

**PERBANDINGAN KLASIFIKASI KETEPATAN WAKTU
KELULUSAN MAHASISWA MENGGUNAKAN REGRESI
LOGISTIK BINER DAN *NAÏVE BAYES CLASSIFIER***

(Studi kasus : Data Kelulusan Mahasiswa S1 tahun akademik 2016/2017

Universitas Islam Indonesia)

TUGAS AKHIR



Disusun Oleh:

Erene Fajrila

14 611 144

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2018**

**PERBANDINGAN KLASIFIKASI KETEPATAN WAKTU
KELULUSAN MAHASISWA MENGGUNAKAN REGRESI
LOGISTIK BINER DAN *NAÏVE BAYES CLASSIFIER***

(Studi kasus : Data Kelulusan Mahasiswa S1 tahun akademik 2016/2017

Universitas Islam Indonesia)

TUGAS AKHIR



Disusun Oleh:

Erene Fajrila

14 611 144

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2018

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistik Biner dan *Naive Bayes Classifier* (Studi Kasus: Data Kelulusan Mahasiswa S1 tahun akademik 2016/2017 Universitas Islam Indonesia)

Nama : Erene Fajrila

Nomor Mahasiswa : 14 611 144

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

Yogyakarta, 19 April 2018

Mengetahui,
Dosen Pembimbing



Tuti Purwaningsih, S.Stat., M.Si.

**HALAMAN PENGESAHAN
TUGAS AKHIR**

**PERBANDINGAN KLASIFIKASI KETEPATAN WAKTU
KELULUSAN MAHASISWA MENGGUNAKAN REGRESI
LOGISTIK BINER DAN *NAÏVE BAYES CLASSIFIER***

(Studi Kasus : Data Kelulusan Mahasiswa S1 tahun akademik 2016/2017

Universitas Islam Indonesia)

Mahasiswa : Erene Fajrila

Nomor Mahasiswa : 14 611 144

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 23 MEI 2018**

Nama Penguji :

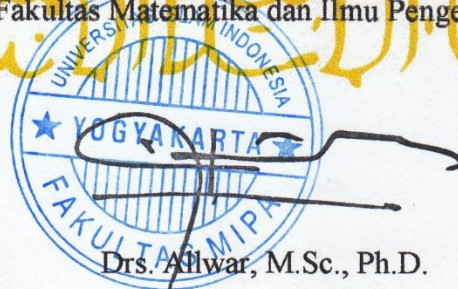
1. Andrie Pasca Hendradewa, S.T., M.T
2. Muhammad Muhajir, S.Si., M.Sc
3. Tuti Purwaningsih, S.Stat., M.Si

Tanda Tangan

.....
.....
.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Drs. Aliwar, M.Sc., Ph.D.

HALAMAN PERSEMBAHAN

Tugas akhir ini penulis persembahkan untuk orang-orang yang disayang:

- ✓ *Bapak Drs. Erni Trisna dan Ibu Neneng Triana, selaku orang tua saya yang selalu mendoakan dan selalu memenuhi kebutuhan yang saya inginkan, yang selalu menyemangati saya dalam menyelesaikan tugas akhir ini, yang selalu mendukung saya dalam setiap hal yang saya lakukan selama itu di jalan positif, yang selalu percaya kepada saya selama saya menempuh pendidikan di Yogyakarta. Terima kasih untuk semua yang telah diberikan kepada saya sampai detik ini, meskipun saya tidak bisa membalas budi kepada orang tua. Semoga karya ini bisa menjadi salah satu hadiah terindah meskipun masih banyak kekurangan.*
- ✓ *Ibu Tuti Purwaningsih, S.Stat, M.Si selaku dosen pembimbing Tugas Akhir saya, yang sudah membimbing saya dengan sangat sabar dan selalu menyemangati saya dan anak-anak bimbingan TA lainnya. Mengajari saya ilmu yang tidak bisa terbayarkan, mengajari saya artinya berjuang untuk mendapatkan sesuatu yang diinginkan. Terima kasih Ibu, jasamu akan saya kenang selalu, Ilmu yang Ibu berikan akan saya terapkan seumur hidup saya.*
- ✓ *Muhammad Fikri Erlangga dan Muhammad Fadhil Sultan Zaki, kedua Adik laki-laki saya tercinta dan tersayang, yang selalu mendukung kakak-nya untuk berjuang sampai akhir menempuh pendidikan, terimakasih pula atas doanya. Semoga kalian berdua sukses pula sekolahnya, dan menempuh pendidikan yang setinggi-tingginya serta menjadi kebanggaan Ayah dan Bunda.*
- ✓ *Siapapun dibalik perjuangan saya, sahabat, maupun teman-teman dekat saya. Terima kasih sudah memberikan cerita dan pengalaman hidup buat saya.*
- ✓ *Teman-teman seperjuangan Statistika 2014, terima kasih banyak atas kebersamaan dan dukungannya.*

KATA PENGANTAR



Assalamu'laikum Wr. Wb.

Alhamdulillah puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, serta shalawat dan salam kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “**Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistik Biner dan Naïve Bayes Classifier (Studi Kasus : Data Kelulusan Mahasiswa S1 tahun akademik 2016/2017 Universitas Islam Indonesia)**” dengan baik.

Tugas akhir ini dilakukan sebagai salah satu persyaratan yang harus dipenuhi dalam menyelesaikan Program Strata Satu di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Selama mengerjakan dan menyusun tugas akhir ini, penulis banyak memperoleh bantuan dari berbagai pihak, baik berupa saran, kritik, ataupun bimbingan. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Drs. Allwar, M.Sc., Ph.D selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
2. Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si selaku Ketua Jurusan Statistika.
3. Tuti Purwaningsih, S.Stat., M.Si selaku dosen pembimbing Tugas Akhir atas bimbingan dan kesabarannya selama menyelesaikan tugas akhir ini.
4. Seluruh dosen dan Staff Program Studi Statistika di Universitas Islam Indonesia yang telah memberikan ilmunya selama delapan semester.
5. Kedua orang tua yang selalu setia mendoakan, mendukung, menemani serta memberikan motivasi kepada penulis dalam menyelesaikan tugas akhir ini.

6. Kedua adik laki-laki penulis yang selalu mendoakan dan mendukung hingga selesai penulisan tugas akhir ini.
7. Sahabat seperjuangan: Inayatus, Nur, Riza, Gustiara, Khoffiya, Ayu dan Ina. Terimakasih atas kebersamaan dan kekeluargaan yang selalu dijaga selama di Jogja.
8. Sahabat Abangaboeh yang selalu memberi semangat: Dea, Yulinar, Arum, Caca, Widia dan Sari. Terimakasih selalu mendukung penulis dalam menyelesaikan tugas akhir ini.
9. Sahabat seperantauan: Wahyuni dan Hajrin yang selalu mendukung dan memberi semangat selama di Jogja.
10. Sahabat dan teman – teman penulis yang telah hadir dan selalu menemani saya selama 3 tahun lebih kuliah serta selalu mendoakan dan mendukung penulis dalam keadaan apapun.
11. Terima kasih untuk semua teman prodi statistika angkatan 2014 dan teman statistika lainnya yang tidak bisa penulis sebutkan satu per satu yang selalu menjadi semangat bagi penulis.
12. Pihak lain yang tidak bisa disebutkan satu per satu, terima kasih atas dukungan dan dorongan yang telah diberikan.

Semoga Allah SWT selalu memberikan rahmat dan anugerah-Nya kepada mereka semua atas segala bantuan, bimbingan, dan pengajaran yang telah diberikan kepada penulis. Penulis menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu penulis menerima kritik dan saran yang bersifat membangun dari semua pihak demi kesempurnaan tugas akhir ini. Semoga tugas akhir ini bermanfaat bagi semua pihak.

Wassalamu'alaikum Wr.Wb.

Yogyakarta, 19 April 2018

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
PERNYATAAN.....	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI.....	22
3.1 Universitas Islam Indonesia.....	22
3.2 Logo Universitas Islam Indonesia	22
3.3 Visi dan Misi Universitas Islam Indonesia.....	24
3.4 Mahasiswa	24
3.5 Kelulusan Mahasiswa	25
3.6 Statistika Deskriptif.....	25
3.7 Regresi Logistik Biner.....	26
3.7.1 Pengujian Parameter dan Maximum Likelihood untuk Regresi Logistik	28

3.7.2 Uji Kesesuaian Model	32
3.7.3 Ods Ratio	33
3.8 Variabel Dummy	33
3.9 Data Mining	34
3.10 Klasifikasi	35
3.11 <i>Naïve Bayes</i>	36
3.12 <i>Confusion Matrix</i>	39
BAB IV METODELOGI PENELITIAN.....	40
4.1 Populasi	40
4.2 Metode Pengumpulan Data dan Sumber Data	40
4.3 Variabel Penelitian	40
4.4 Metode Analisis Data	43
4.5 Tahap Analisis Data	43
BAB V PEMBAHASAN.....	46
5.1 Analisis Deskriptif	46
5.2 Analisis Regresi Logistik Biner	51
5.2.1 Uji Signifikansi Parameter	52
5.2.2 Uji Kecocokan Model.....	56
5.2.3 Model Regresi Logistik Biner	57
5.2.4 Klasifikasi Regresi Logistik Biner	61
5.3 <i>Naïve Bayes Classifier</i>	63
5.3.1 Klasifikasi Tabel <i>Naïve Bayes</i>	70
5.4 Perbandingan Kalsifikasi Regresi Logistik Biner dan <i>Naïve Bayes</i>	73
PENUTUP.....	74
6.1 Kesimpulan	74
6.2 Saran	75
LAMPIRAN.....	81

DAFTAR TABEL

Nomor	Judul	Halaman
Tabel 2.1	Tinjauan Pustaka Penelitian Sebelumnya	13
Tabel 4.1	Confusion Matrix	39
Tabel 5.1	Pembagian Data Training dan Testing Regresi Logistik Biner	52
Tabel 5.2	Uji G (Overall)	52
Tabel 5.3	Uji Wald (Partial)	55
Tabel 5.4	Uji Kecocokan Model	56
Tabel 5.5	Peluang Kejadian Lama Studi Mahasiswa	58
Tabel 5.6	Odds Ratio	60
Tabel 5.7	Prediksi Data Training Regresi Logistik	62
Tabel 5.8	Kalsifikasi Data Training Regresi Logistik Biner	62
Tabel 5.9	Prediksi Data Testing Regresi Logistik Biner	63
Tabel 5.10	Klasifikasi Data Testing Regrsi Logistik Biner	63
Tabel 5.11	Pembagian Data Training dan Data Testing Naïve Bayes	64
Tabel 5.12	Proporsi Data Training dan Testing Naïve Bayes	64
Tabel 5.13	Probabilitas Lama Studi Mahasiswa	64
Tabel 5.14	Probabilitas Lama Studi dengan Jenis Kelamin	65
Tabel 5.15	Probabilita Lama Studi dengan Asal Daerah	65
Tabel 5.16	Probabilitas Lama Studi dengan Jenis SMA	66
Tabel 5. 17	Probabilitas Lama Studi dengan Jurusan Saat SMA	66
Tabel 5. 18	Probabilitas Lama Studi dengan IPK	67
Tabel 5. 19	Probabilitas Lama Studi dengan Fakultas	67
Tabel 5. 20	Probabilitas Lama Studi dengan Pekerjaan Orang Tua	68
Tabel 5.21	Prediksi dari Data Training Naïve Bayes	71
Tabel 5.22	Klasifikasi Tabel Data Training Naïve Bayes	71
Tabel 5.23	Prediksi Data Testing Naïve Bayes	72

Nomor	Judul	Halaman
Tabel 5. 24	Prediksi Data Testing Naïve Bayes	72
Tabel 5.25	Perbandingan Klasifikasi Regresi Logistik dan Naïve Bayes	73

DAFTAR GAMBAR

Nomor	Judul	Halaman
Gambar 1.1	Ketepatan Waktu Kelulusan Mahasiswa Tahun Aademik 2016/207	2
Gambar 3. 1	Logo Universitas Islam Indonesia	23
Gambar 3.2	Langkah-langkah untuk melakukan Data Mining	34
Gambar 5. 1	Persentase Lama Studi Mahasiswa Tahun Akademik 2016/2017	46
Gambar 5. 2	Lama Studi Mahasiswa Berdasarkan Jenis Kelamin	47
Gambar 5. 3	Lama Studi Mahasiswa Berdasarkan Asal Daerah	47
Gambar 5. 4	Lama Studi Mahasiswa Berdasarkan Jenis SMA	48
Gambar 5. 5	Lama Studi Mahasiswa Berdasarkan Jurusan SMA	49
Gambar 5. 6	Lama Studi Mahasiswa Berdasarkan IPK	49
Gambar 5. 7	Lama Studi Mahasiswa Berdasarkan Fakultas	50
Gambar 5. 8	Lama Studi Mahasiswa Berdasarkan Pekerjaan Orang Tua	51

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian	82
Lampiran 2. Sintaks Regresi Logistik Biner	84
Lampiran 3 Sintaks Naïve Bayes Classifier	86
Lampiran 4. Output Regresi Logistik Biner	88
Lampiran 5. Output Naïve Bayes Classifier	90

PERNYATAAN

Dengan ini penulis menyatakan bahwa dalam tugas akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepengetahuan penulis juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 19 April 2018



Penulis

**PERBANDINGAN KLASIFIKASI KETEPATAN WAKTU KELULUSAN
MAHASISWAMENGGUNAKAN REGRESI LOGISTIK BINER DAN *NAÏVE
BAYES CLASSIFIER***

(Studi kasus : Data Kelulusan Mahasiswa S1 tahun akademik 2016/2017 Universitas
Islam Indonesia)

Erene Fajrila

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

INTISARI

Pendidikan merupakan salah satu kebutuhan manusia yang sangat penting, karena pendidikan memiliki tugas untuk menyiapkan sumber daya manusia bagi pembangunan bangsa dan negara. Dalam pembangunan bangsa dan negara tersebut, perlu adanya peningkatan dari pendidikan yaitu kualitas pendidikan. Kualitas pendidikan dapat dilihat dari suatu pencapaian pembelajaran orang tersebut. Salah satu pencapaian pembelajaran orang tersebut dapat dilihat dari kelulusan mahasiswa selama menempuh pendidikan. Kelulusan adalah hasil akhir pencapaian seseorang dalam menempuh suatu pendidikan, dimana lama studi seorang mahasiswa secara umum ditempuh waktu dalam 8 semester atau 4 tahun. Tetapi masih banyak mahasiswa yang menempuh waktu lebih dari 4 tahun. Universitas Islam Indonesia merupakan salah satu Universitas Swasta tertua di Indonesia, dimana memiliki jumlah mahasiswa maupun alumni yang sangat banyak. Terdapat banyak faktor-faktor yang mempengaruhi lama studi mahasiswa dalam menempuh suatu pendidikan. Tujuan dari penelitian ini adalah untuk mengetahui faktor-faktor yang berpengaruh secara signifikan terhadap lama studi mahasiswa Universitas Islam Indonesia dan melakukan klasifikasi ketepatan waktu kelulusan mahasiswa menggunakan Regresi Logistik Biner serta melakukan perbandingan klasifikasi dengan Algoritma *Naïve Bayes Classifier*, karena data yang digunakan cukup besar. Berdasarkan analisis Regresi Logistik Biner didapatkan faktor-faktor yang berpengaruh secara signifikan terhadap lama studi mahasiswa yaitu, variabel Jenis Kelamin, Asal Daerah, Jenis SMA, Jurusan SMA, IPK mahasiswa. Klasifikasi menggunakan Regresi Logistik Biner menggunakan variabel yang signifikan tersebut mendapatkan akurasi sebesar 82,29%, sedangkan hasil klasifikasi menggunakan Algoritma *Naïve Bayes* sebesar 89,10%. Hal ini dapat dikatakan pengklasifikasian menggunakan *Naïve Bayes Classifier* lebih baik dibandingkan Regresi Logistik Biner.

Kata Kunci : *Pendidikan, Faktor, Klasifikasi, Lama Studi, Regresi Logistik Biner, Naïve Bayes Classifier*

**COMPARISON CLASSIFICATION OF TIME GRADUATION STUDENTS
USING BINARY LOGISTIC REGRESSION AND NAÏVE BAYES CLASSIFIER**

(Case study: undergraduate student Graduation Data for the academic in 2016/2017
Islamic University of Indonesia)

Erene Fajrila

Department of Statistics Faculty of Mathematics and Science

Islamic University of Indonesia

ABSTRACT

Education is one of the most important human needs, because education has the task to prepare human resources for the development of the nation and the country. In the development of the nation and the country, the need for improvement of education, namely quality of education. The quality of education can be seen from a person's learning achievement. One of the achievements of the person's learning can be seen from the graduation of the students during their study, and then the graduation is an achievement of the student, which the education periods of the students is generally taken for 8 semesters or 4 years, but there are still many students who finished their study more than 4 years. Islamic University of Indonesia is the oldest private universities in Indonesia, with a large number of the students and alumni. There are many factors who affect the periods of the student to finish their study. The purpose of this research is to know the factors that influence significantly the education periods for the students of Islamic University of Indonesia and conducting classification of the student's graduation using Binary Logistic Regression, and then conducting classification comparison with the classification of algorithms Naïve Bayes Classifier, because the data used is quiet large. Based on the Binary Logistic Regression, the factors that influence significantly to education period of the students are variables gender, origin, type of area high school, the high school department, grade-Pont average student. Classification of Binary Logistic Regression using the significant variables that got the accuracy of 82.29%, while the results of the classification using Algorithm Naïve Bayes Classifier is 89.10%. It can be said that classification using Algorithm of Naïve Bayes Classifier is better than Binary Logistic Regression.

Keywords: *Education, Factor, Classification, The Education Periods, Binary Logistic Regression, Naïve Bayes Classifier*

BAB I

PENDAHULUAN

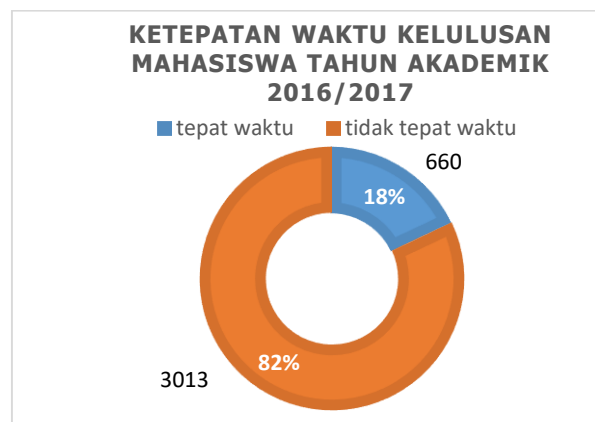
1.1 Latar Belakang Masalah

Pendidikan merupakan kebutuhan manusia yang sangat penting karena pendidikan mempunyai tugas untuk menyiapkan SDM bagi pembangunan bangsa dan Negara. Menurut Syah dalam Chandra (2009), dikatakan bahwa pendidikan berasal dari kata dasar “didik” yang mempunyai arti memelihara dan memberi latihan. Kedua hal tersebut memerlukan adanya ajaran, tuntunan, dan pimpinan tentang kecerdasan pikiran. Pengertian pendidikan adalah proses perubahan sikap dan perilaku seseorang atau sekelompok orang dalam usaha mendewasakan manusia melalui upaya pengajaran dan pelatihan. Dengan melihat definisi tersebut, sebagian orang mengartikan bahwa pendidikan adalah pengajaran karena pendidikan pada umumnya membutuhkan pengajaran dan setiap orang berkewajiban mendidik. Secara sempit mengajar adalah kegiatan secara formal menyampaikan materi pelajaran sehingga peserta didik menguasai materi ajar.

