebesar 72.10%. Persentase lulusan tepat waktu yang tertinggi yaitu dengan kriteria menjadi asisten, persentasenya sebesar 52.22%.



Gambar 5.5 : Grafik Tepat Studi Dengan Asisten

### 5.1.6. Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan IPK

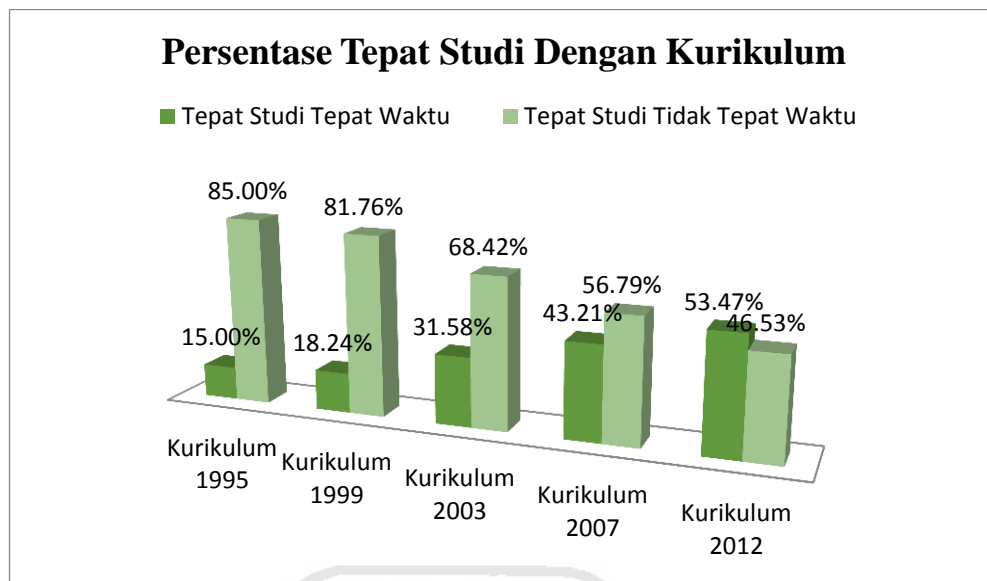
Berdasarkan Gambar 5.6 dapat dilihat bahwa persentase tepat studi di *class* tepat waktu yang tertinggi yaitu kriteria IPK dengan pujian, persentase lulusannya sebesar 65.61%. Persentase tepat studi tepat waktu dengan kriteria IPK Dengan pujian sebesar 65.61%, IPK sangat memuaskan sebesar 22.27%, IPK memuaskan sebesar 3.51%, dan IPK cukup sebesar 0.00%. Persentase tepat studi di *class* tidak tepat waktu dengan IPK cukup lulusannya sebesar 100%, IPK memuaskan sebesar 96.49%, IPK sangat memuaskan sebesar 77.73%, dan IPK dengan pujian sebesar 34.39%.



Gambar 5.6 : Grafik Persentase Tepat Studi Dengan IPK

### 5.1.7. Profil Lulusan Statistika UII Berdasarkan Tepat Studi Dengan Kurikulum

Berdasarkan Gambar 5.7 dapat dilihat bahwa persentase tepat studi di *class* tepat waktu yang tertinggi yaitu kriteria kurikulum 2012, persentase lulusannya sebesar 53.47%. Persentase tepat studi di *class* tepat waktu dengan kriteria kurikulum 2012 sebesar 54.47%, kurikulum 2007 sebesar 43.21%, kurikulum 2003 sebesar 31.58%, kurikulum 1999 sebesar 18.24%. dan kurikulum 1995 sebesar 15.00%.



Gambar 5.7 : Persentase Tepat Studi Dengan Kurikulum

## 5.2. Pemodelan Data Akademik Alumni Statistika UII dan Mahasiswa Statistika UII Angkatan 2012 Dengan Regresi Logistik

Regresi logistik adalah salah satu model untuk menduga hubungan antara peubah respon kategori dengan satu atau lebih peubah prediktor yang kontinyu ataupun kategori (Agresti, 2002). Pada bagian Regresi Logistik ini akan dicari model serta ketepatan klasifikasi tepat studi. Dalam pemodelan ini digunakan dua jenis data yaitu data *training* dan data *testing*, yang mana data alumni akan dijadikan sebagai data *training* dan data mahasiswa angkatan 2012 sebagai data *testing*.