Pendidikan Tinggi adalah jenjang pendidikan setelah pendidikan menengah yang mencakup program diploma, program sarjana, program magister, program doktor, dan program profesi, serta program spesialis, yang diselenggarakan oleh perguruan tinggi berdasarkan kebudayaan bangsa Indonesia (UU No 12, 2012) tentang Pendidikan Tinggi. Kemudian, untuk memenuhi standar kompetensi lulusan bagi mahasiswa program sarjana (S1) beban wajib yang harus ditempuh adalah paling sedikit 144-160 satuan kredit semester (sks). Lama masa studi mahasiswa program sarjana (S1) secara normal menurut kurikulum adalah 8 semester atau selama 4 tahun. Namun faktanya tidak sedikit mahasiswa yang menempuh masa studi melebihi standar kelulusan yang telah ditetapkan.

Dalam dunia pendidikan, setiap perguruan tinggi memiliki kewajiban untuk mengontrol prestasi belajar setiap mahasiswanya dan menghasilkan lulusan yang berkualitas. Seluruh perguruan tinggi juga dituntut untuk menjamin mutu lulusan, dimana mutu ini dimaksudkan bahwa lulusan perguruan tinggi dapat langsung dimanfaatkan oleh *stakeholders*.

Universitas Islam Indonesia (UII) adalah salah satu perguruan tinggi swasta terkemuka di Indonesia. Terinspirasi oleh semangat nasionalisme dan berpedoman pada nilai-nilai *perennial*, UII didirikan satu bulan sebelum proklamasi kemerdekaan Indonesia pada tahun 1945. Saat ini, UII telah tumbuh berkembang menjadi tempat pembelajaran yang terpilih. Berlokasi di utara Yogyakarta, jantung kebudayaan Jawa, kampus utama UII tepat menghadap ke keindahan Gunung Merapi yang menakjubkan, yang menjadikan UII sebagai tempat yang tepat untuk belajar. Pada tahun akademik 2017/2018, UII memiliki 8 Fakultas, dengan 4 (empat) program Diploma Tiga, 25 (dua puluh lima) program Sarjana. Setiap tahun, Universitas Islam Indonesia mengadakan perayaan wisuda untuk mahasiswa yang dinyatakan lulus, tetapi masih banyak mahasiswa Universitas Islam Indonesia di 8 Fakultas tersebut yang lulus tidak tepat waktu.



Gambar 1.1 Ketepatan Waktu Kelulusan Mahasiswa Tahun Akademik 2016/2017

Pada tahun akademik 2016/2017 Universitas Islam Indonesia sudah memiliki jumlah lulusan 3.673 mahasiswa. Jumlah tersebut cukup besar bagi Universitas yang merupakan Universitas Swasta tertua di Indonesia setiap tahunnya. Dari **gambar 1.1**

diasas dapat dilihat bahwa, masih banyak mahasiswa Universitas Islam Indonesia yang lulus tidak tepat waktu. Oleh karena itu, Universitas Islam Indonesia memerlukan tindakan yang tepat untuk mengetahui faktor-faktor yang mempengaruhi lama studi mahasiswa tersebut. Untuk mengetahui faktor-faktor yang berpengaruh secara signifikan pada lama studi mahasiswa, maka perlunya dilakukan analisis regresi. Analisis regresi digunakan untuk mengetahui pengaruh dan mengukur hubungan antara variabel respon dan variabel prediktor. Salah satu teknik regresi adalah analisis Regresi Logistik Biner. Regresi Logistik Biner adalah salah satu metode regresi yang digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Regresi Logistik Biner tidak mengasumsikan hubungan linier antar variabel bebas dan terikat dikarenakan bentuk variabel terikat yang bersifat kategorik. Dalam Regresi Logistik Biner, model yang didapat dengan respon kualitatif pada variabel dependen bertujuan untuk menentukan probabilitas individu dalam keputusan yang bersifat kualitatif. Ada banyak faktor yang mempengaruhi lama (masa) studi mahasiswa dalam menempuh suatu pendidikan. Dalam penelitian ini diasumsikan bahwa faktor-faktor yang mempengaruhi lama studi mahasiswa antara lain jenis kelamin, asal daerah, jenis SMA, jurusan saat SMA, fakultas, IPK, serta Pekerjaan Orang Tua.

Selain itu, dikarenakan data mahasiswa yang ada cukup besar hingga mencapai ribuan untuk data kelulusan mahasiswa dalam satu tahunnya, maka penting untuk menggali informasi-informasi berharga dalam data tersebut. Kemudian, untuk membantu dalam menemukan informasi-informasi berharga tersebut diperlukan adanya teknik *data mining*.

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat

yang tersimpan di dalam database besar (Turban et al, 2005). Salah satu teknik *data mining* adalah pengklasifikasian. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek (Bustami, 2014).

Ada beberapa metode yang dapat digunakan dalam pengklasifikasian dari suatu objek penelitian, yakni Regresi Logistik Biner dan *Naïve Bayes*. Selain mencari hubungan antara 1 variabel respon dengan beberapa variabel bebas, Regresi Logistik Biner mampu mengklasifikasi suatu data, dimana variabel dependen yang bersifat kategorik. Metode Regresi Logistik Biner adalah salah satu dari sekian banyak metode yang dapat digunakan untuk menyelesaikan berbagai jenis permasalahan termasuk klasifikasi. Dalam metode Regresi Logistik Biner mampu menjelaskan faktor-faktor yang mempengaruhi variabel dependen dan pengklasifikasian dari suatu data (Samosir, 2015).

Naïve Bayes Classifier merupakan salah satu teknik klasifikasi. Klasifikasi *Naïve Bayes* adalah sebuah metode penggolongan berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen). Pada klasifikasi *Naïve Bayes*, proses pembelajaran lebih ditekankan pada mengestimasi probabilitas. Keuntungan dari pendekatan ini yaitu pengklasifikasian akan mendapatkan nilai error yang lebih kecil ketika data set berjumlah besar (Berry, 2006). Selain itu menurut Han and Kamber (2006) klasifikasi *Naïve Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar.

Berdasarkan latar belakang yang telah di uraikan diatas, maka dilakukan penelitian yang berjudul “Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistik Biner dan *Naïve Bayes Classifier*”. (Studi Kasus : Data Kelulusan Mahasiswa Strata Satu Universitas Islam Indonesia tahun akademik 2016/2017).

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini, meliputi :

1. Bagaimana gambaran umum dari faktor-faktor yang mendukung ketepatan waktu kelulusan mahasiswa Strata satu Universitas Islam Indonesia tahun akademik 2016/2017?
2. Apa saja faktor-faktor yang mempengaruhi lama (masa) studi mahasiswa Universitas Islam Indonesia menggunakan regresi logistik biner dan bagaimana klasifikasi kelulusan mahasiswa Strata satu Universitas Islam Indonesia menggunakan Regresi Logistik Biner?
3. Bagaimana klasifikasi status kelulusan mahasiswa Strata satu Universitas Islam Indonesia di seluruh fakultas dengan menggunakan metode *Naïve Bayes Classifier*?
4. Bagaimana Perbandingan Klasifikasi Regresi Logistik Biner dan *Naïve Bayes Classifier*?

1.3 Batasan Masalah

Adapun batasan-batasan masalah yang diberikan, agar penelitian ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi sebagai berikut :

1. Data yang digunakan adalah data Kelulusan Mahasiswa Strata Satu tahun akademik 2016/2017 di Universitas Islam Indonesia. Data diperoleh dari *database* Badan Sistem Informasi, Akademik Rektorat Universitas Islam Indonesia.
2. Variabel Dependen (y) antara lain mahasiswa Tepat Waktu dan Tidak Tepat Waktu. Sedangkan variabel independen (x) yang digunakan meliputi Jenis Kelamin, Asal Daerah, Jenis SMA, Jurusan SMA, IPK, Fakultas dan Pekerjaan Orang Tua.
3. Analisis yang digunakan adalah Regresi Logistik Biner dan *Naïve Bayes Classifier*.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini, sebagai berikut:

1. Untuk mengetahui gambaran umum dari faktor-faktor yang mendukung ketepatan waktu kelulusan mahasiswa Strata satu Universitas Islam Indonesia tahun akademik 2016/2017.
2. Untuk mengetahui faktor-faktor yang mempengaruhi masa (lama) studi mahasiswa Universitas Islam Indonesia dan klasifikasi ketepatan waktu kelulusan mahasiswa Strata satu Universitas Islam Indonesia dengan menggunakan Regresi Logistik Biner.
3. Untuk mengetahui klasifikasi ketepatan waktu kelulusan mahasiswa Strata satu Universitas Islam Indonesia dengan menggunakan metode *Naïve Bayes Classifier*.
4. Untuk mengetahui perbandingan klasifikasi Regresi Logistik Biner dan *Naïve Bayes Classifier*.

1.5 Manfaat Penelitian

Berdasarkan tujuan yang ada, maka diperoleh manfaat dari penelitian ini, sebagai berikut :

1. Dengan mengetahui faktor-faktor yang mempengaruhi masa (lama) studi mahasiswa Universitas Islam Indonesia, maka memberikan informasi untuk mengambil suatu keputusan dengan lebih meningkatkan kembali kualitas pendidikan yang ada di Universitas Islam Indonesia.
2. Dengan didapatkannya hasil pelatihan maka dapat diketahui seberapa baik hasil pembelajaran yang diberikan pada data, semakin baik hasil pelatihan maka semakin baik pula hasil pengujian.
3. Sebagai pengembangan dan pengaplikasian ilmu statistika, khususnya tentang Regresi Logistik Biner dan *Naïve Bayes Classifier*.

BAB II

TINJAUAN PUSTAKA

Penelitian terdahulu sangat penting menjadi landasan untuk penulis melakukan penelitian baru, supaya dapat diketahui hubungan antara penelitian sebelumnya dengan penelitian yang dilakukan oleh penulis, agar tidak terjadi plagiasi sehingga, dapat diketahui kontribusi penelitian ini terhadap perkembangan ilmu pengetahuan. Penelitian dengan menggunakan klasifikasi Regresi Logistik Biner dan Naïve Bayes serta dengan topic pendidikan pada umumnya telah banyak dilakukan oleh beberapa peneliti-peneliti sebelumnya. Adapun beberapa penelitian yang berhubungan dengan penelitian ini sebagai berikut.

Penelitian oleh Bustami (2013) dengan judul “Penerapan Algoritma *Naïve Bayes* Untuk Mengklasifikasikan Data Nasabah Asuransi”. Data yang digunakan adalah data nasabah dengan beberapa variabel yaitu jenis kelamin, usia, status, pekerjaan, penghasilan, cara pembayaran premi dan masa pembayaran premi. Adapun tujuan dari penelitian ini yaitu mengetahui klasifikasi Nasabah Asuransi menggunakan Naïve Bayes. Penelitian ini menghasilkan sistem klasifikasi nasabah dengan klasifikasi lancar, kurang lancar atau tidak lancarnya nasabah dalam membayar premi asuransi dengan menggunakan algoritma *Naïve Bayes*.

Penelitian selanjutnya oleh Yuni (2013) dengan judul “Klasifikasi Keputusan Nasabah Dalam Pengambilan Kredit Menggunakan Model Regresi Logistik Biner dan Metode *Classification And Regression Trees (Chart)*”. Tujuan penelitian ini yakni pengklasifikasian keputusan nasabah dalam pengambilan kredit untuk melihat karakteristik dan faktor-faktor yang berpengaruh secara signifikan di bank bjb Cabang Utama Bandung. Hasil penelitian ini adalah variabel-variabel yang berpengaruh secara signifikan terhadap keputusan

nasabah dalam pengambilan kredit dengan Regresi Logistik Biner yakni variabel usia, pendidikan, informasi kredit dan prosedur kredit. Dari model Regresi Logistik Biner mendapatkan hasil akurasi sebesar 72,8%. Selain itu penelitian ini juga menghasilkan pohon optimum dengan dua simpul terminal untuk metode pohon klasifikasi *cart*. Variabel prediktor yang masuk dalam pohon klasifikasi adalah variabel prosedur kredit. Hasil dari pohon klasifikasi *cart* menghasilkan akurasi sebesar 66,8%. Klasifikasi dengan Regresi Logistik Biner lebih cocok digunakan untuk kasus klasifikasi keputusan nasabah dalam pengambilan kartu kredit karena nilai akurasi nya lebih tinggi dibandingkan dengan pohon klasifikasi *cart*.

Kemudian penelitian oleh Mauriza (2014) dengan judul “Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan *Naïve Bayes*”. Tujuan dari penelitian ini adalah untuk memprediksi masa studi mahasiswa di Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta menggunakan *Naïve Bayes*. Diketahui bahwa mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta memiliki jumlah 2358 mahasiswa dan memiliki toleransi ketidakteelitian 5%. Maka jumlah sampel yang diambil yaitu 342 mahasiswa. Hasil penelitian ini adalah mendapatkan mahasiswa sebesar 25,15% dinyatakan lulus tepat waktu atau sebanyak 86 mahasiswa, sedangkan yang lulus tidak tepat waktu mencapai 74,85% atau sebanyak 256 mahasiswa.

Selanjutnya penelitian oleh Fais dkk (2014), dengan judul “Klasifikasi Calon Pendorong Darah dengan Metode *Naïve Bayes*”. Tujuan penelitian ini adalah untuk menentukan peluang kemungkinan besar seseorang dapat mendonorkan darah. Data *training* pendonor sebanyak 75 dan non pendonor sebanyak 100, sedangkan data testing pendonor sebanyak 25 dan non pendonor sebanyak 25 dengan persentase yang dihasilkan sebesar 84%. Jika data *training* pendonor dan non pendonor sebesar 125 dan 100 serta untuk data testing pendonor maupun non pendonor sebesar 25, maka menghasilkan presentase sebesar 58%. Jika data training pendonor dan non pendonor masing-masing 125 serta data testing yang digunakan masing-masing 25, maka

presentase sebesar 78%. Jika untuk data training, ketika pendonor sebanyak 200 dan non pendonor sebanyak 150 serta data testing yang digunakan untuk pendonor dan non pendonor masing-masing sebanyak 25, menghasilkan presentase sebesar 76%. Maka rata-rata akurasi dari keempat percobaan tersebut sebesar 74%.

Selanjutnya penelitian oleh Diaprina dan Suhartono (2014) dari Institut Teknologi Sepuluh Nopember, dengan judul “Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner dan *Radial Basis Funtion Network* di Bank “X” Cabang Kediri”. Data yang diunakan merupakan data sekunder yang diambil di Bank X Cabang Kediri yaitu data nasabah. Tujuannya untuk memprediksi klasifikasi risiko kredit nasabah di Bank X Cabang Kediri menggunakan Regresi Logistik Biner dan *Radial Basis Function Network*. Hasil penelitian ini mendapatkan ketepatan klasifikasi menggunakan metode regresi logistik biner dengan model terbaik adalah sebesar 99,38% untuk data dengan jumlah kredit macet sebanyak 6 dan 99,76% untuk data dengan jumlah kredit macet sebanyak 8. Rata-rata ketepatan klasifikasi untuk metode *Radial Basis Function Network* adalah sebesar 99,03% untuk data dengan jumlah kredit macet sebanyak 6 dan 98,51% untuk data dengan jumlah kredit macet sebanyak 8. Metode yang lebih baik digunakan untuk kasus klasifikasi adalah dengan menggunakan regresi logistik biner karena memiliki nilai ketepatan klasifikasi yang lebih besar dibandingkan dengan menggunakan *Radial Basis Function Network*.

Selanjutnya penelitian oleh Rajagukguk (2015), dengan judul “Perbandingan Metode Klasifikasi Regresi Logistik Biner dan Naïve Bayes Pada Status Pengguna KB di Kota Tegal Tahun 2014”. Tujuan dari penelitian ini adalah mendapatkan model dan ketepatan klasifikasi menggunakan Regresi Logistik Biner dan Naïve Bayes serta membandingkan ketepatan klasifikasi Regresi Logistik Biner dan Naïve Bayes pada status pengguna KB di Kota Tegal. Hasil penelitian ini mendapatkan ketepatan klasifikasi dengan metode Regresi Logistik Biner sebesar 83,33 % sedangkan ketepatan klasifikasi dengan metode Naïve Bayes sebesar 81,75 %. Pengklasifikasian

menggunakan metode Regresi Logistik Biner lebih baik dibandingkan dengan metode Naïve Bayes.

Penelitian di bidang Pendidikan oleh Muin (2016), dengan judul “Metode *Naïve Bayes* Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)”. Tujuannya untuk melakukan prediksi peluang kelulusan mahasiswa baru. Hasil penelitian ini mendapatkan nilai persentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naïve Bayes* mencapai 94 %.

Penelitian oleh Rumaenda (2016), dari Universitas Diponegoro dengan judul “Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner dan Algoritma C4.5”. Tujuan dari penelitian ini adalah menentukan faktor yang mempengaruhi penyakit hipertensi di UPT Puskesmas Ponjong menggunakan model regresi logistik biner serta mengetahui ketepatan klasifikasi menggunakan Regresi Logistik Biner dan Algoritma C4.5 kemudian membandingkan ketepatan klasifikasi tersebut. Data yang digunakan adalah data sekunder, yaitu data pasien hipertensi di UPT Puskesmas Ponjong 1, Gunungkidul Hasil penelitian mendapatkan faktor-faktor yang mempengaruhi terjadinya jenis penyakit hipertensi yakni jenis kelamin, tekanan darah sistolik, serta penyakit lain menggunakan Regresi Logistik Biner serta memperoleh ketepatan klasifikasi sebesar 72,5352%. Kemudian untuk algoritma C4.5 mendapatkan pohon keputusan yang terbentuk menghasilkan pohon sebanyak 13 simpul, yang terdiri dari sebuah simpul akar yaitu atribut tekanan darah sistolik, 5 simpul keputusan, serta 7 simpul daun dengan faktor-faktor yang mempengaruhi terjadinya jenis penyakit hipertensi adalah tekanan darah sistolik, riwayat berobat, penyakit lain, tekanan darah diastolik, dan jenis kelamin, sedangkan faktor umur tidak berpengaruh dan mendapatkan ketepatan klasifikasi sebesar sebesar 64,0845%. Dari hasil tersebut dapat dikatakan bahwa regresi logistik biner lebih baik dibandingkan algoritma C4.5 dalam mengklasifikasikan jenis penyakit hipertensi di UPT Puskesmas Ponjong I, Gunungkidul.

Kemudian penelitian oleh Ma'arif (2016), dari SMIK Jenderal Achmad Yani Yogyakarta dengan judul “Perbandingan *Naïve Bayes Classifier* dan *Support Vector Mechine* Untuk Klasifikasi Judul Artikel”. Tujuan dari penelitian ini adalah untuk membandingkan hasil klasifikasi judul artikel menggunakan *Naïve Bayes Classifier* dan *Support Vector Mechine*. Data yang digunakan dalam penelitian ini adalah data judul untuk artikel-artikel tentang agama Islam yang diperoleh dari berbagai website. Dataset terdiri dari 1827 judul artikel *unique* yang dibagi menjadi 1600 data untuk *training* dan 274 data untuk *testing*. Penelitian ini menghasilkan Klasifikasi *Naïve Bayes Classifier* menggunakan fitur *trigram* dan *term frequency* memiliki nilai akurasi yang paling tinggi yaitu 81%. Sementara itu, model yang dibangun dengan algoritma SVM terbaik 73% dengan menggunakan fitur *unigram*. Dapat dikatakan bahwa klasifikasi yang dihasilkan oleh *Naïve Bayes Classifier* memiliki performa yang lebih tinggi dibandingkan dengan model yang dihasilkan dari yang dihasilkan dari algoritma *Support Vector Mechine*.

Kemudian penelitian oleh Fatmawati (2016), dari STMIK Nusa Mandiri Jakarta dengan judul “Perbandingan Algoritma Klasifikasi *Data Mining* Model C4.5 dan *Naïve Bayes* Untuk Prediksi Penyakit Diabetes”. Tujuan dari penelitian ini adalah untuk mengetahui diantara model algoritma C4.5 dan *Naïve Bayes* yang memiliki akurasi paling tinggi untuk memprediksi penyakit diabetes. Hasil penelitian tersebut didapatkan bahwa klasifikasi menggunakan C4.5 memperoleh akurasi sebesar 73,30% dan nilai AUC sebesar 0,733, sedangkan nilai akurasi *Naïve Bayes* sebesar 75,13% dan nilai AUC adalah 0,810. Dapat disimpulkan bahwa dengan menggunakan model *Naïve Bayes* lebih tinggi tingkat akurasinya, dengan peningkatan akurasi sebesar 1.83% dan peningkatan nilai AUC sebesar 0.077 dan hasil pengujian dari prediksi diabetes hasilnya termasuk *Good Clasification*

Selanjutnya penelitian oleh Ramandhani (2017), dari Universitas Diponegoro, dengan judul “Metode *Bootstrap Aggregating* Regresi Logistk Biner Untuk Ketepatan Klasifikasi Kesejahteraan Rumah Tangga Di Kota Pati”. Tujuan penelitian ini adalah

mengetahui model dan ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati berdasarkan analisis Regresi Logistik Biner dan mengetahui ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati dengan metode *Bootstrap Aggregating (Bagging)* Regresi Logistik Biner. Hasil penelitian ini memperoleh ketepatan klasifikasi sebesar 79,55% untuk Regresi Logistik Biner, sedangkan hasil analisis *Bagging* Regresi Logistik biner menunjukkan bahwa pada 90 kali replikasi *bootstrap* diperoleh nilai ketepatan klasifikasi terbesar yaitu 79,87%.

Kemudian penelitian oleh Fajri (2018), dari Sekolah Tinggi Teknologi Nurul Jadid dengan judul “Perbandingan Sistem Klasifikasi *Naïve Bayes* dan *Decision Tree* Untuk Diagnosa Penyakit Diabetes”. Tujuan dari penelitian ini adalah untuk melakukan proses *comparasi* atau membandingkan antara metode *Decision Tress* dan *Naïve Bayes* dalam klasifikasi diagnosa penyakit diabetes sebagai sistem keputusan yang digunakan untuk membantu mendiagnosa penyakit diabetes. Hasil dari penelitian ini memperoleh model klasifikasi dengan menggunakan *Desicion Tree* mempunyai nilai optimal akurasi sebesar 71.75 % dengan menggunakan split rasio 30 % sedangkan dengan menggunakan *Naïve Bayes* mempunyai nilai optimal akurasi sebesar 75.28 % dengan split rasio 30 %. Hal ini menunjukkan bahwa metode yang mempunyai nilai akurasi yang paling optimal adalah metode naive bayes untuk klasifikasi diagnosa penyakit diabetes.

Perbedaan penelitian tugas akhir ini dengan penelitian terdahulu yaitu penelitian ini menggunakan data kelulusan mahasiswa Strata Satu Universitas Islam Indonesia tahun akademik 2016/2017 dan menggunakan metode analisis Regresi Logistik Biner dan *Naïve Bayes Classifier*. Penelitian ini bertujuan untuk mengetahui faktor-faktor yang berpengaruh secara signifikan terhadap lama studi mahasiswa Universitas Islam Indonesia dan mengetahui akurasi dari klasifikasi lama studi mahasiswa Universitas Islam Indonesia serta mengetahui perbandingan dari klasifikasi metode Regresi Logistik Biner dan Algoritma *Naïve Bayes Classifier*.

Dari uraian tersebut dapat, dilihat melalui tabel 2.1 tentang tinjauan pustaka penelitian sebelumnya.

Tabel 2.1 Tinjauan Pustaka Penelitian Sebelumnya

No	Penelitian	Metode	Tujuan	Hasil
1	Bustami (2013), Dosen Teknik Informatika Universitas Malikussaleh.	Teknik klasifikasi dengan Algoritma Naïve Bayes	Adapun tujuan dari penelitian ini yaitu mengetahui klasifikasi Nasabah Asuransi menggunakan Naïve Bayes.	Penelitian ini menghasilkan sistem klasifikasi nasabah dengan klasifikasi lancar, kurang lancar atau tidak lancarnya nasabah dalam membayar premi asuransi dengan menggunakan algoritma <i>Naïve Bayes</i> .
2	Yuni (2013), Universitas Pendidikan Indonesia.	Regresi Logistik Biner dan Metode <i>Classification And Regression Trees (Chart)</i>	Tujuan penelitian ini yakni pengklasifikasian keputusan nasabah dalam pengambilan kredit untuk melihat karakteristik dan faktor-faktor yang berpengaruh secara signifikan di bank bjb Cabang Utama Bandung.	Hasil penelitian ini adalah variabel-variabel yang berpengaruh secara signifikan terhadap keputusan nasabah dalam pengambilan kredit dengan Regresi Logistik Biner yakni variabel usia, pendidikan, informasi kredit dan prosedur kredit. Dari model Regresi Logistik Biner mendapatkan hasil akurasi sebesar 72,8%. Selain itu penelitian ini juga menghasilkan pohon optimum dengan dua simpul terminal untuk metode pohon klasifikasi <i>cart</i> . Variabel prediktor yang

No	Penelitian	Metode	Tujuan	Hasil
				masuk dalam pohon klasifikasi adalah variabel prosedur kredit. Hasil dri pohon klasifikasi <i>cart</i> menghasilkan akurasi sebesar 66,8%. Klasifikasi dengan Regresi Logistik Biner lebih cocok digunakan untuk kasus klasifikasi keputusan nasabah dalam pengambilan kartu kredit karena nilai akurasi nya lebih tinggi dibandingkan dengan pohon klasifikasi <i>cart</i> .
3	Mauriza (2014), Universitas Muhammadiyah Surakarta.	Teknik klasifikasi dengan <i>Naïve Bayes</i> .	Tujuan dari penelitian ini adalah untuk memprediksi masa studi mahasiswa di Fakultas Komunikasi dan Informatika Universita Muhammadiyah Surakarta menggunakan <i>Naïve Bayes</i> .	Diketahui bahwa mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta memiliki jumlah 2358 mahasiswa dan memiliki toleransi ketidakteelitian 5%. Maka jumlah sampel yang diambil yaitu 342 mahasiswa. Hasil penelitian ini adalah mendapatkan mahasiswa sebesar 25,15% dinyatakan lulus tepat waktu atau sebayak 86 mahasiswa, sedangkan yang lulus

No	Penelitian	Metode	Tujuan	Hasil
				tidak tepat waktu mencapai 74,85% atau sebanyak 256 mahasiswa.
4	Fais, dkk (2014), Universitas Brawijaya.	Teknik klasifikasi dengan Regresi Logistik Biner dan Naïve Bayes	Tujuan penelitian ini adalah untuk menentukan peluang kemungkinan besar seseorang dapat mendonorkan darah.	Data <i>training</i> pendonor sebanyak 75 dan non pendonor sebanyak 100, sedangkan data testing pendonor sebanyak 25 dan non pendonor sebanyak 25 dengan persentase yang dihasilkan sebesar 84%. Jika data <i>training</i> pendonor dan non pendonor sebesar 125 dan 100 serta untuk data testing pendonor maupun non pendonor sebesar 25, maka menghasilkan presentase sebesar 58%. Jika data training pendonor dan non pendonor masing-masing 125 serta data testing yang digunakan masing- masing 25, maka presentase sebesar 78%. Jika untuk data training, ketika pendonor sebanyak 200 dan non pendonor sebanyak 150 serta data testing yang digunakan untuk pendonor dan non pendonor masing-masing sebanyak 25,

No	Penelitian	Metode	Tujuan	Hasil
				menghasilkan presentase sebesar 76%. Maka rata-rata akurasi dari keempat percobaan tersebut sebesar 74%.
5	Diaprina dan Suhartono (2014), Institut Teknologi Sepuluh Nopember.	Teknik klasifikasi dengan Regresi Logistik Biner dan <i>Radial Basis Function Network</i> .	Tujuannya untuk memprediksi klasifikasi risiko kredit nasabah di Bank X Cabang Kediri menggunakan Regresi Logistik Biner dan <i>Radial Basis Function Network</i> .	Hasil penelitian ini mendapatkan ketepatan klasifikasi menggunakan metode regresi logistik biner dengan model terbaik adalah sebesar 99,38% untuk data dengan jumlah kredit macet sebanyak 6 dan 99,76% untuk data dengan jumlah kredit macet sebanyak 8. Rata-rata ketepatan klasifikasi untuk metode <i>Radial Basis Function Network</i> adalah sebesar 99,03% untuk data dengan jumlah kredit macet sebanyak 6 dan 98,51% untuk data dengan jumlah kredit macet sebanyak 8. Metode yang lebih baik digunakan untuk kasus klasifikasi adalah dengan menggunakan regresi logistik biner karena memiliki nilai ketepatan klasifikasi yang lebih besar

No	Penelitian	Metode	Tujuan	Hasil
				dibandingkan dengan menggunakan <i>Radial Basis Function Network</i> .
6	Rajagukguk (2015), Universitas Diponegoro	Teknik klasifikasi dengan Regresi Logistik Biner dan <i>Naïve Bayes Classifier</i>	Tujuan dari penelitian ini adalah mendapatkan model dan ketepatan klasifikasi menggunakan Regresi Logistik Biner dan <i>Naïve Bayes</i> serta membandingkan ketepatan klasifikasi Regresi Logistik Biner dan <i>Naïve Bayes</i> pada status pengguna KB di Kota Tegal.	Hasil penelitian ini mendapatkan ketepatan klasifikasi dengan metode Regresi Logistik Biner sebesar 83,33 % sedangkan ketepatan klasifikasi dengan metode <i>Naive Bayes</i> sebesar 81,75 %. Pengklasifikasian menggunakan metode Regresi Logistik Biner lebih baik dibandingkan dengan metode <i>Naïve Bayes</i> .
7	Muin (2016), Universitas Al Asyariah Mandar.	Teknik Klasifikasi Menggunakan <i>Naïve Bayes</i> .	Tujuannya untuk melakukan prediksi peluang kelulusan mahasiswa baru.	Hasil penelitian ini mendapatkan nilai persentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa

No	Penelitian	Metode	Tujuan	Hasil
				Baru yang diterapkan ke dalam metode <i>Naïve Bayes</i> mencapai 94 %.
8	Rumaenda (2016), Universitas Diponegoro.	Teknik Klasifikasi menggunakan Regresi Logistik Biner dan ALgoritma C4.5	Tujuan dari penelitian ini adalah menentukan faktor yang mempengaruhi penyakit hipertensi di UPT Puskesmas Ponjong menggunakan model regresi logistik biner serta mengetahui ketepatan klasifikasi menggunakan Regresi Logistik Biner dan Algoritma C4.5 kemudian membandingkan ketepatan klasifikasi tersebut.	Hasil penelitian mendapatkan faktor-faktor yang mempengaruhi terjadinya jenis penyakit hipertensi yakni jenis kelamin, tekanan darah sistolik, serta penyakit lain menggunakan Regresi Logistik Biner serta memperoleh ketepatan klasifikasi sebesar 72,5352%. Pada algoritma C4.5 mendapatkan pohon keputusan yang terbentuk menghasilkan pohon sebanyak 13 simpul, yang terdiri dari sebuah simpul akar yaitu atribut tekanan darah sistolik, 5 simpul keputusan, serta 7 simpul daun dengan faktor-faktor yang mempengaruhi terjadinya jenis penyakit hipertensi adalah tekanan darah sistolik, riwayat berobat, penyakit lain, tekanan darah diastolik, dan jenis kelamin, sedangkan faktor umur tidak berpengaruh dan

No	Penelitian	Metode	Tujuan	Hasil
				mendapatkan ketepatan klasifikasi sebesar 64,0845%.
9	Ma'arif (2016), STMIK Jenderal Achmad Yani Yogyakarta.	Teknik Klasifikasi menggunakan <i>Naïve Bayes Classifier</i> dan <i>Support Vector Mechine</i>	Tujuan dari penelitian ini adalah untuk membandingkan hasil klasifikasi judul artikel menggunakan <i>Naïve Bayes Classifier</i> dan <i>Support Vector Mechine</i> .	Penelitian ini menghasilkan Klasifikasi <i>Naïve Bayes Classifier</i> menggunakan fitur <i>trigram</i> dan <i>term frequency</i> memiliki nilai akurasi yang paling tinggi yaitu 81%. Sementara itu, model yang dibangun dengan algoritma SVM terbaik 73% dengan menggunakan fitur <i>unigram</i> . Dapat dikatakan bahwa klasifikasi yang dihasilkan oleh <i>Naïve Bayes Classifier</i> memiliki performa yang lebih tinggi dibandingkan dengan model yang dihasilkan dari yang dihasilkan dari algoritm <i>Support Vector Mechine</i> .
10	Fatmawati (2016), STMIK Nusa Mandiri Jakarta.	Teknik Klasifikasi menggunakan	Tujuan dari penelitian ini adalah untuk mengetahui diantara model algoritma C4.5 dan <i>Naïve Bayes</i> yang	Hasil penelitian tersebut didapatkan bahwa klasifikasi menggunakan C4.5 memperoleh akurasi sebesar 73,30% dan nilai AUC sebesar

No	Penelitian	Metode	Tujuan	Hasil
		<i>C4.5</i> dan <i>Naïve Bayes</i> .	memiliki akurasi paling tinggi untuk memprediksi penyakit diabetes	0,733, sedangkan nilai akurasi <i>Naïve Bayes</i> sebesar 75,13% dan nilai AUC adalah 0,810.
10	Ramandhani (2017), Universitas Diponegoro	Teknik Klasifikasi dengan Metode <i>Bootstrap Aggregating</i> Regresi Logistik Biner.	Tujuan penelitian ini adalah mengetahui model dan ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati berdasarkan analisis Regresi Logistik Biner dan mengetahui ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati dengan metode <i>Bootstrap Aggregating (Bagging)</i> Regresi Logistik Biner.	Hasil penelitian ini memperoleh ketepatan klasifikasi sebesar 79,55% untuk Regresi Logistik Biner, sedangkan hasil analisis <i>Bagging</i> Regresi Logistik biner menunjukkan bahwa pada 90 kali replikasi <i>bootstrap</i> diperoleh nilai ketepatan klasifikasi terbesar yaitu 79,87%.

11	Fajri (2018), Sekolah Tinggi Teknologi Nurul Jadid.	Teknik Klasifikasi menggunakan <i>Naïve Bayes</i> dan <i>Decision Tree</i>	Tujuan dari penelitian ini adalah untuk melakukan proses <i>comparasi</i> atau membandingkan antara metode <i>Decision Tress</i> dan <i>Naïve Bayes</i> dalam klasifikasi diagnosa penyakit diabetes.	Hasil dari penelitian ini memperoleh model klasifikasi dengan menggunakan <i>Desicion Tree</i> mempunyai nilai optimal akurasi sebesar 71.75 % dengan menggunakan split rasio 30 % sedangkan dengan menggunakan <i>Naïve Bayes</i> mempunyai nilai optimal akurasi sebesar 75.28 % dengan split rasio 30 %. Dapat dikatakan bahwa <i>Naïve Bayes</i> lebih unggul dibandingkan <i>Decision Tree</i> dalam mengklasifikasikan penyakit diabetes.
----	---	--	---	---

BAB III

LANDASAN TEORI

3.1 Universitas Islam Indonesia

Universitas Islam Indonesia (UII) adalah salah satu perguruan tinggi swasta terkemuka di Indonesia. Terinspirasi oleh semangat nasionalisme dan berpedoman pada nilai-nilai *perennial*, UII didirikan satu bulan sebelum proklamasi kemerdekaan Indonesia pada tahun 1945. Saat ini, UII telah tumbuh berkembang menjadi tempat pembelajaran yang terpilih. Pembelajaran berarti menggali ‘harta karun’ yang ada di dalam diri sendiri. Dengan 3 program Doktor, 9 program Magister, 4 program Profesi, 25 program studi Sarjana, dan 4 program Diploma yang mencakup spektrum ilmu pengetahuan yang luas, UII memungkinkan mahasiswanya untuk menemukan dan meningkatkan potensi dirinya untuk meraih masa depan yang cemerlang. Kesempurnaan adalah hasil dari inovasi seumur hidup dan usaha yang tanpa henti. Dari sisi internal, usaha UII dalam menyempurnakan pelayanan adalah dengan menyediakan fasilitas modern dan meningkatkan jaminan kualitas. Untuk mendukung proses pembelajaran, 10 perpustakaan, 82 laboratorium, dan 27 pusat studi disediakan untuk penelitian dan pendidikan. Selain itu, Masjid Besar, toko buku, gelanggang olahraga, auditorium, rumah sakit dan pusat konvensi mahasiswa juga merupakan fasilitas yang tersedia di beberapa kampus UII. (Akademik UII, 2018)

3.2 Logo Universitas Islam Indonesia

Panduan Penggunaan Merek (*Branding Guidelines*) Universitas Islam Indonesia (UII) adalah dokumen yang berdasar pada SK Rektor No. 1279/SK-Rek/DOSDM/X/2014 yang berisi panduan dan aturan mengenai penggunaan lambang, signature, dan identitas visual Universitas Islam Indonesia. Panduan ini menjadi pedoman utama dalam menempatkan identitas universitas dan diperbarui secara berkala. Berikut ini logo Universitas Islam Indonesia:



Gambar 3. 1 Logo Universitas Islam Indonesia

Elemen yang terdapat di dalam logo UII berupa simbol dan warna dijelaskan sebagai berikut:

1. Gambar: Gambar lambang universitas berwujud perisai yang di dalamnya terdapat setangkai bunga berpolakan pena, buku, lunas kapal dan bertuliskan *Al-Jami'ah Al-Islamiyah Al-Indonesiyah* berwarna kuning emas jika warna dasarnya gelap dan berwarna biru jika warna dasarnya terang.
2. Bentuk: Bentuk perisai melambangkan pertahanan dan ketahanan. Bentuk bunga di tengah yang digayakan (distilisasi) menjadi kubah masjid melambangkan kebudayaan Indonesia yang sesuai dengan ajaran Islam. Bentuk bunga dengan lima mahkota melambangkan Dasar Negara Pancasila dan Rukun Islam. Bentuk melati berkelopak empat dan pada putik sari di atas bunga melambangkan perguruan tinggi atas dasar Catur Dharma. Bentuk yang digayakan (distilisasi) menjadi buku pada kelopak daun tengah melambangkan Kitab Suci Al Qur'an. Bentuk dua penyangga pada kelopak daun yang paling bawah melambangkan Dua Kalimat Syahadat. Bentuk lunas kapal sebagai pintu masjid di bawah kelopak melambangkan unsur budaya Islam.
3. Warna: Warna biru melambangkan ketegasan dan/atau kewibawaan. Warna kuning emas melambangkan harapan dan/atau pendidikan. Warna putih melambangkan ketulusan, kejujuran, dan ketekunan.

3.3 Visi dan Misi Universitas Islam Indonesia

Adapun visi dan misi Universitas Islam Indonesia sebagai berikut:

a. Visi

Terwujudnya Universitas Islam Indonesia sebagai *rahmatan lil 'alamin*, memiliki komitmen pada kesempurnaan (keunggulan), risalah Islamiyah, di bidang pendidikan, penelitian, pengabdian masyarakat dan dakwah, setingkat universitas yang berkualitas di negara-negara maju

b. Misi

Menegakkan *Wahyu Illahi* dan *Sunnah Nabi* sebagai sumber kebenaran abadi yang membawa rahmat bagi alam semesta melalui pengembangan dan penyebaran ilmu pengetahuan, teknologi, budaya, sastra dan seni yang berjiwa Islam, dalam rangka membentuk cendekiawan muslim dan pemimpin bangsa yang bertakwa, berakhlak mulia, berilmu amaliah dan beramal ilmiah, yang memiliki keunggulan dalam keislaman, keilmuan, kepemimpinan, keahlian, kemandirian dan profesionalisme.