### 5.2.1. Uji Secara Bersama

Uji serentak dilakukan untuk memeriksa peranan apakah variabel bebas (X) secara bersama-sama berpengaruh terhadap variabel terikat (Y). Statistik uji yang digunakan adalah statistik uji G, jika diperoleh statistic uji  $G > X_{(\alpha, p)}^2$  maka  $H_0$  ditolak, sehingga model dikatakan variabel bebas secara bersama-sama berpengaruh terhadap variabel terikat. Hipotesis yang digunakan adalah :

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_0 : \text{minimal terdapat satu } \beta_i \neq 0$$



Tabel 5.1. Statistik Uji G

G	$\chi^2_{(\alpha,p)}$	Keterangan
246.714	19.575	Tolak $H_0$

Berdasarkan hasil statistik uji G yang terdapat pada tabel 5.1, maka didapat nilai G sebesar 246.714 dan  $\chi^2_{(\alpha,p)}=19.675$  dengan nilai  $\alpha=0.05$ . Sehingga berdasarkan hipotesis uji G menghasilkan keputusan bahwa  $G(246.714) > \chi^2_{(\alpha,p)}(19.675)$ . Artinya variabel bebas secara bersama-sama mempengaruhi variabel terikat.

### 5.2.2. Uji Wald

Uji wald bertujuan untuk mengetahui peranan masing-masing variabel prediktor terhadap variabel respon. Statistic uji yang digunakan yakni statistik uji  $W$  jika nilai statistik uji  $W > \chi^2_{(\alpha,1)}$  maka tolak  $H_0$ , atau dengan kata lain variabel prediktor mempengaruhi variabel respon. Pengujian pada masing masing variabel didasari oleh hipotesis :

$H_0 : \beta_j = 0$ , dengan  $j = 1, 2, \dots, p$  (artinya variabel prediktor tidak berpengaruh terhadap variabel respon)

$H_1 : \beta_j \neq 0$ , dengan  $j = 1, 2, \dots, p$  (artinya variabel prediktor berpengaruh secara individu terhadap variabel respon)

Hasil pengujian Uji Wald sebagai berikut :

Tabel 5.2. Statistik Uji Wald

Variabel	Wald	P-value	Keterangan
Jenis Kelamin	9.444	0.002	Tolak $H_0$
KR3	4.732	0.030	Tolak $H_0$
KR4	50.453	0.000	Tolak $H_0$
IPK2 (1)	7.227	0.007	Tolak $H_0$
IPK3 (1)	20.241	0.000	Tolak $H_0$

Dari hasil pengujian pada tabel 5.2 diatas dengan tingkat kepercayaan 95% diperoleh variabel jenis kelamin, KR3, KR4, IPK2, dan IPK3 (Indeks Prestasi Kumulatif) masing-masing signifikan berpengaruh terhadap variabel Y (tepat studi). Berdasarkan hasil output nilai wald untuk masing-masing variabel jenis kelamin, KR3, KR4, IPK2, dan IPK3 adalah 9.444, 4.732, 50.453, 7.227, dan 20.241 yang dimana nilai wald lebih besar dari  $\chi^2_{(\alpha,1)}$ . sehingga hipotesis awal  $H_0$  ditolak. Artinya variabel jenis kelamin, KR3, KR4, IPK2 dan IPK3 berpengaruh terhadap variabel terikat.

Setelah dilakukan uji secara keseluruhan maupun individual diperoleh hasil bahwa variabel Jurusan asal sekolah dan IPK berpengaruh signifikan terhadap variabel Lama Studi. Berikut adalah model terbaik :

$$\pi(x) = \frac{\exp(-4.098 - 1.011 \text{ jenis kela min} + 0.666KR3 + 22.720KR4 + 2.760IPK2 + 4.631IPK3)}{1 + \exp(-4.098 - 1.011 \text{ jenis kela min} + 0.666KR3 + 22.720KR4 + 2.760IPK2 + 4.631IPK3)}$$

Dengan model logit :

$$\hat{g}(x) = -4.098 - 1.011 \text{ jenis kela min} + 0.666KR3 + 22.720KR4 + 2.760IPK2 + 4.631IPK3$$

Dimana :

KR3 = Nilai *dummy* ketiga untuk variabel kurikulum pada observasi ke-i.

KR4 = Nilai *dummy* keempat untuk variabel kurikulum pada observasi ke-i.

IPK2 = Nilai *dummy* kedua untuk variabel IPK pada observasi ke-i.

IPK3 = Nilai *dummy* ketiga untuk variabel IPK pada observasi ke-i.

Dari model yang terbentuk, maka dapat diprediksi nilai peluang untuk masing-masing kriteria adalah sebagai berikut :

Tabel 5.3. Estimasi Probabilitas Regresi Logistik

Jenis kelamin	Kurikulum	IPK	Tepat Studi	
			Tepat Waktu	Tidak Tepat Waktu
Laki-Laki	Selain KR3 dan KR4	IPK1	0.006	0.994
		IPK2	0.087	0.913
		IPK3	0.383	0.617
	KR3	IPK1	0.012	0.988
		IPK2	0.157	0.843
		IPK3	0.547	0.453

Jenis kelamin	Kurikulum	IPK	Tepat Studi	
			Tepat Waktu	Tidak Tepat Waktu
Wanita	KR4	IPK1	1.000	0.000
		IPK2	1.000	0.000
		IPK3	1.000	0.000
	Selain KR3 dan KR4	IPK1	0.016	0.984
		IPK2	0.208	0.792
		IPK3	0.630	0.370
	KR3	IPK1	0.031	0.969
		IPK2	0.338	0.662
		IPK3	0.768	0.232
KR4	IPK1	1.000	0.000	
	IPK2	1.000	0.000	
	IPK3	1.000	0.000	

Dari Tabel 5.3 maka dapat dijelaskan bahwa mahasiswa yang tepat studinya tepat waktu adalah mahasiswa yang memiliki peluang lebih besar dari 0.5. Terlihat bahwa mahasiswa yang memiliki kriteria jenis kelamin wanita, kurikulum selain 2012, IPK memuaskan memiliki peluang lulus tidak tepat waktu. Keadaan yang sama juga terjadi pada mahasiswa dengan jenis kelamin wanita, kurikulum selain 2012, IPK memuaskan. Selain dari dua kriteria yang telah dijelaskan sebelumnya, maka akan lulus tepat waktu.

### 5.2.3. Odds Ratio

Nilai-nilai  $\beta$  pada model di atas akan lebih mudah diinterpretasikan melalui *odds ratio*. Analisisnya sebagai berikut :

Tabel 5.4. Odds Ratio

Variabel	Odds Ratio
Jenis Kelamin	0.364
KR3	1.946
KR4	7364964008.357
IPK2	15,796
IPK3	102,608

1. Dalam kasus variabel jenis kelamin, dengan *odds ratio* sebesar 0.443 dapat diartikan bahwa mahasiswa dengan kriteria jenis kelamin laki-laki peluang lulus tepat waktu adalah 0.443 lebih kecil dari mahasiswa dengan jenis kelamin laki-laki
2. Pada variabel kurikulum 2007 (KR3), dengan *odds ratio* sebesar 1.946 dapat diartikan bahwa mahasiswa yang menggunakan kurikulum 12 memiliki peluang lulus tepat waktu sebesar 1.946 kali dibandingkan dengan mahasiswa dengan kriteria kurikulum yang lainnya.
3. Pada variabel kurikulum 2012 (KR4), dengan *odds ratio* sebesar 7364964008.357 dapat diartikan bahwa mahasiswa yang menggunakan kurikulum 2012 memiliki peluang lulus tepat waktu sebesar 7364964008.357 kali dibandingkan dengan mahasiswa dengan kriteria kurikulum yang lainnya.
4. Pada variabel IPK (IPK2), dengan *odds ratio* sebesar 15,796 dapat diartikan bahwa mahasiswa dengan kriteria IPK sangat memuaskan memiliki peluang lulus tepat waktu sebesar 15,796 kali dibandingkan dengan mahasiswa dengan kriteria IPK Cukup atau memuaskan.
5. Pada variabel IPK (IPK3), dengan *odds ratio* sebesar 102,608 dapat diartikan bahwa mahasiswa yang memiliki kriteria IPK dengan pujian mempunyai peluang lulus tepat waktu sebesar 102,608 kali dibandingkan dengan mahasiswa dengan kriteria IPK Cukup, memuaskan dan sangat memuaskan.

#### 5.2.4. Prediksi Tepat Studi

Berikut ini adalah hasil prediksi tepat studi mahasiswa statistika UII angkatan 2012 :

Tabel 5.5. Hasil Prediksi Tepat Studi

<b>Tepat Studi</b>	
<b>Tepat Waktu</b>	<b>Tidak Tepat Waktu</b>
116	0

Berdasarkan Tabel 5.5 maka diperoleh hasil prediksi tepat studi mahasiswa statistika UII angkatan 2012. Dari 116 mahasiswa yang diprediksi, didapatkan seluruh mahasiswa statistika angkatan 2012 lulus tepat waktu.

### 5.2.5. Ketepatan Klasifikasi

Berdasarkan Tabel 5.6, maka didapat hasil ketepatan klasifikasi tepat studi di *class* tepat waktu sebanyak 53 mahasiswa tepat dalam pengklasifikasian dari 53 mahasiswa yang lulus tepat waktu. Tidak ada mahasiswa yang tepat dalam klasifikasi untuk tepat studi di *class* tidak tepat waktu. Dengan demikian ketepatan klasifikasi tepat studi mahasiswa statistika UII dengan menggunakan regresi logistik sebesar 45.69%.