3.4 Mahasiswa

Mahasiswa adalah seseorang yang sedang dalam proses menimba ilmu ataupun belajar dan terdaftar sedang menjalani pendidikan pada salah satu bentuk perguruan tinggi yang terdiri dari akademik, politeknik, sekolah tinggi, institut dan universitas (Hartaji, 2012). Menurut Undang – Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan pasal 1 ayat 15 , mahasiswa adalah Peserta didik pada jenjang Pendidikan Tinggi. Menurut Siswoyo (2007) mahasiswa dapat didefinisikan sebagai individu yang sedang menuntut ilmu ditingkat perguruan tinggi, baik negeri maupun swasta atau lembaga lain yang setingkat dengan perguruan tinggi. Mahasiswa dinilai memiliki tingkat intelektualitas yang tinggi, kecerdasan dalam berpikir dan perencanaan dalam bertindak. Berpikir kritis dan bertindak dengan cepat dan tepat merupakan sifat yang cenderung melekat pada diri setiap mahasiswa, yang merupakan prinsip yang saling melengkapi.

3.5 Kelulusan Mahasiswa

Berdasarkan ketetapan Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi tentang Sistem Pendidikan Tinggi disebutkan bahwa untuk memenuhi standar kompetensi lulusan bagi mahasiswa program sarjana (S1) beban wajib yang harus ditempuh adalah paling sedikit 144-160 satuan kredit semester (sks) dengan masa studi selama 8-12 semester atau 4-6 tahun.

Seorang mahasiswa dinyatakan lulus program apabila telah menyelesaikan minimal SKS sesuai dengan kurikulum masing-masing Program Studi dengan Indeks Prestasi Kumulatif (IPK) minimal 2.00 dan menyelesaikan Tugas Akhir dan/atau skripsi dan telah mempublikasi-kan karya ilmiah untuk program S1 atau telah menyelesaikan Tugas Akhir atau Karya Tulis. Setelah melakukan Evaluasi Akhir Studi, Ketua Program Studi melaporkan kepada Dekan mahasiswa yang telah memenuhi persyaratan untuk lulus untuk diajukan pengesahan kelulusannya kepada Rektor. Rektor menerbitkan ijazah bagi mahasiswa yang telah dinyatakan lulus. Seorang mahasiswa dinyatakan lulus berhak memakai gelar sarjana untuk Program Strata Satu. Indeks Prestasi Kelulusan sebagai dasar penentuan predikat kelulusan ditentukan sebagai berikut :

1. IPK 2.00-2.75 : Lulus dengan predikat cukup
2. IPK 2.76-3.00 : Lulus dengan predikat memuaskan
3. IPK 3.01-3.05 : Lulus dengan predikat sangat memuaskan
4. IPK 3.51-4.00 : Lulus dengan predikat cumlaude.

3.6 Statistika Deskriptif

Statistika deskriptif adalah bagian dari statistika yang mempelajari alat, teknik, atau prosedur yang digunakan untuk menggambarkan atau mendeskripsikan kumpulan data atau hasil pengamatan yang telah dilakukan. Kegiatan-kegiatan tersebut antara lain adalah kegiatan pengumpulan data, pengelompokan data, penentuan nilai dan fungsi statistik, serta pembuatan grafik, diagram dan gambar. Statistika deskriptif ini merupakan metode-metode yang berkaitan dengan pengumpulan, peringkasan, dan

penyajian suatu data sehingga memberikan informasi yang berguna dan juga menatanya ke dalam bentuk yang siap untuk dianalisis. Dengan kata lain, statistika deskriptif ini merupakan fase yang membicarakan mengenai penjabaran dan penggambaran termasuk penyajian data. Dalam fase ini dibahas mengenai ukuran-ukuran statistik seperti ukuran pusat, ukuran sebaran, dan ukuran lokasi dari persebaran atau distribusi data.

Metode statistik adalah prosedur-prosedur yang digunakan dalam pengumpulan, penyajian, analisis dan penafsiran data. Kemudian metode tersebut dibagi menjadi dua, yaitu statistika deskriptif dan statistik inferensial (Walpole dkk, 1995).

Statistik deskriptif adalah statistik yang berfungsi untuk mendiskripsikan atau memberi gambaran terhadap obyek yang diteliti melalui data sampel atau populasi sebagaimana adanya, tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum (Sugiyono, 2007).

Statistik deskriptif adalah bagian dari statistika yang mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Statistika deskriptif hanya berhubungan dengan hal menguraikan atau memberikan keterangan-keterangan mengenai suatu data atau keadaan (Hasan, 2004). Dengan kata statistika deskriptif berfungsi menerangkan keadaan, gejala, atau persoalan. Penarikan kesimpulan pada statistika deskriptif (jika ada) hanya ditujukan pada kumpulan data yang ada.

3.7 Regresi Logistik Biner

Regresi logistik biner adalah suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon yang bersifat biner dengan variabel prediktor (Hosmer dan Lemeshow, 2000). Variabel respon y terdiri dari 2 kategori yaitu sukses dan gagal yang dinotasikan dengan $y=1$ (sukses) dan $y=0$ (gagal). Dalam keadaan demikian, variabel y mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Fungsi Probabilitas untuk setiap observasi adalah diberikan sebagai berikut.

$$f(y) = \pi^y(1-\pi)^{1-y}; \quad y=0, 1 \quad (3.1)$$

Dimana jika $y = 0$ maka $f(y) = 1 - \pi$ dan jika $y = 1$ maka $f(y) = \pi$. Fungsi regresi logistiknya dapat dituliskan sebagai berikut.

$$f(y) = \frac{e^y}{1+e^y} \quad (3.2)$$

dengan $y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

Nilai y antara $-\infty$ dan $+\infty$ sehingga nilai $f(y)$ terletak antara 0 dan 1 untuk setiap nilai y yang diberikan. Hal tersebut menunjukkan bahwa model logistik sebenarnya menggambarkan probabilitas atau risiko dari suatu objek. Menurut Hosmer dan Lemeshow (2000), model regresi logistik dengan p buah variabel predictor dibentuk dengan nilai $\pi(x) = E(Y = 1 | x)$. Model regresi logistiknya adalah sebagai berikut.

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}} \quad (3.3)$$

Dimana :

$\pi(x)$: probabilitas “sukses” atau peluang kejadian sukses

p : banyaknya variabel predictor

X_1, \dots : Variabel independen kuantitatif atau kualitatif

β_0 : Konstanta dari model

$\beta_1 \dots \beta_p$: Parameter koefisien regresi

Model transformasi logit dari $\pi(x)$ dari persamaan diatas dapat dituliskan sebagai berikut:

$$g(X) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (3.4)$$

dengan $g(X)$: transformasi logit dari $\pi(x)$

Dalam penelitian Imaslihhah (2013), tentang Faktor-faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya menggunakan Regresi Logistik, bahwa mengasumsikan variabel prediktor diantaranya, Jenis Kelamin, Fakultas, Asal Daerah, Jalur Penerimaan, Status SMA, Jenis SMA, Pekerjaan Ibu dan Ayah, serta Pendapatan. Pengujian secara serentak didapatkan bahwa, variabel

fakultas, jenis kelamin, jalur penerimaan, pekerjaan ayah, pekerjaan ibu dan pendapatan. Pengujian secara individu, faktor yang berpengaruh adalah fakultas, jalur penerimaan, pekerjaan ayah, pekerjaan ibu dan pendapatan. Dimana, ketepatan klasifikasi dari model yang didapat sebesar 77,41% yang berarti sudah cukup baik.

3.7.1 Pengujian Parameter dan *Maximum Likelihood* untuk Regresi Logistik

Estimasi parameter dalam regresi logistik dilakukan dengan metode *Maximum Likelihood*. Metode tersebut mengestimasi parameter β dengan cara memaksimalkan fungsi likelihood dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Pada regresi logistik, setiap pengamatan mengikuti distribusi bernoulli sehingga dapat ditentukan fungsi *likelihood*nya.

Misalkan suatu sampel terdiri dari n percobaan yang saling bebas, dengan y_i adalah variabel respon dari observasi ke- i ($i = 1, 2, \dots, n$) berdistribusi binomial dengan probabilitas sukses $\pi(x_i)$ dan probabilitas gagal $1 - \pi(x_i)$, serta x_i merupakan variabel terikat pada pengamatan ke- i ($i=1,2,\dots,n$) maka fungsi probabilitas untuk setiap pasangan adalah sebagai berikut.

$$f(x_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad (3.5)$$

dimana $(y_i) = 0,1$

dengan,

$$\pi(x_i) = \frac{e^{\left(\sum_{j=0}^p \beta_j x_j\right)}}{1 + e^{\left(\sum_{j=0}^p \beta_j x_j\right)}} \quad (3.6)$$

Dimana, ketika $j=0$, maka nilai $x_{ij} = x_{i0} = 1$

Setiap pasangan pengamatan diasumsikan independen, sehingga fungsi likelihoodnya merupakan gabungan dari fungsi distribusi masing-masing pasangan yaitu sebagai berikut:

$$l(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad (3.7)$$

Fungsi likelihood tersebut lebih mudah dimaksimalkan dalam bentuk $\log l(\beta)$ dan dinyatakan dengan $L(\beta)$.

$$\begin{aligned}
L(\beta) &= \ln l(\beta) \\
&= \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \\
&= \sum_{i=1}^n (\sum_{j=1}^p y_i x_{ij}) \beta_j - \sum_{i=1}^n \ln \left(1 + e^{(\sum_{j=0}^p \beta_j x_j)} \right) \quad (3.8)
\end{aligned}$$

Nilai β maksimum didapatkan melalui turunan $L(\beta)$ terhadap β dan hasilnya adalah sama dengan nol.

$$\begin{aligned}
\frac{\partial L(\beta)}{\partial \beta_j} &= \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left(\frac{e^{(\sum_{j=0}^p \beta_j x_j)}}{1 + e^{(\sum_{j=0}^p \beta_j x_j)}} \right) \\
\sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \hat{\pi}(x_i) &= 0 \quad (3.9)
\end{aligned}$$

dengan $j = 0, 1, \dots, p$

Estimasi varians dan kovarians dikembangkan melalui teori MLE (*Maximum Likelihood Estimation*) dari koefisien parameternya (Rao, 1973 dalam Hosmer dan Lemeshow, 1989). Teori tersebut menyatakan bahwa estimasi varians kovarians didapatkan melalui turunan kedua $L(\beta)$.

$$\frac{\partial^2 L(\beta)}{\partial L \beta_j \beta_u} = - \sum_{i=1}^n x_{ij} x_{iu} \pi(x_i) (1 - \pi(x_i)) \quad (3.9)$$

dengan $j, u = 0, 1, \dots, p$

Matriks varians kovarians berdasarkan estimasi parameter diperoleh melalui invers matriks dan diberikan sebagai berikut:

$$\hat{Cov}(\hat{\beta}) = \{x^T \text{Diag} [\hat{\pi}(x_i)(1 - \hat{\pi}(x_i))] x\}^{-1} \text{ dan } x^T \text{ diberikan oleh,}$$

$$x^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{nk} \\ \dots & \dots & & \dots \\ x_{1k} & x_{21} & \dots & x_{nk} \end{bmatrix}$$

$Diag [\hat{\pi}(x_i)(1 - \hat{\pi}(x_i))]$ merupakan matriks diagonal ($n \times n$) dengan diagonal utamanya adalah $[\hat{\pi}(x_i)(1 - \hat{\pi}(x_i))]$. Penaksir $SE(\hat{\beta})$ diberikan oleh akar kuadrat diagonal utama. Untuk mendapatkan nilai taksiran β dari turunan pertama fungsi $L(\beta)$ yang non linier maka digunakan metode iterasi Newton Raphson. Persamaan yang digunakan adalah

$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1}q^{(t)} \quad (3.10)$$

dimana $t = 1, 2, \dots$ sampai konvergen

$$\text{dengan, } q^T = \left(\frac{\partial L(\beta)}{\partial \beta_0}, \frac{\partial L(\beta)}{\partial \beta_1}, \dots, \frac{\partial L(\beta)}{\partial \beta_k} \right)$$

dan H merupakan matriks *Hessian*. Elemen-elemennya adalah $h_{ju} = \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_u}$, $j = 1, 2, \dots, n$, dan $u = 1, 2, \dots, n$; $(H^t)^{-1} = \frac{1}{|h|} adj(H)$ sehingga,

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1k} \\ h_{21} & h_{22} & \dots & h_{2k} \\ \dots & \dots & \ddots & \dots \\ h_{k1} & h_{k2} & \dots & h_{kk} \end{bmatrix} \quad (3.11)$$

dimana k merupakan banyaknya iterasi dan ju merupakan variabel prediktor. Matriks *Hessian* merupakan matriks simetris (*upper diagonal=lower diagonal*) atau dapat dikatakan matriks yang transposenya sama dengan dirinya sendiri. Dari persamaan tersebut, diperoleh :

$$\beta^{(t+1)} = \beta^{(t)} + \{x^T Diag [\pi(x_i)^{(t)}(1 - \pi(x_i)^{(t)})] x\}^{-1} x^T (y - m^{(t)}) \quad (3.12)$$

Dengan $m^{(0)} = \pi(x_i)^{(0)}$. Langkah-langkah iterasi *Newton Raphson* seperti berikut.

- Menentukan nilai dugaan awal $\beta^{(0)}$ kemudian dengan menggunakan persamaan (3.11) maka didapatkan $\pi(x_i)^{(0)}$.
- Dari $\pi(x_i)^{(0)}$ pada langkah a, diperoleh matriks *Hessian* $H^{(0)}$ dan vektor $q^{(0)}$.

- c. Proses selanjutnya untuk $t > 0$ digunakan persamaan (3.11) dan (3.12) hingga $\pi(x_i)^{(t)}$ dan $\beta^{(t)}$ konvergen.

Ada dua cara mengestimasi model regresi logistik, yaitu metode secara menyeluruh dan secara bertahap (*stepwise*). (Widarjono, 2010)

a. Uji Likelihood

Uji rasio Likelihood digunakan untuk menguji kelayakan model yang diperoleh dari estimasi parameter yang bertujuan untuk mengetahui apakah variabel independen yang terdapat dalam model berpengaruh nyata atau tidak secara keseluruhan (Hosmer and Lemeshow, 2000). Uji ini membandingkan model lengkap (model dengan variabel prediktor) terhadap model yang hanya dengan konstanta (model tanpa variabel prediktor) untuk melihat apakah model yang hanya dengan konstanta secara signifikan lebih baik dari model lengkap dengan rumus sebagai berikut.

$$G^2 = -2 \ln \left[\frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right] \quad (3.13)$$

$$G^2 = -2 \ln \left(\frac{\binom{n_1}{n} n_1 \binom{n_0}{n} n_0}{\prod_{i=1}^n \pi_1^{y_i} (1-\pi_1)^{(1-y_i)}} \right) \quad (3.14)$$

dengan:

n_1 : banyaknya observasi berkategori 1

n_0 : banyaknya observasi berkategori 0

n : banyaknya observasi (n_1+n_0)

Hipotesa :

H_0 : $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$ (Secara simultan variabel bebas tidak berpengaruh terhadap variabel terikat)

H_1 : $\beta_j \neq 0$, $j = 1, 2, \dots, p$ (minimal ada satu variabel bebas yang berpengaruh secara simultan terhadap variabel terikat)

Kriteria ini mengambil taraf nyata α maka H_0 ditolak jika $G^2 > X_{(\alpha, v)}^2$ dimana v adalah banyaknya variabel prediktor.

b. Uji Wald

Uji Wald digunakan untuk mengetahui variabel-variabel bebas mana yang mempunyai hubungan lebih kuat dengan variabel respon nya (Hosmer and Lemeshow, 2000). Statistik uji Wald dihitung dengan membagi parameter yang ditaksir oleh galat baku dari parameter yang ditaksir.

$$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2 \quad (3.15)$$

Dimana:

$\hat{\beta}_j$: nilai dugaan untuk parameter β_j

SE : dugaan galat baku untuk koefisien β_j

Hipotesa :

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

H_1 : paling sedikit satu $\beta_j \neq 0$, $j = 1, 2, 3, \dots, p$

Kriteria ini mengambil taraf nyata α maka H_0 ditolak jika $|W_j| > X_{(\alpha,1)}^2$

3.7.2 Uji Kesesuaian Model

Uji kesesuaian model digunakan untuk menilai apakah model sesuai atau tidak. Pada uji ini, statistik uji yang digunakan adalah statistik *devians* (D) yang dirumuskan sebagai berikut: (Nugraha, 2013)

$$D = -2 \sum_1^n = 1 \left\{ y_i \ln \left(\frac{n_i \hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(n_i - \frac{n_i \hat{\pi}_i}{ni - y_i} \right) \right\} \quad (3.16)$$

Dimana :

$\hat{\pi}_i = \hat{\pi}(x_i)$: Peluang observasi ke-i pada kategori ke-i

y_i : Banyaknya sukses

n_i : Banyaknya pengamatan

Hipotesa :

H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Statistik D berdistribusi *Chi-Square* dengan derajat bebas uji ini adalah $(k - (p + 1))$ dimana k merupakan jumlah kovariat dan p merupakan jumlah variabel independen

3.7.3 Odds Ratio

Odds adalah cara penyajian probabilitas, yang menjelaskan probabilitas bahwa kejadian tersebut akan terjadi dibagi dengan probabilitas bahwa kejadian itu tidak akan terjadi. *Odds* adalah rasio probabilitas sukses (π) terhadap probabilitas gagal ($1 - \pi$). Pada data populasi, nilai *odds*-nya adalah :

$$odds = \frac{\pi}{1-\pi} \quad (3.17)$$

Sedangkan untuk sampel dihitung dengan menggunakan rumus :

$$odds = \frac{p}{1-p} \quad (3.18)$$

Odds bernilai positif,

$$0 < odds < \infty$$

Ketika *odds* bernilai satu, berarti probabilitas sukses sama dengan probabilitas gagal, kemudian jika nilai *odds* bernilai kurang dari satu berarti probabilitas sukses lebih kecil daripada probabilitas gagal. Demikian juga sebaliknya jika nilai *odds* lebih dari satu berarti probabilitas sukses lebih besar dari pada probabilitas gagal.

3.8 Variabel Dummy

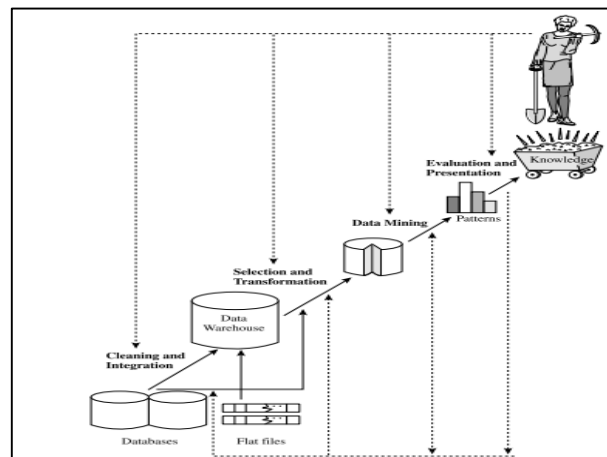
Dummy variabel adalah variabel data yang bersifat kualitatif yang digunakan untuk membuat kategori (Nachowi da Usman, 2002). Menurut Supranto (2004) variable *dummy* disebut juga variabel indicator biner, kategorik, kualitatif, boneka atau variabel dikotomi. Suatu persamaan regresi tidak hanya menggunakan variabel kategorik sebagai variabel bebas, tetapi dapat pula disertai oleh variabel bebas lain

yang numerik. Variabel *dummy* bisa digunakan pada variabel tak bebas Y , sehingga Y bernilai 0 atau 1, yang memiliki arti ya atau tidak (bersifat dikotomi)

3.9 Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (R Shyara dan Saroj, 2013). Selain itu, kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar disebut dengan *Data Mining* (Mujiasih, 2011). Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa *Data Mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan.

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada **Gambar 3.1**. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.



(Sumber: Han dan Kamber, 2006)

Gambar 3.2 Langkah-langkah untuk melakukan Data Mining

1) Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.

2) Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.