Tabel 5.6. Ketepatan Klasifikasi

Tepat Studi	Tepat Waktu	Tidak Tepat Waktu	<i>Percentage Correct</i>
Tepat Waktu	53	63	$\frac{(53+0)}{116} = 45.69\%$
Tidak Tepat Waktu	0	0	

### 5.3. Pemodelan Data Akademik Alumni Statistika UII dan Mahasiswa Statistika UII Angkatan 2012 Dengan Naïve Bayes

Dalam menerapkan algoritma *Naïve Bayes*, data dibagi menjadi 2 bagian yaitu data *training* dan data *testing*. Oleh algoritma *Naive Bayes Classification*, data training digunakan untuk membentuk tabel probabilitas, dan data *testing* digunakan untuk menguji tabel probabilitas yang terbentuk.

#### 5.3.1. Pengujian Data *Training*

Pengujian data *training* dilakukan untuk mendapatkan model Naïve Bayes yang kemudian model tersebut akan digunakan untuk memprediksi lama studi mahasiswa. Berikut ini data *training* yang digunakan dalam penelitian ini :

Tabel 5.7. Data *Training*

No	Jenis Kelamin	Asal Daerah	Jurusan Sekolah	Asisten	Kurikulum	IPK	Tepat Studi
1	LAKI-LAKI	Yogyakarta	IPA	TIDAK	Kurikulum 1995	Sangat Memuaskan	TIDAK TEPAT WAKTU
2	LAKI-LAKI	Luar Yogyakarta	IPA	TIDAK	Kurikulum 1995	Sangat Memuaskan	TIDAK TEPAT WAKTU
3	WANITA	Yogyakarta	IPA	TIDAK	Kurikulum 1995	Memuaskan	TIDAK TEPAT WAKTU
:							
487	WANITA	Luar Yogyakarta	IPA	TIDAK	Kurikulum 2011	Dengan Pujian	TEPAT WAKTU
488	WANITA	Luar Yogyakarta	IPA	TIDAK	Kurikulum 2011	Dengan Pujian	TEPAT WAKTU

Dari tabel 5.7 kemudian dicari probabilitas *prior* masing-masing *class* lama studi (Y) dan probabilitas bersyarat untuk setiap variabel independen (X) terhadap Y. berikut ini probabilitas setiap kriteria :

**a. Probabilitas Tepat Studi (Y)**

Tabel 5.7. menjelaskan tentang probabilitas tepat studi “Tepat Waktu” dan “Tidak Tepat Waktu” mahasiswa prodi statistika UII angkatan 1995-2011.

Tabel 5.8. Probabilitas Tepat Studi

Probabilitas	
Tepat Waktu	Tidak Tepat Waktu
0.323	0.677

Dari Tabel 5.8. diperoleh gambaran probabilitas tepat studi di *class* tepat waktu sebesar 0.323, dan tepat studi di *class* tidak tepat waktu sebesar 0.677. Dari tabel diatas dapat ditarik kesimpulan bahwa tepat studi mahasiswa yang tidak tepat waktu lebih besar dari tepat studi mahasiswa yang tepat waktu.

### b. Probabilitas Tepat Studi Dengan Jenis Kelamin

Tabel 5.9 menggambarkan tentang probabilitas tepat studi mahasiswa prodi statistika UII angkatan 1995-2011 dengan kriteria jenis kelamin.

Tabel 5.9. Probabilitas Tepat Studi Dengan Jenis Kelmain

Jenis Kelamin	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Wanita	0.750	0.606
Laki - Laki	0.250	0.394

Berdasarkan Tabel 5.9. Didapatkan gambaran dari data alumni yang mempunyai probabilitas terbesar adalah mahasiswa dengan jenis kelamin wanita dan tepat studi “Tepat Waktu”, sedangkan probabilitas yang terkecil adalah mahasiswa dengan jenis kelamin laki-laki dan tepat studi “Tepat Waktu”. Berdasarkan Tabel 5.9 juga dapat disimpulkan bahwa mahasiswa yang berjenis kelamin wanita memiliki peluang untuk lulus tepat waktu lebih besar dari mahasiswa dengan jenis kelamin laki-laki.

### c. Probabilitas Tepat Studi Dengan Asal Daerah

Tabel 5.10 menjelaskan probabilitas tepat studimahasiswa prodi statistika angkatan 1995-2011 dengan kriteria tepat studi dan asal daerah.

Tabel 5.10. Probabilitas Tepat Studi Dengan Asal Daerah

Asal Daerah	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Yogyakarta	0.087	0.146
Luar Yogyakarta	0.913	0.854

Berdasarkan Tabel 5.10. Diperoleh gambaran data alumni yang mempunyai probabilitas terbesar dengan kriteria asal daerah luar Yogyakarta adalah lulusan tepat waktu, sedangkan probabilitas terbesar dengan kriteria asal daerah Yogyakarta adalah tidak tepat waktu. Dari Tabel 5.10 juga dapat dikatakan bahwa mahasiswa yang bersal dari luar Yogyakarta memiliki peluang untuk lulus tepat waktu lebih besar dari mahasiswa yang berasal dari Yogyakarta.

#### d. Probabilitas Tepat Studi Dengan Jurusan Sekolah

Tabel 5.11 menjelaskan terkait probabilitas tepat studimahasiswa prodi statistika angkatan 1995-2011 dengan jurusan sekolah.

Tabel 5.11. Probabilitas Tepat Studi Dengan Jurusan Sekolah

Jurusan Sekolah	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
IPA	0.894	0.725
IPS	0.062	0.152
Lain-Lain	0.044	0,122

Berdasarkan Tabel 5.11 maka probabilitas terbesar adalah mahasiswa dengan jurusan IPA dan tepatstudi “Tepat Waktu”, kemudian disusul oleh mahasiswa dengan kriteria jurusan IPA dan tepat studi “Tidak Tepat Waktu”. Probabilitas terkecil yaitu mahasiswa dengan kriteria Jurusan lain-lain dan tepat studi “Tepat Waktu”. Dari Tabel 5.11 juga dapat dikatakan bahwa mahasiswa dengan kriteria jurusan IPA cenderung untuk lulus tepat waktu.

#### e. Probabilitas Tepat Studi Dengan Asisten

Tabel 5.12 menjelaskan tentang probabilitas mahasiswa tepat studi prodi statistika UII angkatan 1995-2011 dengan kriteria asisten.

Tabel 5.12. Probabilitas Tepat Studi Dengan Asisten

Asisiten	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
YA	0.294	0.128
TIDAK	0.706	0.872

Berdasarkan Tabel 5.12 maka probabilitas terbesar adalah mahasiswa dengan kriteria asisten“Tidak” dan tepat studi “ Tidak Tepat Waktu”.Probabilitas terkecil yaitu mahasiswa dengan kriteria asisten “YA” dantepat studi “Tidak Tepat Waktu”. Dari Tabel 5.12 juga dapat dikatakan bahwa mahasiswa yang menjadi asisten praktikumakan cenderung lulus tepat waktu.



#### f. Probabilitas Tepat Studi Dengan Kurikulum

Tabel 5.13 menjelaskan terkait probabilitas tepat studimahasiswa prodi statistika angkatan 1995-2011 dengan kurikulum.

Tabel 5.13. Probabilitas Tepat Studi Dengan Kurikulum

Kurikulum	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Kurikulum 1995	0.037	0.101
Kurikulum 1999	0.181	0.388
Kurikulum 2003	0.225	0.233
Kurikulum 2007	0.219	0.137
Kurikulum 2012	0.337	0.140

Berdasarkan Tabel 5.13 maka probabilitas terbesar adalah mahasiswa dengan kriteria kurikulum 1999 dan tepat studi “Tidak Tepat Waktu”, kemudian disusul oleh mahasiswa dengan kriteria kurikulum 2012 dan tepat studi “Tepat Waktu”. Probabilitas terkecil yaitu mahasiswa dengan kriteria kurikulum 2012 dan tepat studi “Tidak Tepat Waktu”. Dari Tabel 5.13 juga dapat dikatakan bahwa mahasiswa yang memiliki kriteria kurikulum 2012 akan cenderung lulus tepat waktu.

#### g. Probabilitas Tepat Studi Dengan IPK

Tabel 5.14 menjelaskan terkait probabilitas tepat studimahasiswa prodi statistika angkatan 1995-2011 dengan IPK.

Tabel 5.14. Probabilitas Tepat Studi Dengan IPK

IPK	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Cukup	0.000	0.101
Memuaskan	0.012	0.164
Sangat Memuaskan	0.344	0.573
Dengan Pujian	0.644	0.161

Berdasarkan Tabel 5.14 maka probabilitas terbesar adalah mahasiswa dengan kriteria IPK dengan pujian dan tepat studi “Tepat Waktu”, kemudian disusul oleh mahasiswa dengan kriteria IPK sangat memuaskan dan tepat studi “Tidak Tepat Waktu”. Probabilitas terkecil yaitu mahasiswa dengan kriteria

IPK cukup dan tepat studi “Tepat Waktu”. Dari Tabel 5.14 juga dapat dikatakan bahwa mahasiswa yang memiliki kriteria IPK dengan pujian akancenderung lulus tepat waktu.