3) Seleksi data (*data selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

4) Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.

5) Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data

6) Evaluasi Pola (*Pattern evolution*)

Untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik.

7) *Knowledge presentation* (Pengetahuan Presentasi)

Di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah ditambah pada pengguna.

3.10 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu:

1. Proses *Training*

Pada proses training digunakan training set yang telah diketahui label-labelnya untuk membangun model atau fungsi.

2. Proses *Testing*

Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses *training*, maka digunakan data yang disebut dengan testing set untuk memprediksi label-labelnya

3.11 *Naïve Bayes*

Algoritma Klasifikasi Naïve Bayes menggunakan data dengan target (*class/label*) yang berupa nilai kategorikal/nominal. *Naïve Bayes Classifier* (NBC) adalah salah satu algoritma dalam teknik data mining yang menerapkan teori Bayes dalam klasifikasi. *Naïve Bayes Classifier* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Teori keputusan Bayes merupakan pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), teori Bayesian pada dasarnya adalah kemungkinan kejadian di masa depan yang bisa dihitung dengan menentukan frekuensi pengalaman sebelumnya. Penggunaan algoritma *Bayes* dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Metode *Naïve Bayes* sering disebut dengan algoritma HMAP (*Hypothesis Maximum Apriori Probability*) yang merupakan penyederhanaan dari metode bayes. Metode ini menyatakan hipotesa dari penghitungan menggunakan probabilitas berdasarkan kondisi prior (Prasetyo, 2012).

Dalam penelitian Jananto (2013) tentang perkiraan waktu lama studi mahasiswa, algoritma *Naïve Bayes* mampu mengklasifikasikan lama studi mahasiswa dengan tingkat kesalahan berkisar pada 20% hingga 34% yang hal ini dapat dipengaruhi oleh jumlah data *training* maupun *testing* yang digunakan serta tingkat konsisten data. Faktor-faktor yang diasumsikan mempengaruhi lama studi mahasiswa antara lain Indeks Prestasi Kumulatif, Jenis Kelamin, Kota Lahir, Tipe Sekolah serta Kota Sekolah. Oleh karena itu fungsi prediksi dengan memanfaatkan teknik *data mining* menggunakan algoritma *Naïve Bayes* telah dapat dibuat dan digunakan untuk memprediksi dari masa studi atau ketepatan masa studi dari mahasiswa dengan *data training* dan data *testing* yang telah diperoleh. Penelitian oleh Fatmawati (2013),

menyatakan bahwa dalam pengklasifikasian penyakit diabetes menggunakan *Naïve Bayes* memiliki akurasi yang lebih tinggi dibandingkan algoritma C4.5, dimana *Naïve Bayes* memiliki akurasi sebesar 75,13%, sedangkan algoritma C4.5 sebesar 73,30% serta pengujian dari prediksi diabetes hasilnya termasuk *Good Classification*. Kelebihan dari *Naïve Bayes Classifier* adalah sederhana tetapi memiliki akurasi yang tinggi (Rish, 2006). Selain itu *Naïve Bayes* juga dapat menangani kuantitatif dan data diskrit, serta hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi.

Persamaan dari teorema *Bayes* adalah :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (3.19)$$

Dimana :

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis data merupakan suatu *class* spesifik
- $P(H|X)$: Probabilitas hipotesis *H* berdasarkan kondisi X (posteriori probabilitas)
- $P(H)$: Probabilitas hipotesis *H* (prior probabilitas)
- $P(X|H)$: Probabilitas *X* berdasarkan kondisi pada hipotesis *H*
- $P(X)$: Probabilitas *X*

Untuk menjelaskan teorema *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (3.20)$$

Dimana variabel *C* merepresentasikan kelas, sementara variabel F_1, \dots, F_n merepresentasikan karakteristik-karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel dengan 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*)

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (3.21)$$

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $P(F_1, \dots, F_n | C)$ menggunakan aturan perkalian, menjadi sebagai berikut:

$$\begin{aligned} P(F_1, \dots, F_n | C) &= P(F_1 | C) P(F_2, \dots, F_n | C, F_1) \\ &= P(F_1 | C) P(F_2 | C, F_1) P(F_3, \dots, F_n | C, F_1, F_2) \\ &= P(F_1 | C) P(F_2 | C, F_1), \dots, P(F_n | C, F_1, F_2, \dots, F_{n-1}) \end{aligned} \quad (3.22)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, Di sinilah digunakan asumsi independensi yang sangat tinggi (*naïf*), 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:

$$P(F_i | F_j) = \frac{P(F_i \cap F_j)}{P(F_i)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (3.23)$$

untuk $i \neq j$, sehingga

$$P(F_i | C, F_j) = P(F_i | C) \quad (3.24)$$

Dari persamaan (3.24) 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:

$$\begin{aligned} P(F_1, \dots, F_n | C) &= P(F_1 | C) P(F_2 | C), \dots, P(F_n | C) \\ P(F_1, \dots, F_n | C) &= \prod_{i=1}^n P(F_i | C) \end{aligned} \quad (3.25)$$

3.12 Confusion Matrix

Confusion matrix adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Contoh *confusion matrix* untuk klasifikasi biner ditunjukkan pada **tabel 4.1**.

Tabel 4. 1 Confussion Matrix

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan untuk tabel 4.1 dinyatakan sebagai berikut:

- *True Positive* (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
- *True Negative* (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
- *False Positive* (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
- *False Negative* (FN) yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Perhitungan akurasi dinyatakan dalam persamaan (3.20).

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (3.20)$$

Kemudian dapat juga menghitung APER atau yang disebut laju error merupakan ukuran evaluasi yang digunakan untuk melihat peluang kesalahan klasifikasi yang dihasilkan oleh suatu fungsi klasifikasi. Semakin kecil nilai APER maka hasil pengklasifikasian semakin baik (Prasetyo, 2012). Formulasi untuk menghitung APER yaitu :

$$APPER = \frac{FP+FN}{TP+FP+TN+FN} \times 100\% \quad (3.21)$$

BAB IV

METODELOGI PENELITIAN

4.1 Populasi dan Sampel

Pada penelitian ini populasi yang dipilih erat kaitannya dengan masalah yang akan diteliti. (Arikunto, 2002) menyatakan bahwa populasi adalah keseluruhan objek penelitian. Populasi dari penelitian ini adalah seluruh alumni Universitas Islam Indonesia. Kemudian sampel yang digunakan dalam penelitian ini adalah data alumni (kelulusan mahasiswa) Universitas Islam Indonesia Tahun Akademik 2016/2017 sebesar 3.673 mahasiswa.

4.2 Metode Pengumpulan Data dan Sumber Data

Metode pengumpulan data pada penelitian ini merupakan metode dengan data sekunder yang bersumber dari *database* Akademik Rektorat Universitas Islam Indonesia di akses pada Januari 2018.

4.3 Variabel Penelitian

Pada penelitian ini, himpunan data yang digunakan adalah Data 8 fakultas alumni mahasiswa Universitas Islam Indonesia tahun akademik 2016/2017. Dimana variabel yang digunakan antara lain meliputi variabel dependen dan variable independen. Berikut daftar variabel penelitian dapat dilihat pada **tabel 4.1**.

Tabel 4.1 Variabel Penelitian

No	Jenis Variabel	Variabel	Pengertian	Kategori
1	Dependen	Lama Studi Mahasiswa	Lama studi mahasiswa merupakan status kelulusan mahasiswa dalam beberapa tahun.	a. ≤ 4 Tahun : Tepat waktu b. > 4 tahun: Tidak Tepat Waktu

2	Independen	Jenis Kelamin	Karakteristik mahasiswa yang dapat dibedakan	<ul style="list-style-type: none"> a. Laki-laki b. Perempuan
3	Independen	Asal Daerah	Tempat tinggal mahasiswa yang tercatat di dalam kependudukan.	<ul style="list-style-type: none"> a. Yogyakarta : Mahasiswa yang berasal dari daerah Yogyakarta b. Luar Yogyakarta : Mahasiswa yang bukan berasal dari Yogyakarta
4	Independen	Jenis SMA	Jenis sekolah mahasiswa ketika masih duduk di bangku SMA yang merupakan tempat belajar.	<ul style="list-style-type: none"> a. Negeri b. Swasta Lainnya (Selain Negeri dan Swasta, misalnya SMA di Luar Negeri)
5	Independen	Jurusan saat SMA	Jurusan ketika masih duduk di bangku SMA.	<ul style="list-style-type: none"> a. IPA b. IPS Lainnya (Selain IPA dan IPS)
6	Independen	Jenis Fakultas	Jenis fakultas yang ada di Universitas Islam Indonesia. Fakultas merupakan	<ul style="list-style-type: none"> a. FIAI b. FMIPA c. FPSB d. FK

			divisi dalam universitas yang terdiri dari sejumlah bidang studi atau jurusan terkait.	<ul style="list-style-type: none"> e. FTI f. FTSP g. FE h. FH
7	Independen	Indeks Prestasi Kumulatif	Alat ukur prestasi di bidang akademik atau pendidikan yang dihitung berdasarkan sks yang ditempuh.	<ul style="list-style-type: none"> a. Cumlaude yaitu nilai antara 3.51-4.00 b. Sangat memuaskan yaitu nilai diantar 3.00-3.50 c. Memuaskan yaitu nilai diantar 2.75-2.99 d. Cukup yaitu nilai dibawah 2.75
8	Independen	Pekerjaan Orang Tua	Usaha yang dilakukan orang tua untuk memenuhi kebutuhan sehari-hari keluarganya.	<ul style="list-style-type: none"> a. Pegawai Negeri Sipil b. Pegawai Swasta c. Wiraswasta d. Petani e. Pensiun Tidak Bekerja

4.4 Metode Analisis Data

Penulis menggunakan beberapa metode dalam analisis data tersebut, diantaranya :

1. Analisis Statistika Deskriptif yaitu digunakan untuk mendapatkan gambaran umum dari lama studi mahasiswa berdasarkan karakteristik mahasiswa seperti Jenis Kelamin, Asal Daerah , IPK, Jenis SMA, Jurusan SMA, Jenis Fakultas serta Jenis Pekerjaan Orang Tua
2. Analisis Regresi Logistik yang digunakan untuk mengetahui data variable independen mana saja yang mempengaruhi variable dependen (lama studi mahasiswa) secara signifikan, serta klasifikasi dengan menggunakan Regresi Logistik, digunakan untuk mengetahui mengetahui tingkat akurasi dari klasifikasi tersebut.
3. Klasifikasi dengan menggunakan *Naive Bayes*, digunakan untuk mengetahui klasifikasi lama studi mahasiswa serta untuk mengetahui tingkat akurasi dari klasifikasi tersebut.

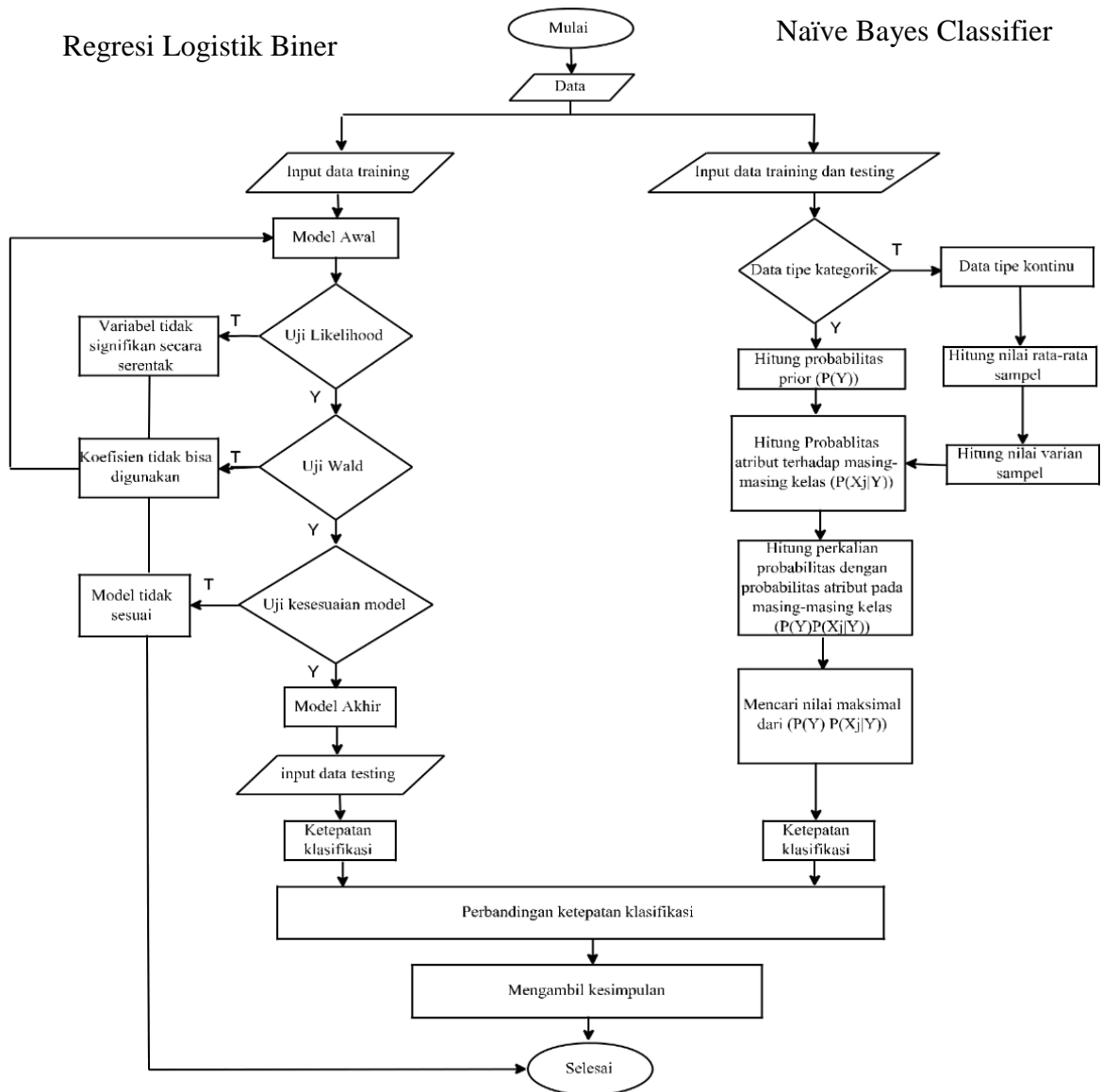
3.5 Tahap Analisis Data

Berikut ini adalah tahap-tahap analisis data yang digunakan :

- a. Tahapan analisis data menggunakan analisis regresi logistik biner :
 1. Input data training
 2. Uji Likelihood
 3. Uji Wald
 4. Uji kesesuaian model
 5. Model akhir
 6. Input data testing
 7. Ketepatan klasifikasi
- b. Tahapan analisis data menggunakan *Naive Bayes Classifier*:
 1. Input data *training* dan data *testing*
 2. Data kategorik, lakukan perhitungan probabilitas prior $P(Y)$.

3. Kemudian, data kategorik, hitung probabilitas atribut terhadap masing-masing kelas $P(X_j|Y)$. Hitung perkalian probabilitas dengan probabilitas atribut pada masing-masing kelas
4. Mencari nilai maksimal dari $P(Y)P(X_j|Y)$
5. Ketepatan klasifikasi

Dari tahapan diatas, berikut ini gambaran *flow chart* untuk melihat tahap-tahap analisis data dapat dilihat pada **gambar 4.1**.



BAB V

PEMBAHASAN

Dalam bab ini, akan dijelaskan hasil dari penelitian yang telah dilakukan, yakni analisis Deskriptif yang digunakan untuk mendeskripsikan atau menggambarkan suatu data, kemudian analisis Regresi Logistik Biner yang digunakan untuk mengetahui faktor-faktor yang mempengaruhi secara signifikan dan klasifikasi lama studi mahasiswa serta analisis Naïve Bayes yang digunakan untuk pengklasifikasian lama studi mahasiswa juga.

5.1 Analisis Deskriptif

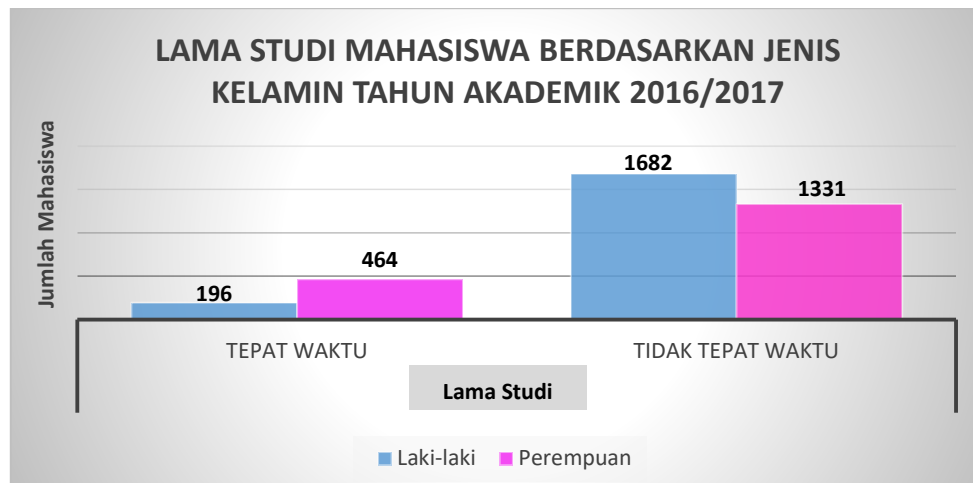
Analisis Deskriptif yang digunakan untuk melihat gambaran umum pada data yang digunakan. Data yang digunakan adalah data Alumni Mahasiswa Universitas Islam Indonesia Tahun Akademik 2016/2017. Berikut ini merupakan persentase lama studi mahasiswa Tahun Akademik 2016/2017 dapat dilihat pada gambar 5.1 dibawah ini.



Gambar 5. 1 Persentase Ketepatan Waktu Kelulusan Mahasiswa Tahun Akademik 2016/2017

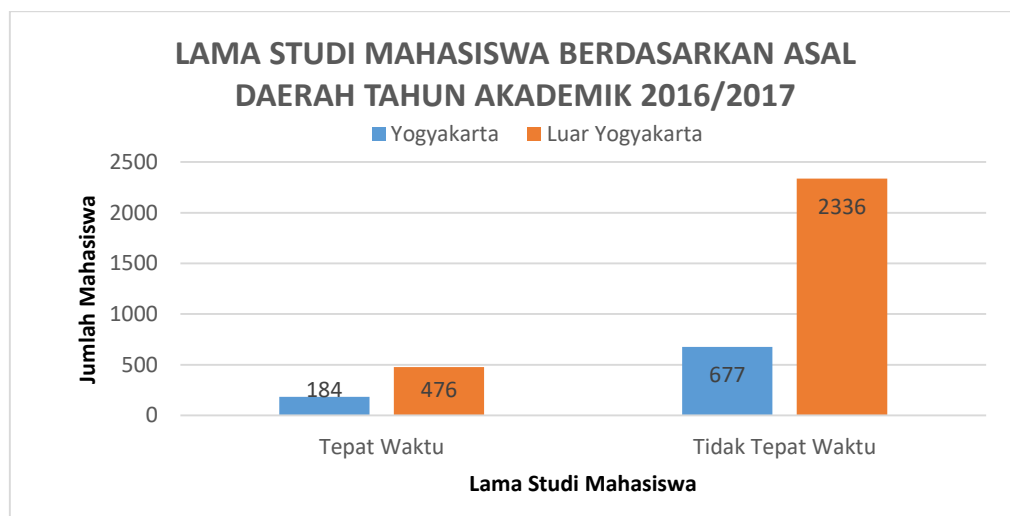
Berdasarkan **gambar 5.1** diatas, dapat diketahui bahwa persentase lama studi mahasiswa Tepat Waktu hanya sebesar 18% sedangkan untuk mahasiswa tidak tepat

waktu sebesar 82%. Dari gambaran tersebut dapat dikatakan bahwa masih banyak mahasiswa yang lulus tidak tepat waktu, atau masa studinya kurang dari 4 tahun.