### 5.3.2. Pengujian Data *Testing*

Setelah model Naïve Bayes terbentuk, maka akan dilakukan klasifikasi tepat studi mahasiswa statistika UII angkatan 2012, Berikut ini diberikan salah seorang mahasiswa statistika UII angkatan 2012 untuk dilakukan klasifikasi secara manual. karakteristiknya sebagai berikut:

Tabel 5.15. Data *Testing*

Jenis Kelamin	Asal Daerah	Jurusan sekolah	Asisten	Kurikulum	IPK	Tepat Studi
Laki-laki	Luar Yogyakarta	IPA	Ya	Kurikulum 2012	Dengan Pujian	?

#### a. Menghitung Probabilitas Besyarat Sesuai Kondisi Yang Diberikan Oleh Data *Testing*.

$$P(C_{\text{laki-laki}} | F_{\text{Tepat Waktu}}) = 0.250$$

$$P(C_{\text{luar Yogyakarta}} | F_{\text{Tepat Waktu}}) = 0.913$$

$$P(C_{\text{IPA}} | F_{\text{Tepat Waktu}}) = 0.894$$

$$P(C_{\text{YA}} | F_{\text{Tepat Waktu}}) = 0.294$$

$$P(C_{\text{kurikulum 2012}} | F_{\text{Tepat Waktu}}) = 0.337$$

$$P(C_{\text{Dengan Pujian}} | F_{\text{Tepat Waktu}}) = 0.644$$

$$P(C_{\text{laki-laki}} | F_{\text{Tidak Tepat Waktu}}) = 0.394$$

$$P(C_{\text{Luar Yogyakarta}} | F_{\text{Tidak Tepat Waktu}}) = 0.854$$

$$P(C_{\text{IPA}} | F_{\text{Tidak Tepat Waktu}}) = 0.725$$

$$P(C_{\text{YA}} | F_{\text{Tidak Tepat Waktu}}) = 0.128$$

$$P(C_{\text{kurikulum 2012}} | F_{\text{TidakTepat Waktu}}) = 0.140$$

$$P(C_{\text{Dengan Pujian}} | F_{\text{Tidak Tepat Waktu}}) = 0.161$$

#### b. Melakukan Perkalian Pada Semua Hasil Probabilitas Untuk Setiap *Class* Tepat Studi

Untuk tepat studi = Tepat Waktu

$$P = 0.250 * 0.913 * 0.894 * 0.294 * 0.337 * 0.644 * 0.323 = 0.0042$$

Untuk tepat studi = Tidak Tepat Waktu

$$P = 0.394 * 0.854 * 0.725 * 0.128 * 0.140 * 0.161 * 0.677 = 0.0004$$

### c. Membandingkan Probabilitas Tepat Studi

Berdasarkan peluang terjadinya masing-masing *class* tepat studi berdasarkan kriteria yang diajukan, dapat dilihat bahwa probabilitas tepat studi di *class* tepat waktu lebih besar dari tepat studi di *class* tidak tepat waktu. Dengan demikian prediksi tepat studi untuk data *testing* yang diberikan adalah tepat waktu.

Berikut ini hasil pengklasifikasian tepat studi mahasiswa statistika UII angkatan 2012 menggunakan *software R* :

Tabel 5.16. Hasil Prediksi Tepat Studi Mahasiswa Statistika UII Angkatan 2012

No	Jenis Kelamin	Asal Daerah	Jurusan Sekolah	Asisten	Kurikulum	IPK	Tepat Studi
1	LAKI-LAKI	Luar Yogyakarta	IPS	TIDAK	Kurikulum 2012	Cukup	TIDAK TEPAT WAKTU
2	WANITA	Luar Yogyakarta	IPA	TIDAK	Kurikulum 2012	Sangat Memuaskan	TIDAK TEPAT WAKTU
3	LAKI-LAKI	Yogyakarta	IPA	TIDAK	Kurikulum 2012	Dengan Pujian	TIDAK TEPAT WAKTU
:							
115	WANITA	Yogyakarta	IPA	TIDAK	Kurikulum 2012	Memuaskan	TIDAK TEPAT WAKTU
116	LAKI-LAKI	Yogyakarta	IPA	TIDAK	Kurikulum 2012	Sangat Memuaskan	TIDAK TEPAT WAKTU

Berdasarkan Tabel 5.16 menunjukkan hasil prediksi tepat studi mahasiswa prodi statistika UII angkatan 2012. Dari 116 mahasiswa yang diprediksi didapatkan mahasiswa yang tepat studinya “Tepat Waktu” sebanyak 35 mahasiswa, dan yang tepat studinya “Tidak Tepat Waktu” sebanyak 81 mahasiswa. Untuk hasil yang lebih lengkap dapat dilihat pada lampiran.