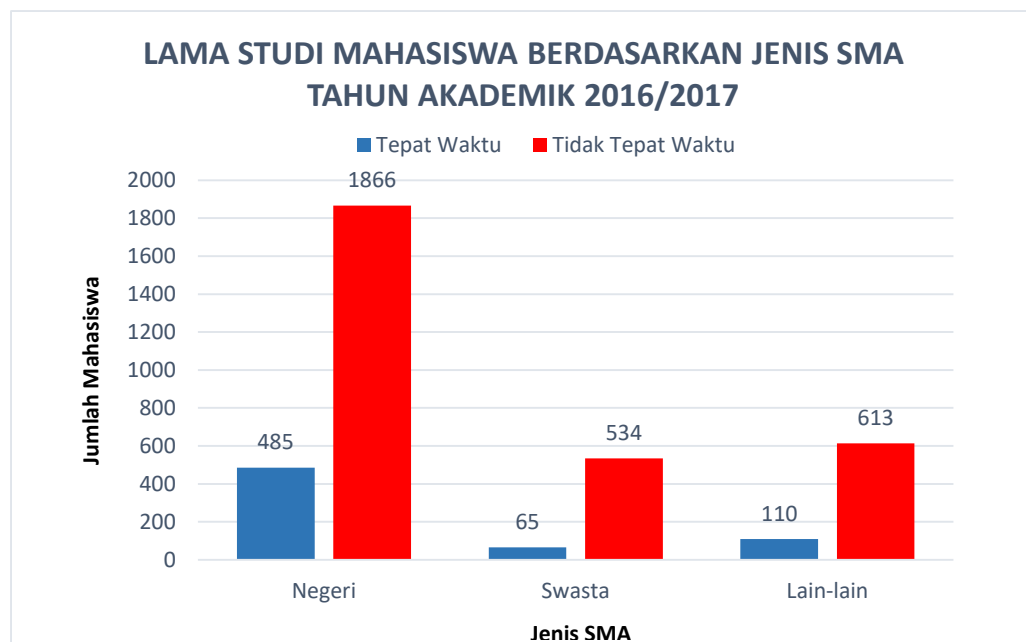
Gambar 5. 2 Lama Studi Mahasiswa Berdasarkan Jenis Kelamin

Berdasarkan **gambar 5.2** dapat dilihat bahwa mahasiswa Universitas Islam Indonesia yang paling banyak adalah jumlah mahasiswa Laki-laki. Tetapi untuk kelulusan mahasiswa secara tepat waktu atau mahasiswa menyelesaikan masa studinya ≤ 4 tahun adalah mahasiswa perempuan yakni sebesar 464 mahasiswa. Hal ini dapat disimpulkan bahwa mahasiswa berjenis kelamin perempuan lebih cenderung menyelesaikan masa studi tepat waktu dibandingkan mahasiswa laki-laki.



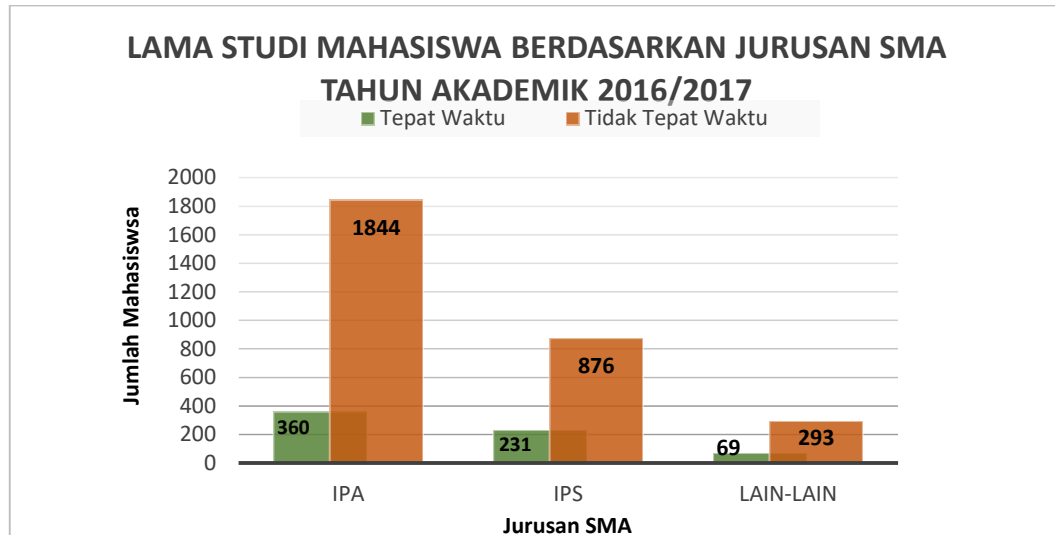
Gambar 5. 3 Lama Studi Mahasiswa Berdasarkan Asal Daerah

Berdasarkan **gambar 5.3** diatas dapat dijelaskan bahwa jumlah kelulusan tidak tepat waktu paling banyak adalah daerah luar Yogyakarta sebesar 2336 mahasiswa. Kemudian untuk jumlah kelulusan tepat waktu paling banyak juga terdapat pada asal daerah luar Yogyakarta sebesar 476 mahasiswa. Dapat dikatakan bahwa mahasiswa yang berasal dari luar Yogyakarta merupakan jumlah yang sangat besar untuk menyelesaikan massa studi mahasiswa tersebut, karena banyak mahasiswa yang merantau untuk menempuh studi di Yogyakarta. Dari gambar 5.4 tersebut dapat dikatakan pula bahwa mahasiswa yang berasal dari Yogyakarta maupun Luar Yogyakarta cenderung menempuh masa studi dengan tidak tepat waktu atau > 4 tahun.



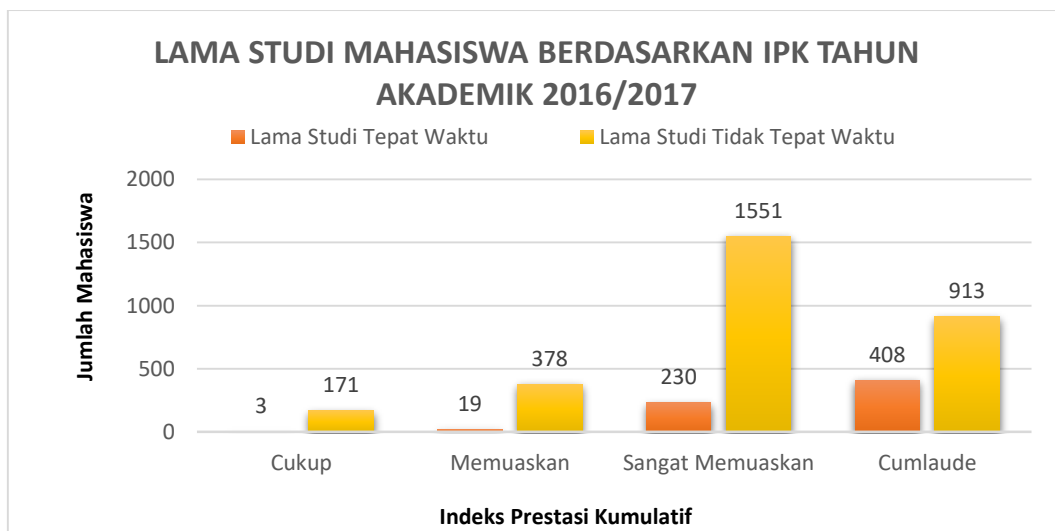
Gambar 5. 4 Lama Studi Mahasiswa Berdasarkan Jenis SMA

Berdasarkan **gambar 5.4** dapat dijelaskan bahwa alumni mahasiswa Universitas Islam Indonesia paling banyak berasal dari sekolah Negeri dengan jumlah kelulusan tepat waktu sebesar 485 mahasiswa dan tidak tepat waktu sebesar 1.866 mahasiswa. Sedangkan mahasiswa yang berasal dari sekolah Swasta paling sedikit untuk kelulusan mahasiswa tepat waktu dan tidak tepat waktu sebesar 65 mahasiswa dan 534 mahasiswa.



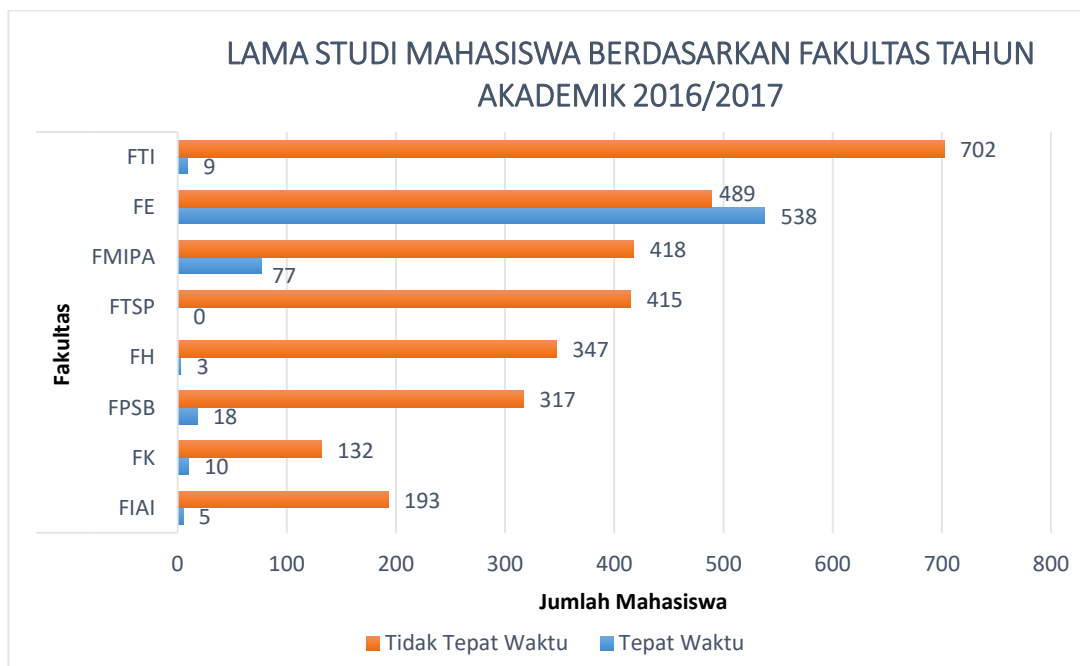
Gambar 5. 5 Lama Studi Mahasiswa Berdasarkan Jurusan SMA

Berdasarkan **gambar 5.5** dapat dilihat bahwa mahasiswa alumni yang jurusan ketika SMA adalah mahasiswa dengan jurusan IPA. Diketahui bahwa dengan jurusan IPA mahasiswa lulus tidak tepat waktu sebanyak 1.844 mahasiswa. Begitu juga dengan lama studi tepat waktu yang paling banyak adalah jurusan IPA sebesar 360 mahasiswa. Sedangkan untuk alumni mahasiswa yang paling sedikit lulus dengan tepat waktu dan tidak tepat waktu terdapat pada jurusan lain-lain yakni sebesar 69 mahasiswa dan 293 mahasiswa.



Gambar 5. 6 Lama Studi Mahasiswa Berdasarkan IPK

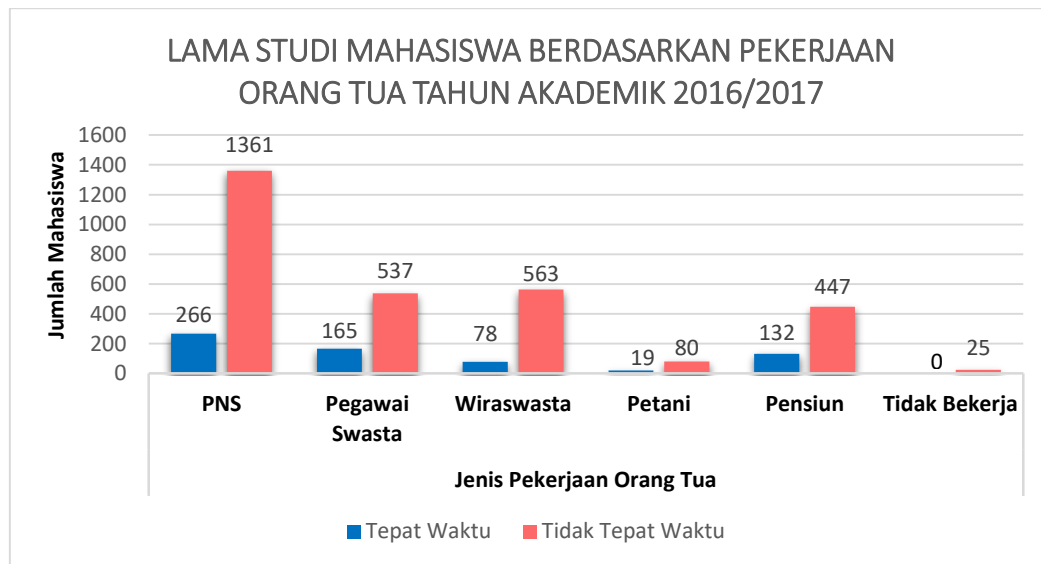
Pada **gambar 5.6** menjelaskan bahwa terdapat 4 kategori IPK yakni cukup, memuaskan, sangat memuaskan serta cumlaude. Berdasarkan gambar tersebut bahwa jumlah mahasiswa lulus tepat waktu paling banyak adalah kategori Cumlaude yakni 408 mahasiswa. Mahasiswa dengan IPK sangat memuaskan, memuaskan dan cukup untuk lulus tepat waktu sebanyak 230 mahasiswa, 19 mahasiswa dan 3 mahasiswa. Kemudian untuk mahasiswa yang tidak lulus tepat waktu paling banyak terdapat pada kategori sangat memuaskan yakni 1.551 mahasiswa. Kemudian untuk mahasiswa dengan IPK cumlaude, memuaskan dan cukup untuk mahasiswa tidak tepat waktu sebanyak 913 mahasiswa, 378 mahasiswa dan 171 mahasiswa.



Gambar 5. 7 Lama Studi Mahasiswa Berdasarkan Fakultas

Pada **gambar 5.7** dapat dilihat bahwa jumlah mahasiswa yang lulus tepat waktu terbanyak adalah Fakultas Ekonomi sebesar 538 mahasiswa, sedangkan yang paling sedikit yaitu Fakultas Teknik Sipil dan Perencanaan tidak memiliki mahasiswa yang lulus secara tepat waktu. Kemudian untuk mahasiswa yang lulus tidak tepat waktu terbanyak ada pada Fakultas Teknologi Industri sebesar 702 mahasiswa, sedangkan

untuk mahasiswa yang paling sedikit lulus tidak tepat waktu ada di Fakultas Kedokteran sebesar 132 mahasiswa.



Gambar 5. 8 Lama Studi Mahasiswa Berdasarkan Pekerjaan Orang Tua

Pada **gambar 5.8** dapat dilihat bahwa dari ke-6 kategori pekerjaan orang tua bahwa pekerjaan Pegawai Negeri dengan dengan kelulusan tidak tepat waktu maupun tepat waktu memiliki jumlah paling banyak. Dapat diketahui bahwa Pegawai Negeri Sipil dengan lama studi mahasiswa tepat waktu sebesar 266 mahasiswa dan lama studi mahasiswa tidak tepat waktu sebesar 1.361 mahasiswa. Berdasarkan **gambar 5.8** tersebut dapat dijelaskan bahwa pekerjaan orang tua terhadap lama studi tepat waktu dan tidak tepat waktu cenderung dengan pekerjaan sebagai Pegawai Negeri Sipil.

5.2 Analisis Regresi Logistik Biner

Analisis regresi logistik digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data dikotomik/biner dengan variabel bebas yang berupa data berskala interval dan atau kategorik (Hosmer dan Lemeshow, 1989). Data yang digunakan merupakan data alumni mahasiswa Universitas Islam Indonesia. Dari data tersebut akan dilakukan pengujian untuk mendapatkan suatu model, serta mengetahui peluang lama studi mahasiswa Universitas Islam Indonesia berdasarkan faktor-faktor

yang mempengaruhi serta menghasilkan klasifikasi berdasarkan akurasi dari regresi logistik tersebut.

Pada analisis Regresi Logistik Biner, data yang digunakan merupakan data alumni mahasiswa Universitas Islam Indonesia tahun akademik 2016/2017 dengan total data sebesar 3.673 mahasiswa. Data tersebut dibagi menjadi 2 bagian, yakni data *training* dan data *testing*. Data *training* adalah data yang siap untuk di-*mining* yang telah melewati data *preprocessing*. Sedangkan data *testing* adalah data yang digunakan untuk menguji *rule* klasifikasi yang diperoleh dari data *training*.

Tabel 5.1 Pembagian Data *Training* dan *Testing* Regresi Logistik Biner

Pembagian	Persentase	Total
Data pelatihan	80%	2.939
Data pengujian	20%	734
Total	100%	3.673

Berdasarkan tabel 5.1 diatas dapat diketahui bahwa persentase data pelatihan sebesar 80% lebih besar dari data pengujian yang hanya 20%, hal ini dikarenakan supaya algoritma pembelajaran dapat lebih terlatih ketika model didapatkan dari hasil pelatihan diujikan dapat memperoleh hasil yang lebih akurat.

5.2.1 Uji Signifikansi Parameter

5.2.1.1 Uji G (Overall)

Dalam pengujian model logistic dilakukan pengujian model secara keseluruhan dengan Uji G (metode likelihood). Statistik uji G menyebar menurut sebaran Chi Square (χ^2).

Tabel 5.2 Uji G (Overall)

χ^2	DF	Significant
310,03	9	0,000

Berdasarkan **tabel 5.2** dilakukan uji hipotesis sebagai berikut :

1. Hipotesis

$H_0: \beta_k = 0$, dengan $k = 1,2,3 \dots, p$. (Secara simultan variabel bebas tidak berpengaruh terhadap variabel terikat)

$H_1: \beta_k \neq 0$, dengan $k = 1,2,3 \dots, p$. (Minimal ada satu variabel bebas yang berpengaruh secara simultan terhadap variabel terikat)

2. Tingkat signifikansi

$$\alpha = 5\%$$

3. Daerah kritis

H_0 ditolak jika nilai *Significant* $< \alpha$, atau H_0 ditolak jika nilai $G^2 > \chi^2_{(db;\alpha)}$

4. Statistik uji

$$\textit{Significant} = 0.000$$

$$G^2 = 310,03$$

$$\chi^2_{(db;\alpha)} = \chi^2_{(9;0.05)} = 16,92$$

5. Keputusan

$\textit{Significant}(0.000) < \alpha (0.05)$, atau $G^2 (310,03) > \chi^2_{(9;0.05)}(16,92)$ maka tolak H_0

6. Kesimpulan

Dengan menggunakan tingkat kepercayaan 95% diketahui bahwa minimal ada satu variabel yang signifikan mempengaruhi variabel terikat atau variabel dependen (lama studi mahasiswa) pada ketepatan kelulusan mahasiswa Universitas Islam Indonesia, sehingga dapat disimpulkan bahwa model dapat digunakan untuk analisis lebih lanjut.

5.2.1.2 Uji Wald (Uji Parsial)

Dalam analisis regresi logistik, pengujian secara parsial menggunakan uji wald. Berikut ini merupakan variabel yang mempengaruhi lama studi mahasiswa : (Dilihat dari **tabel 5.3**)

1. Hipotesis

$H_0 : \beta_j = 0$, dengan $j = 1, 2, \dots, p$. (Tidak terdapat pengaruh variabel bebas terhadap variabel terikat)

$H_1: \beta_j \neq 0$, dengan $j = 1, 2, \dots, p$. (Terdapat pengaruh variabel bebas terhadap variabel terikat)

2. Tingkat signifikansi $\alpha = 5\%$

3. Daerah kritis

Tolak H_0 jika $P\text{-value} < \alpha$, atau Tolak H_0 jika $|Wald| > Z_{\frac{\alpha}{2}}$

4. Statistik uji

$$Z_{\frac{\alpha}{2}} = 1,96$$

a. Jenis Kelamin

- Perempuan

$$Wald = 7,142, P\text{-value} = 0,000$$

b. Asal Daerah

- Yogyakarta

$$Wald = 2,122, P\text{-value} = 0,034$$

c. Jenis SMA

- SMA Swasta

$$Wald = 1,07, P\text{-value} = 0,283$$

- SMA Negeri

$$Wald = 2,98, P\text{-value} = 0,003$$

d. Jurusan saat SMA

- IPS

$$Wald = 0,45, P\text{-value} = 0,654$$

- IPA

$$Wald = 3,54, P\text{-value} = 0,000$$

e. IPK

- Memuaskan

Wald = 1,046, *P-value* = 0.295

- Sangat Memuaskan

Wald = 3,062, *P-value* = 0.002

- Cumlaude

Wald = 4,572, *P-value* = 0.000

5. Keputusan

Tabel 5. 3 Uji Wald (Parsial)

Variabel	<i>P-value</i>	Tanda	α	Keputusan	Wald	Tanda	$Z\alpha/2$
Jenis kelamin (Perempuan)	0,000	<	0,05	Tolak H0	7,142	>	1,96
Asal daerah (Yogyakarta)	0,034	<	0,05	Tolak H0	2,122	>	1,96
Jenis_SMA (Swasta)	0,283	>	0,05	Gagal Tolak H0	1,07	<	1,96
Jenis SMA (Negeri)	0,003	<	0,05	Tolak H0	2,98	>	1,96
Jurusan saat SMA (IPS)	0,654	<	0,05	Gagal Tolak H0	0,45	<	1,96
Jurusan saat SMA (IPA)	0,000	>	0,05	Tolak H0	3,54	>	1,96
IPK (Memuaskan)	0,295	>	0,05	Gagal Tolak H0	34,51 1	>	1,96
IPK (Sangat Memuaskan)	0,002	<	0,05	Tolak H0	3,06	>	1,96
IPK (Cumlaude)	0,000	<	0,05	Tolak H0	4,57	>	1,96

6. Kesimpulan

Dengan menggunakan tingkat kepercayaan 95%, dapat disimpulkan bahwa semua variabel bebas yang diteliti berpengaruh terhadap variabel terikat yaitu Jenis Kelamin(Laki-laki), asal daerah (Yogyakarta), jenis SMA (Negeri), jurusan SMA (IPA), IPK (Sangat Memuaskan dan Cumlaude).

5.2.2 Uji Kecocokan Model

Uji kecocokan model atau uji goodness of fit digunakan untuk melihat ketepatan model yang digunakan. Uji goodness of fit dapat dilihat dari nilai *Hosmer dan Lemeshow test*. Nilai *Hosmer dan Lemeshow test* digunakan untuk melihat apakah data empiris cocok atau tidak dengan model.

Tabel 5.4 Uji Kecocokan Model

Chi Square (χ^2)	DF	Significant
7,814	8	0,451

Berdasarkan **tabel 5.4** dilakukan uji hipotesis sebagai berikut:

1. Hipotesis

H_0 : $\hat{\pi}_i = y_i$ atau model sesuai (tidak ada perbedaan nyata antara hasil observasi dengan kemungkinan prediksi model)

H_1 : $\hat{\pi}_i \neq y_i$ atau model tidak sesuai (ada perbedaan nyata antara hasil observasi dengan kemungkinan prediksi model)

2. Tingkat Signifikansi $\alpha = 5\%$

3. Daerah Kritis

Tolak H_0 jika $P - value < \alpha$, atau Tolak H_0 jika $\chi^2_{hitung} > \chi^2_{(db;\alpha)}$

4. Statistik Uji

$P - value = 0,451$

$\chi^2_{hitung} = 7,814$ $\chi^2_{(db;\alpha)} = \chi^2_{(8;0.05)} = 15,51$

5. Keputusan

$$P - \text{value}(0.451) > \alpha(0.05)$$

$\chi^2 \text{ hitung}(7,814) < \chi^2_{(8;0,05)}(15,51)$, maka gagal tolak H_0 .

6. Kesimpulan

Dengan menggunakan tingkat kepercayaan 95%, data yang ada menunjukkan bahwa data empiris cocok dengan model yang didapat, dengan kata lain model sesuai dengan data.

5.2.3 Model Regresi Logistik Biner

Dari hasil uji wald (uji parsial) dapat dinyatakan bahwa semua variabel yang diteliti berpengaruh signifikan terhadap ketepatan lama studi mahasiswa. Maka dapat dinyatakan dengan persamaan regresi logistik, sebagai berikut:

$$\begin{aligned} L_n \left(\frac{p}{1-p} \right) = & -3,994 + 0,796 Jns_{kelamin(perempuan)} + 0,246 asal_{daerah(Yogyakarta)} \\ & + 0,447 jenis_{SMA(Negeri)} - 0,673 jurusan_{SMA(IPA)} \\ & + 1,809 IPK(sangat memuaskan) + 2,702 IPK(cumlaude) \end{aligned}$$

Berdasarkan persamaan regresi logistik tersebut, maka dapat dihitung peluang kejadian terjadinya lama studi mahasiswa. Hasil persamaan regresi logistik kemudian diekspresikan untuk mengetahui kecocokan dengan nilai lama studi mahasiswa yaitu 1 (lulus tepat waktu) atau 0 (lulus tidak tepat waktu).

Peluang kejadian terjadinya ketepatan lama studi mahasiswa dapat dihitung menggunakan persamaan berikut:

$$\begin{aligned} \pi(x) &= \frac{\exp(-3,994 + 0,796 jns_{kelamin(perempuan)} + 0,2464 asal_{daerah(Yogyakarta)} + \\ & \quad 0,447 jenis_{SMA(Negeri)} - 0,673 jurusan_{SMA(IPA)} \\ & \quad + 1,809 IPK(sangat memuaskan) + 2,702 IPK(cumlaude)}{\exp(-3,994 + 0,796 jns_{kelamin(perempuan)} + 0,2464 asal_{daerah(Yogyakarta)} + \\ & \quad 1 + 0,447 jenis_{SMA(Negeri)} - 0,673 jurusan_{SMA(IPA)} \\ & \quad + 1,809 IPK(sangat memuaskan) + 2,702 IPK(cumlaude)} \end{aligned}$$

Dari persamaan diatas, maka peneliti dapat menghitung peluang kejadian terjadinya lama studi mahasiswa Universitas Islam Indonesia serta nilai *odds ratio*. Berikut ini dapat dilihat peluang kejadian lama studi mahasiswa dan nilai *odds ratio* pada **tabel 5.5** dan **tabel 5.6**.

a. Peluang kejadian lama studi mahasiswa Universitas Islam Indonesia

Tabel 5. 5 Peluang Kejadian Lama Studi Mahasiswa

Jenis Kelamin	Asal Daerah	Jenis SMA	Jurusan SMA	IPK	Tepat Waktu	Tidak Tepat Waktu
Laki-laki	Luar Yogyakarta	Selain Negeri	Selain IPA	Selain SM dan Cum	0.018	0.982
				Sangat Memuaskan	0.101	0.899
				Cumlaude	0.215	0.785
			IPA	Selain SM dan Cum	0.009	0.991
				Sangat Memuaskan	0.054	0.946
				Cumlaude	0.122	0.878
		Negeri	Selain IPA	Selain SM dan Cum	0.028	0.972
				Sangat Memuaskan	0.149	0.851
				Cumlaude	0.300	0.700
			IPA	Selain SM dan Cum	0.014	0.986
				Sangat Memuaskan	0.082	0.918
				Cumlaude	0.179	0.821
	Yogyakarta	Selain Negeri	Selain IPA	Selain SM dan Cum	0.023	0.977
				Sangat Memuaskan	0.125	0.875
				Cumlaude	0.259	0.741
			IPA	Selain SM dan Cum	0.011	0.989
				Sangat Memuaskan	0.068	0.932

Laki-laki	Yogyakarta	Negeri	Selain IPA	Cumlaude	0.152	0.848
				Selain SM dan Cum	0.035	0.965
				Sangat Memuaskan	0.183	0.817
			IPA	Cumlaude	0.354	0.646
				Selain SM dan Cum	0.018	0.982
				Sangat Memuaskan	0.102	0.898
Cumlaude	0.218	0.782				
Perempuan	Luar Yogyakarta	Selain Negeri	Selain IPA	Selain SM dan Cum	0.039	0.961
				Sangat Memuaskan	0.199	0.801
				Cumlaude	0.378	0.622
		IPA	Selain SM dan Cum	0.020	0.980	
			Sangat Memuaskan	0.112	0.888	
			Cumlaude	0.237	0.763	
	Negeri	Selain IPA	Selain SM dan Cum	0.060	0.940	
			Sangat Memuaskan	0.280	0.720	
			Cumlaude	0.487	0.513	
		IPA	Selain SM dan Cum	0.031	0.969	
			Sangat Memuaskan	0.165	0.835	
			Cumlaude	0.326	0.674	
	Yogyakarta	Selain Negeri	Selain IPA	Selain SM dan Cum	0.049	0.951
				Sangat Memuaskan	0.241	0.759
				Cumlaude	0.437	0.563
			IPA	Selain SM dan Cum	0.025	0.975
				Sangat Memuaskan	0.139	0.861
				Cumlaude		

Perempuan	Yogyakarta	Negeri	Cumlaude	0.284	0.716	
			Selain IPA	Selain SM dan Cum	0.075	0.925
				Sangat Memuaskan	0.332	0.668
		IPA	Cumlaude	0.549	0.451	
			Selain SM dan Cum	0.04	0.96	
			Sangat Memuaskan	0.202	0.798	
				Cumlaude	0.383	0.617

b. Odds Ratio

Odds adalah cara penyajian probabilitas yang menjelaskan probabilitas bahwa kejadian tersebut akan terjadi jika dibagi dengan probabilitas bahwa kejadian itu tidak akan terjadi, Nugraha (2013). Dengan kata lain Odds adalah rasio probabilitas sukses (π) terhadap probabilitas gagal ($1 - \pi$).

Tabel 5.6 Odds Ratio

Variabel	Exp (B)
Jenis kelamin (Perempuan)	2,216524
Asal daerah (Yogyakarta)	1,27936
Jenis SMA (Negeri)	1,564287
Jurusan SMA (IPA)	0,510002
IPK (Sangat Memuaskan)	6,104401
IPK (Cumlaude)	14,90967

Berdasarkan **tabel 5.6** Diatas dapat diketahui bahwa nilai *odds ratio* didapatkan dari nilai eksponensial (B). Berikut ini akan dijelaskan nilai *odds ratio* berdasarkan variabel-variabel yang mempengaruhi secara signifikan lama studi mahasiswa Universitas Islam Indonesia.

1. Variabel Jenis kelamin (perempuan) diketahui nilai *odds ratio* sebesar 2,216. Angka tersebut menjelaskan bahwa peluang pengaruh variabel jenis kelamin

(perempuan) terhadap mahasiswa lulus tepat waktu sebesar 2,216 kali dibandingkan dengan mahasiswa berjenis kelamin laki-laki.

2. Variabel Asal daerah (Yogyakarta) diketahui nilai *odds ratio* sebesar 1,279. Angka tersebut menjelaskan bahwa pengaruh variabel asal daerah (Yogyakarta) terhadap mahasiswa lulus tepat waktu sebesar 1,279 kali dibandingkan dengan mahasiswa berasal dari selain Yogyakarta.
3. Variabel jenis SMA (Negeri) diketahui nilai *odds ratio* sebesar 1,564. Angka tersebut menjelaskan bahwa pengaruh variabel jenis SMA (Negeri) terhadap mahasiswa lulus tepat waktu sebesar 1,564 kali dibandingkan dengan mahasiswa dari jenis SMA lainnya.
4. Variabel jurusan SMA (IPA) diketahui nilai *odds ratio* sebesar 0,510. Angka tersebut menjelaskan bahwa pengaruh variabel asal daerah (Yogyakarta) terhadap mahasiswa lulus tepat waktu sebesar 0,510 kali dibandingkan dengan mahasiswa jurusan SMA lainnya.
5. Variabel IPK (sangat memuaskan) diketahui nilai *odds ratio* sebesar 6,104. Angka tersebut menjelaskan pengaruh variabel IPK (sangat memuaskan) terhadap mahasiswa lulus tepat waktu sebesar 6,104 kali dibandingkan dengan IPK cukup dan IPK lainnya.
6. Variabel IPK (cumlaude) diketahui nilai *odds ratio* sebesar 14,909. Angka tersebut menjelaskan bahwa pengaruh variabel IPK (cumlaude) terhadap mahasiswa lulus tepat waktu sebesar 14,909 kali dibandingkan dengan IPK Cukup dan IPK lainnya.

5.2.4 Klasifikasi Regresi Logistik Biner

a. Klasifikasi Data Training

Berdasarkan data training sebanyak 2.939 mahasiswa, didapatkan bahwa mahasiswa yang lulus tepat waktu sebesar 41 dan lulus tidak tepat waktu sebesar 2.497. Dapat dilihat pada **tabel 5.7**.

Tabel 5.7 Prediksi Data Training Regresi Logistik

Tepat Waktu	Tidak Tepat Waktu
41	2.497

Kemudian data testing tersebut dilakukan ketepatan klasifikasi. Berikut ini merupakan klasifikasi dari Regresi Logistik Biner.

Tabel 5.8 Kalsifikasi Data *Training* Regresi Logistik Biner

Lama Studi	Tepat Waktu	Tidak Tepat Waktu	Total
Tepat Waktu	24	504	528
Tidak Tepat Waktu	17	2.394	2.411
Total	41	2.898	2.939

Pada **tabel 5.7** tersebut dapat dihitung nilai akurasi dan nilai error untuk klasifikasi data training Regresi Logistik Biner sebagai berikut.

$$\text{Akurasi} = \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{24 + 2.394}{2.939} = \frac{2.418}{2.939} = 0,8227 \quad (5.1)$$

$$\text{Error Rate} = \frac{\sum \text{Prediksi Salah}}{\sum (\text{benar} + \text{salah})} = \frac{504 + 17}{2.939} = \frac{521}{2.939} = 0.1772 \quad (5.2)$$

Berdasarkan perhitungan **5.1** dan **5.3** dapat dilihat bahwa hasil menunjukkan sebanyak 2.418 mahasiswa diklasifikasikan dengan benar sedangkan sebanyak 521 mahasiswa yang tidak diklasifikasi dengan benar. Diketahui pula persentase akurasi pengklasifikasian lama studi mahasiswa sebesar 82,27%. Sedangkan untuk nilai *error* dari hasil klasifikasi tersebut sebesar 17,72%.

b. Klasifikasi Data *Testing*

Berdasarkan data *testing* sebanyak 734 mahasiswa, didapatkan bahwa mahasiswa yang lulus tepat waktu sebesar 12 dan lulus tidak tepat waktu sebesar 722.

Tabel 5. 9 Prediksi Data *Testing* Regresi Logistik Biner

Tepat Waktu	Tidak Tepat Waktu
12	722

Kemudian dari data testing dilakukan ketepatan klasifikasi. Berikut ini merupakan klasifikasi dari Regresi Logistik Biner.

Tabel 5. 10 Klasifikasi Data *Testing* Regresi Logistik Biner

Lama Studi	Tepat Waktu	Tidak Tepat Waktu	Total
Tepat Waktu	7	125	132
Tidak Tepat Waktu	5	597	602
Total	12	722	734

$$\text{Akurasi} = \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{7 + 597}{734} = \frac{604}{734} = 0,8229 \quad (5.3)$$

$$\text{Error Rate} = \frac{\sum \text{Prediksi Salah}}{\sum (\text{benar} + \text{salah})} = \frac{5 + 125}{734} = \frac{130}{734} = 0.1771 \quad (5.4)$$

Berdasarkan perhitungan 5.3 dan 5.4 dapat dilihat bahwa hasil menunjukkan sebanyak 604 mahasiswa diklasifikasikan dengan benar sedangkan sebanyak 130 mahasiswa yang tidak diklasifikasi dengan benar. Diketahui pula persentase akurasi pengklasifikasian lama studi mahasiswa sebesar 82,29%. Sedangkan untuk nilai *error* dari hasil klasifikasi tersebut sebesar 17,71%.

5.3 *Naïve Bayes Classifier*

Pada analisis *Naïve Bayes Classifier*, data yang digunakan merupakan data alumni mahasiswa Universitas Islam Idonesia tahun 2000-2017 dengan total data sebesar 47.033 mahasiswa. Data tersebut dibagi menjadi 2 bagian, yakni data training dan data testing. Data training adalah data yang siap untuk di-mining yang telah melewati data *preprocessing*. Sedangkan data *testing* adalah data yang digunakan untuk menguji *rule* klasifikasi yang diperoleh dari data *training*.

Tabel 5.11 Pembagian Data *Training* dan Data *Testing* Naïve Bayes

Pembagian	Persentase	Total
Data pelatihan	80%	2.939
Data pengujian	20%	734
Total	100%	3.673

Berdasarkan **tabel 5.11** diatas dapat diketahui bahwa persentase data pelatihan sebesar 80% lebih besar dari data pengujian yang hanya 20%, hal ini dikarenakan supaya algoritma pembelajaran dapat lebih terlatih ketika model didapatkan dari hasil pelatihan diujikan dapat memperoleh hasil yang lebih akurat. Berikut ini adalah proporsi data pelatihan dan data pengujian.

Tabel 5.12 Proporsi Data *Training* dan *Testing* Naïve Bayes

Pembagian	Tepat Waktu	Tidak Tepat Waktu
Data Pelatihan	0,1796529	0,8203471
Data Pengujian	0,1798365	0,8201635

Dari **tabel 5.12** dapat dilihat bahwa masing-masing pembagian data pelatihan tepat waktu dan tidak tepat waktu sebesar 18% dan 82%. Begitupula untuk pembagian data pengujian, proporsi data tepat waktu dan tidak tepat waktu juga sebesar 18% dan 82%. Pada data pelatihan yang bersifat kategorik didapatkan probabilitas prior.

a. Probabilitas Lama Studi Mahasiswa (Y)

Tabel 5.13 Probabilitas Lama Studi Mahasiswa

Tepat Waktu	Tidak Tepat Waktu
0,180	0,820

Dapat dilihat dari **tabel 5.13** bahwa probabilitas kelas lama studi mahasiswa tepat waktu sebesar 0,12 dan probabilitas kelas lama studi mahasiswa tidak tepat waktu sebesar 0,82. Dari hasil probabilitas tersebut dapat

dikatakan bahwa, mahasiswa yang lulus tidak tepat waktu lebih banyak dari pada yang lulus tepat waktu.

b. Probabilitas Lama Studi dengan Jenis Kelamin

Tabel 5.14 Probabilitas Lama Studi dengan Jenis Kelamin

Jenis Kelamin	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Laki-laki	0,303	0,564
Perempuan	0,697	0,436

Berdasarkan **tabel 5.14** bahwa probabilitas kelas lama studi mahasiswa tepat waktu dengan jenis kelamin laki-laki sebesar 0,303 dan probabilitas kelas lama studi mahasiswa tidak tepat waktu dengan jenis kelamin laki-laki sebesar 0,564. Sedangkan untuk probabilitas lama studi mahasiswa tepat waktu dan tidak tepat waktu dengan jenis kelamin perempuan sebesar 0,697 dan 0,436. Oleh karena itu, dapat dikatakan mahasiswa yang lulus tepat waktu dengan jenis kelamin perempuan lebih besar dibandingkan mahasiswa berjenis kelamin laki-laki.

c. Probabilita Lama Studi dengan Asal Daerah

Tabel 5.15 Probabilita Lama Studi dengan Asal Daerah

Asal Daerah	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Luar Yogyakarta	0,714	0,771
Yogyakarta	0,286	0,229

Berdasarkan **tabel 5.15** bahwa probabilitas kelas lama studi mahasiswa tepat waktu dengan asal daerah luar Yogyakarta lebih besar dibandingkan dengan mahasiswa yang berasal dari daerah Yogyakarta yakni sebesar 0,714.