### 5.3.3. Ketepatan Klasifikasi

Berdasarkan Tabel 5.17 dapat dilihat dari 53 mahasiswa yang tepat studinya “Tepat Waktu” terdapat 30 mahasiswa diklasifikasikan secara tepat, dan 23 mahasiswa lainnya tidak tepat. Salah satu klasifikasi juga terjadi pada tepat studi di *class* tidak tepat waktu yaitu sebanyak 5 mahasiswa terjadi salah satu klasifikasi dari 63 mahasiswa yang tepat studinya tidak tepat waktu dan sisanya tepat dalam klasifikasi. Dengan demikian dapat dikatakan bahwa algoritma Naïve Bayes

berhasil mengklasifikasikan tepat studi mahasiswa statistika UII angkatan 2012 dengan persentase akurasi sebesar 75.86%.

Tabel 5.17. Ketepatan Klasifikasi

Tepat Studi	Tepat Waktu	Tidak Tepat Waktu	<i>Percentage Correct</i>
Tepat Waktu	30	5	$\frac{(30+58)}{116} = 75.86\%$
Tidak Tepat Waktu	23	59	

#### 5.4. Perbandingan Ketepatan Klasifikasi

Setelah mengetahui hasil masing-masing ketepatan klasifikasi pada kedua metode, maka langkah selanjutnya adalah membandingkannya. Berikut merupakan perbandingan antara kedua metode berdasarkan tingkat akurasi.

Tabel 5.18. Perbandingan Ketepatan Klasifikasi

Metode	Ketepatan Klasifikasi
Regresi Logistik	45.69%
Naïve Bayes	75.86%

Berdasarkan tabel 5.18 maka dapat dikatakan bahwa metode terbaik dalam pengklasifikasian tepat studi mahasiswa statistika UII angkatan 2012 yaitu algoritma Naïve Bayes dengan tingkat akurasi sebesar 75.86%.

## BAB VI

### KESIMPULAN DAN SARAN

#### 6.1. Kesimpulan

Berdasarkan rumusan masalah yang telah dijelaskan pada bab sebelumnya, maka diperoleh kesimpulan berikut untuk menjawab rumusan masalah yang ada :

1. Dari hasil analisis diperoleh model Regresi Logistik dan algoritma Naïve Bayes untuk memprediksi tepat studi mahasiswa statistika UII angkatan 2012 :
  - a. Model Regresi Logistik untuk memprediksi tepat studi mahasiswa statistika dapat dilihat di halaman 35. Dari model yang terbentuk, terdapat tiga variabel yang signifikan yaitu jenis kelamin, Kurikulum 2007 (KR3), kurikulum 2012 (KR4), IPK sangat memuaskan (IPK2) dan IPK dengan pujian (IPK3). Hasil klasifikasi dengan menggunakan Regresi Logistik didapatkan seluruh mahasiswa statistika UII angkatan 2012 tepat waktu. Ketepatan klasifikasi atau tingkat akurasi model Regresi Logistik dalam mengklasifikasi mahasiswa statistika UII angkatan 2012 sebesar 45.69%.
  - b. Model algoritma Naïve Bayes untuk memprediksi tepat studi mahasiswa statistika UII angkatan 2012 dapat dilihat di halaman 39-42. Dari model tersebut dapat dilakukan klasifikasi tepat studi mahasiswa statistika UII dengan cara menghitung peluang disetiap kondisi, kemudian bandingkan probabilitas masing-masing *class*. *Class* dengan probabilitas yang paling besar itulah yang menjadi tujuan dari hasil klasifikasi. Hasil klasifikasi dengan menggunakan algoritma Naïve Bayes didapatkan sebanyak 35 mahasiswa lulus tepat waktu, dan 81 mahasiswa lulus tidak tepat waktu. ketepatan klasifikasi atau tingkat akurasi model algoritma Naïve Bayes dalam mengklasifikasi mahasiswa statistika UII angkatan 2012 sebesar 75,86%.

2. Dalam penelitian ini, algoritma Naïve Bayes lebih baik dalam mengklasifikasi tepat studi mahasiswa statistika UII angkatan 2012, hal ini dijelaskan dengan melihat nilai akurasi sebesar 75.86.%, dan Regresi Logistik sebesar 45.69%. Kondisi ini dikarenakan pada algoritma Naïve Bayes mengakomodasi semua kriteria untuk melakukan pengklasifikasian, sedangkan Regresi Logistik tidak demikian. Regresi Logistik hanya menggunakan kriteria yang signifikan saja. Jadi dalam penelitian pengklasifikasian tepat studi mahasiswa statistika UII angkatan 2012 lebih baik menggunakan algoritma Naïve Bayes.

## 6.2. Saran

Saran yang dapat diberikan dalam penelitian ini adalah :

1. Banyaknya tepat studi tidak tepat waktu yang terdapat pada data *training* mengakibatkan tingkat ketepatan klasifikasi tidak terlalu tinggi. Untuk penelitian selanjutnya agar diperhatikan data *training* yang akan digunakan.
2. Variabel indepenen dapat ditambah lagi, seperti jumlah mengikuti tes CEPT, dan berorganisasi selama masa kuliah.
3. Berdasarkan hasil penelitian ini, maka peneliti menyarankan agar dibangun aplikasi untuk memprediksi tepat studi dengan menggunakan data akademik mahasiswa statistika UII. Input dari aplikasi adalah data induk mahasiswa dan data akademik mahasiswa. Output dari aplikasi adalah hasil prediksi tepat studi yaitu dalam bentuk tepat waktu atau tidak tepat waktu. Sehingga apabila mahasiswa yang diprediksi lulus tidak tepat waktu maka dapat diberi bimbingan studi.

## DAFTAR PUSTAKA

- Agresti, A., (2002). *Categorical Data Analysis Second Edition*. New York: John Wiley & Sons.
- BAN PT - Badan Akreditasi Nasional Perguruan Tinggi. (2008). *Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi*.
- BAN PT - Badan Akreditasi Nasional Perguruan Tinggi. (2011). *Akreditasi Institusi Perguruan Tinggi - Buku III Pedoman Penyusunan Borang*, pp.4.
- Budiatil, D., & Wilandari, Y. (2013). *Analisis Lama Studi Mahasiswa Fsm Universitas Diponegoro Menggunakan Regresi Logistik Biner*.
- Hosmer, David W. & Lemeshow, Stanley. (2000). *Applied Logistik Regression*. Willey.
- Jananto, A. (2013). *Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa*. *Dinamik-Jurnal Teknologi Informasi*, 18(1).
- Kundriasworo, M. A. (2013). *Perbandingan Metode Regresi Logistik Dan Metode Chaid (Chi-Square Automatic Interaction Detection) Dalam Studi Kasus Ketepatan Masa Studi Mahasiswa Uin Sunan Kalijaga* (Doctoral Dissertation, Uin Sunan Kalijaga).
- Margasari, A. (2014). *Penerapan Metode Cart (Classification And Regression Trees) Dan Analisis Regresi Logistik Biner Pada Klasifikasi Profil Mahasiswa Fmipa Universitas Brawijaya*. *Jurnal Mahasiswa Statistik*, 2(4), Pp-257.
- Mauriza, A. F. (2014). *Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika UMS Menggunakan Metode Naive Bayes* (Doctoral dissertation, Universitas Muhammadiyah Surakarta).
- Meinanda, M. H., Annisa, M., Muhandri, N., & Suryadi, K. (2009). *Prediksi masa studi sarjana dengan artificial neural network*. *Internetworking Indonesia Journal*, 1(2), 31-35.
- Murtopo, A. A. (2016). *Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naive Bayes*. *CSRID Journal*, 7(3), 145-154

- Nugraha J. (2013). *Pengantar Analisis Data Kategorik*. Sleman : CV BUDI UTAMA.
- Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia. (2013) *Tentang Standar Nasional Pendidikan Tinggi*.
- Peraturan Pemerintah Republik Indonesia Nomor 66. (2010). *Tentang Pengelolaan Dan Penyelenggaraan Pendidikan*.
- Rachimawan, A. F., & Ulama, B. S. S. (2016). *Ads Filtering Menggunakan Jaringan Syaraf Tiruan Perceptron, Naïve Bayes Classifier, dan Regresi Logistik*. Jurnal Sains dan Seni ITS, 5(1), D83-D89.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Jurnal EECCIS, 7(1), 59-64.
- Sari, M. K., Ernawati, E., & Wisnubhadra, I. (2016). *Pembangunan Aplikasi Klasifikasi Mahasiswa Baru untuk Prediksi Hasil Studi Menggunakan Naïve Bayes Classifier*. Jurnal Buana Informatika, 7(2).
- Undang Undang Republik Indonesia Nomor 20. (2003). *Tentang Sistem Pendidikan Nasional*.