Begitu pula dengan probabilitas mahasiswa yang lulus tidak tepat waktu dengan asal luar Yogyakarta lebih besar daripada yang berasal dari Yogyakarta yakni sebesar 0,771.

d. Probabilitas Lama Studi dengan Jenis SMA

Tabel 5.16 Probabilitas Lama Studi dengan Jenis SMA

Jenis SMA	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Negeri	0,725	0,621
Swasta	0,098	0,177
Lain-lain	0,176	0,202

Dari **tabel 5.16** bahwa probabilitas kelas lama studi mahasiswa tepat waktu dan tidak tepat waktu dengan jenis SMA Negeri lebih besar dengan jenis SMA Swasta, maupun lainnya yakni sebesar 0,725 dan 0,621. Dapat dikatakan bahwa lama studi mahasiswa berpengaruh terhadap jenis SMA. Sebagian besar mahasiswa Universitas Islam Indonesia berasal dari jenis SMA Negeri.

e. Probabilitas Lama Studi dengan Jurusan Saat SMA

Tabel 5.17 Probabilitas Lama Studi dengan Jurusan Saat SMA

Jurusan Saat SMA	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
IPA	0,559	0,612
IPS	0,331	0,290
Lain-lain	0,110	0,098

Dari **tabel 5.17** dapat dilihat bahwa probabilitas kelas lama studi mahasiswa tepat waktu dengan jurusan IPA lebih besar dibandingkan dengan

jurusan IPS maupun lainnya sebesar 0,559. Begitupula untuk probabilitas kelas lama studi mahasiswa tidak tepat waktu dengan jurusan IPA juga lebih besar dibandingkan jurusan lainnya sebesar 0,612.

f. Probabilitas Lama Sudi dengan IPK

Tabel 5. 18 Probabilitas Lama Studi dengan IPK

IPK	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
Cukup	0,005	0,057
Memuaskan	0,246	0,122
Sangat Memuaskan	0,357	0,516
Cumlaude	0,611	0,303

Berdasarkan **tabel 5.18** bahwa probabilitas kelas lama studi mahasiswa tepat waktu dengan IPK Cumlaude lebih besar dibandingkan jenis IPK lainnya sebesar 0,611. Sedangkan untuk probabilitas kelas lama studi mahasiswa tidak tepat waktu dengan IPK sangat memuaskan lebih besar dibandingkan dengan jenis IPK lainnya sebesar 0,516.

g. Probabilitas Lama Sudi dengan Fakultas

Tabel 5. 19 Probabilitas Lama Studi dengan Fakultas

Fakultas	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
FE	0,814	0,159
FH	0,001	0,115
FIA	0,005	0,068

FK	0,013	0,046
FMIPA	0,123	0,138
FPSB	0,028	0,104
FTI	0,013	0,232
FTSP	0,000	0,138

Berdasarkan **tabel 5.19** bahwa probabilitas kelas lama studi mahasiswa tepat waktu dan tidak tepat waktu dengan Fakultas Ekonomi lebih besar dibandingkan jenis Fakultas lainnya sebesar 0,814 dan 0,159.

h. Probabilitas Lama Sudi dengan Pekerjaan Orang Tua

Tabel 5. 20 Probabilitas Lama Studi dengan Pekerjaan Orang Tua

Pekerjaan Orang Tua	Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu
PNS	0,424	0,454
Pegawai Swasta	0,236	0,174
Wiraswasta	0,112	0,189
Petani	0,027	0,028
Pensiun	0,201	0,148
Tidak Bekerja	0,000	0,008

Dari **tabel 5.20** dapat dilihat bahwa probabilitas kelas lama studi mahasiswa tepat waktu dan tidak tepat waktu dengan jenis pekerjaan orang tua PNS lebih

besar dibandingkan jenis pekerjaan orang tua dibidang lainnya sebesar 0,424 dan 0,454.

Adapun persamaan *Naïve Bayes Classifier*, ketika mengasusmsikan semua jenis variabel independen saling bebas satu sama lain sebagai berikut:

Peluang dengan kelas Tepat Waktu:

$$\begin{aligned} & P(JK, Asal Daerah, Jenis SMA, \dots, P. Ortu | \text{Tepat Waktu}) \\ &= P(0,697 | 0,180) \times P(0,286 | 0,180) \times P(0,559 | 0,180) \times P(0,559 | 0,180) \\ & \times P(0,814 | 0,180) \times P(0,611 | 0,180) \times P(0,112 | 0,180) \quad (5.5) \end{aligned}$$

Peluang dengan kelas Tidak Tepat Waktu:

$$\begin{aligned} & P(JK, Asal Daerah, Jenis SMA, \dots, Pekerjaan Ortu | \text{Tidak Tepat Waktu}) \\ &= P(0,436 | 0,820) \times P(0,229 | 0,820) \times P(0,621 | 0,820) \times P(0,612 | 0,820) \\ & \times P(0,159 | 0,820) \times P(0,303 | 0,820) \times P(0,189 | 0,820) \quad (5.6) \end{aligned}$$

Berikut ini dilakukan simulasi untuk perhitungan data testing.

- Jika diketahui mahasiswa Universitas Islam Indonesia berjenis kelamin perempuan, berasal dari Yogyakarta dengan Jenis SMA adalah Swasta dan Jurusan SMA IPA, Fakultas Ekonomi dengan IPK sangat memuaskan dan pekerjaan orang tua adalah Wiraswasta.

Probabilitas Tepat Waktu Mahasiswa Universitas Islam Indonesia:

$$P(X(\text{perempuan}) | C(\text{Tepat Waktu})) = 0,697$$

$$P(X(\text{Yogyakarta}) | C(\text{Tepat Waktu})) = 0,286$$

$$P(X(\text{Negeri}) | C(\text{Tepat Waktu})) = 0,725$$

$$P(X(\text{IPA}) | C(\text{Tepat Waktu})) = 0,559$$

$$P(X(\text{Ekonomi}) | C(\text{Tepat Waktu})) = 0,814$$

$$P(X(\text{Cumlaude}) | C(\text{Tepat Waktu})) = 0,611$$

$$P(X(\text{wiraswasta}) | C(\text{Tepat Waktu})) = 0,112$$

$$\begin{aligned}
& \text{Peluang Tepat Waktu} \\
& = 0,697 \times 0,286 \times 0,725 \times 0,559 \times 0,814 \times 0,611 \times 0,112 \times 0,180 \\
& = 0,0008 \qquad (5.7)
\end{aligned}$$

Probabilitas Tidak Tepat Waktu Mahasiswa Universitas Islam Indonesia:

$$P(X(\text{perempuan})|C(\text{Tidak Tepat Waktu})) = 0,436$$

$$P(X(\text{Yogyakarta})|C(\text{Tidak Tepat Waktu})) = 0,229$$

$$P(X(\text{Negeri})|C(\text{Tidak Tepat Waktu})) = 0,621$$

$$P(X(\text{IPA})|C(\text{Tidak Tepat Waktu})) = 0,612$$

$$P(X(\text{Ekonomi})|C(\text{Tidak Tepat Waktu})) = 0,159$$

$$P(X(\text{Cumlaude})|C(\text{Tidak Tepat Waktu})) = 0,303$$

$$P(X(\text{wiraswasta})|C(\text{Tidak Tepat Waktu})) = 0,189$$

$$\begin{aligned}
& \text{Peluang Tidak Tepat Waktu} \\
& = 0,436 \times 0,229 \times 0,621 \times 0,612 \times 0,159 \times 0,303 \times 0,189 \times 0,820 \\
& = 0,0002 \qquad (5.8)
\end{aligned}$$

Berdasarkan peluang lama studi mahasiswa tepat waktu maupun tidak tepat waktu, dapat diketahui bahwa peluang tepat waktu lebih besar dibandingkan dengan peluang tidak tepat waktu, maka salah satu data simulasi uji tersebut dikatakan bahwa lama studi mahasiswa Universitas Islam Indonesia berpeluang lulus dengan tepat waktu.

5.3.1 Klasifikasi Tabel *Naïve Bayes*

a. Klasifikasi *Naïve Bayes Data Training*

Berdasarkan model-model probabilitas yang didapatkan, dilakukan pengujian terhadap data training yang merupakan data yang dibagi 80% untuk

data training. Data training tersebut sebesar 2.939 data mahasiswa alumni Universitas Islam Indonesia. Berikut ini hasil prediksi dari data training kelulusan mahasiswa Universitas Islam Indonesia.

Tabel 5.21 Prediksi dari Data *Training Naïve Bayes*

Tepat Waktu	Tidak Tepat Waktu
442	2.497

Berdasarkan data training sebanyak 2.939 mahasiswa, didapatkan bahwa mahasiswa yang lulus tepat waktu sebesar 442 dan lulus tidak tepat waktu sebesar 2.497 mahasiswa. Kemudian data training tersebut dilakukan ketepatan klasifikasi. Berikut ini merupakan klasifikasi dari *Naïve Bayes Classifier*.

Tabel 5.22 Klasifikasi Tabel Data *Training Naïve Bayes*

Lama Studi	Tepat Waktu	Tidak Tepat Waktu	Total
Tepat Waktu	317	125	442
Tidak Tepat Waktu	211	2.286	2.497
Total	528	2.411	2.939

Pada **tabel 5.22** tersebut dapat dihitung nilai akurasi dan nilai error untuk klasifikasi *Naïve Bayes* sebagai berikut.

$$\text{Akurasi} = \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{317 + 2.286}{2.939} = \frac{2.603}{2.939} = 0,885 \quad (5.9)$$

$$\text{Error Rate} = \frac{\sum \text{Prediksi Salah}}{\sum (\text{benar} + \text{salah})} = \frac{125 + 211}{2.939} = \frac{336}{2.939} = 0.114 \quad (5.10)$$

Berdasarkan perhitungan **5.6** dan **5.7** dapat dilihat bahwa hasil menunjukkan sebanyak 2.603 mahasiswa diklasifikasikan dengan benar sedangkan sebanyak 336 mahasiswa yang tidak diklasifikasi dengan benar. Diketahui pula persentase akurasi pengklasifikasian lama studi mahasiswa

secara tepat waktu sebesar 88,5%. Sedangkan untuk nilai *error* dari hasil klasifikasi tersebut sebesar 11,4%.

b. Klasifikasi Naïve Bayes Data Testing

Berdasarkan model-model probabilitas yang didapatkan, dilakukan pengujian terhadap data testing yang dibagi menjadi 20%. Data pengujian tersebut sebesar 734 data mahasiswa alumni Universitas Islam Indonesia. Berikut ini merupakan hasil prediksi dari data testing kelulusan mahasiswa Universitas Islam Indonesia dapat dilihat pada **tabel 5.23**.

Tabel 5.23 Prediksi Data *Testing* Naïve Bayes

Tepat Waktu	Tidak Tepat Waktu
112	622

Berdasarkan data testing sebanyak 734 mahasiswa, didapatkan bahwa mahasiswa yang lulus tepat waktu sebesar 112 dan lulus tidak tepat waktu sebesar 622. Kemudian data testing tersebut dilakukan ketepatan klasifikasi. Berikut ini merupakan klasifikasi dari *Naïve Bayes Classifier*.

Tabel 5. 24 Prediksi Data *Testing* Naïve Bayes

Lama Studi	Tepat Waktu	Tidak Tepat Waktu	Total
Tepat Waktu	82	30	112
Tidak Tepat Waktu	50	572	622
Total	132	602	734

Berdasarkan **tabel 5.24** tersebut dapat dihitung nilai akurasi dan nilai error untuk klasifikasi Naïve Bayes sebagai berikut.

$$\text{Akurasi} = \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{82 + 572}{734} = \frac{654}{734} = 0,891 \quad (5.11)$$

$$\text{Error Rate} = \frac{\sum \text{Prediksi Salah}}{\sum (\text{benar} + \text{salah})} = \frac{30 + 50}{734} = \frac{80}{734} = 0.109 \quad (5.12)$$

Berdasarkan perhitungan 5.8 dan 5.9 dapat dilihat bahwa hasil menunjukkan sebanyak 656 mahasiswa diklasifikasikan dengan benar sedangkan sebanyak 78 mahasiswa yang tidak diklasifikasi dengan benar. Diketahui pula persentase akurasi pengklasifikasian lama studi mahasiswa secara tepat waktu sebesar 89,10%. Sedangkan untuk nilai *error* dari hasil klasifikasi tersebut sebesar 10,90%.

5.4 Perbandingan Kalsifikasi Regresi Logistik Biner dan Naïve Bayes

Berdasarkan analisis Regresi Logistik dan Naïve Bayes Classifier didapatkan hasil klasifikan kelulusan mahasiswa Universitas Islam Indonesia Tahun Akademik 2016/2017. Berikut ini perbandingan dari kedua metode yang telah digunakan.

Tabel 5.25 Perbandingan Klasifikasi Regresi Logisti dan Naïve Bayes

Metode	Ketepatan Klasifikasi
Regresi Logistik Biner	82,29%
Naïve Bayes Classifier	89,10%

Berdasarkan **tabel 5.23** maka dapat dikatakan bahwa metode terbaik dalam pengklasifikasian lama studi mahasiswa Universitas Islam Indonesia tahun akademik 2016/2017 yaitu menggunakan algoritma *Naïve Bayes* dengan tingkat akurasi sebesar 89,10%. Hal ini disebabkan, karena *Naïve Bayes* mengasumsikan bahwa semua variable independen yang digunakan merupakan variabel yang saling bebas satu sama lain dan mempengaruhi variabel Lama Studi Mahasiswa (dependen).

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, maka diperoleh kesimpulan untuk menjawab rumusan masalah yang ada, yaitu:

1. Diketahui bahwa persentase lama studi mahasiswa Tepat Waktu hanya sebesar 18% sedangkan untuk mahasiswa tidak tepat waktu sebesar 82%. Dari gambaran tersebut dapat dikatakan bahwa masih banyak mahasiswa yang lulus tidak tepat waktu, atau masa studinya lebih dari 4 tahun. Selain itu, mahasiswa yang berjenis kelamin perempuan lebih cenderung menyelesaikan masa studi dengan tepat waktu. Mahasiswa yang berasal dari daerah luar Yogyakarta masih banyak yang tidak dapat menyelesaikan studi tepat waktu. Jenis SMA Negeri serta Jurusan saat SMA adalah IPA paling banyak memiliki mahasiswa yang menyelesaikan masa studi tepat waktu dan tidak tepat waktu. Fakultas Teknologi Industri memiliki mahasiswa yang paling banyak menyelesaikan studi tidak tepat waktu, serta Pekerjaan Orang Tua adalah Pegawai Negeri Sipil memiliki mahasiswa yang paling banyak lulus tepat waktu.
2. Berdasarkan Analisis Regresi Logistik Biner, didapatkan faktor-faktor yang signifikan mempengaruhi lama studi mahasiswa antara lain Jenis Kelamin (perempuan), Asal Daerah (Yogyakarta), Jenis SMA (Negeri), Jurusan SMA (IPA), IPK (Sangat Memuaskan), IPK (Cumlaude). Dimana, dapat dijelaskan bahwa ketika mahasiswa perempuan lebih cenderung lulus tepat waktu sebesar 2,216 kali dibandingkan dengan mahasiswa berjenis kelamin laki-laki. Ketika mahasiswa berasal dari daerah Yogyakarta lebih cenderung lulus tepat waktu sebesar 2,279 kali dibandingkan mahasiswa berasal dari Luar Yogyakarta. Mahasiswa dengan latar belakang Jenis SMA Negeri lebih cenderung lulus tepat waktu sebesar 1,564 kali dibandingkan dengan jenis SMA lainnya. Ketika

mahasiswa berlatar belakang jurusan SMA IPA lebih cenderung lulus tepat waktu sebesar 0,510 kali dibandingkan dengan jurusan SMA lainnya. Ketika mahasiswa tersebut memiliki IPK sangat memuaskan, maka lebih cenderung lulus tepat waktu sebesar 6,104 kali dibandingkan IPK cukup. Serta mahasiswa dengan IPK *Cumlaude* akan lebih cenderung lulus tepat waktu sebesar 14,909 kali dibandingkan IPK cukup. Kemudian, diketahui pula ketepatan kalsifikasi dari model Regresi Logistik Biner untuk kelulusan mahasiswa pada data latih sebesar 82,27% dan data uji memberikan akurasi sebesar 82,29%.

3. Hasil klasifikasi menggunakan *Naïve Bayes Classifier* pada data latih memberikan tingkat akurasi sebesar 88,50%. Sedangkan untuk klasifikasi pada data uji memberikan tingkat akurasi sebesar 89,10%
4. Algoritma *Naïve Bayes Classifier* lebih baik dalam mengkalsifikasi lama studi mahasiswa Universitas Islam Indonesia karena akurasi data uji sebesar 89,10% lebih besar dibandingkan dengan akurasi data uji Regresi Logistik Biner sebesar 82,29%.

6.2 Saran

Adapun saran yang diperoleh, setelah dilakukan analisis data sebagai berikut:

1. Perlu penambahan variabel independen dalam memprediksi lama studi mahasiswa untuk mendapatkan hasil akurasi yang lebih baik.
2. Diharapkan dari penelitian ini dapat membantu pihak institusi perguruan tinggi untuk merancang strategi terbaik guna meningkatkan kualitas perguruan tinggi.
3. Mahasiswa yang berasal dari jenis fakultas teknik, kebanyakan lulus tidak tepat waktu, oleh karena itu, perlu adanya pertimbangan khusus kriteria terhadap mahasiswa baru yang ingin mendaftar jurusan teknik. Sebagian besar mahasiswa teknik berjenis kelamin laki-laki, tetapi dalam penelitian ini, mahasiswa yang berjenis kelamin laki-laki kebanyakan lulus tidak tepat waktu, oleh karena itu, perlu diimbangi antara jenis kelamin perempuan dan jenis kelamin laki-laki saat mahasiswa baru ingin mendaftar.

4. Diharapkan adanya aplikasi untuk memprediksi lama studi mahasiswa menggunakan data mahasiswa Universitas Islam Indonesia. Sehingga apabila mahasiswa yang diprediksi lulus tidak tepat waktu maka dapat diberikan bimbingan pembelajaran lebih lanjut.
5. Setelah dilakukan penelitian ini diharapkan dapat dikembangkan penelitian lain yang sejenis dengan metode yang berbeda sehingga dapat dilihat klasifikasi yang mana yang lebih baik.

DAFTAR PUSTAKA

- Akademik UII. 2018. Sekilas UII. Diakses pada tanggal 30 Januari 2018 pukul 09.00 WIB dari <https://www.uui.ac.id/sekilas-uui/>.
- Ahdika, Atina. 2015. Regresi Logistik Biner. Dikases pada tanggal 8 Februari 2018 pukul 20.00 WIB dari <http://atinaahdika.com/wp-content/uploads/2015/09/REGRESI-LOGISTIK-BINER.pdf>.
- Arikunto. 2002. Metodologi Penelitian Suatu Pendekatan Proposal. Jakarta : PT. Rineka Cipta.
- Berry, Michael W. and Murray Browne. (2006). *Lecture Notes in DATA MINING*. USA: *World Scientific*.
- Budiati, Diah dan Yuciana Wilandari. (2013). Analisis Lama Studi Mahasiswa Fsm Universitas Diponegoro Menggunailan Regresi Logistik Biner.
- Chandra, Fransisca. (2009). “Peran Partisipasi Kegiatan di Alam Masa anak, Pendidikan dan Jenis Kelamin sebagai Moderasi Terhadap Perilaku Ramah Lingkungan”. Disertasi S3. Program Magister Psikologi Fakultas Psikologi. Unversita Gadjah Mada Yogyakarta.
- Diaprina, Sistya Rosi dan Suhartono. (2014). Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner Dan *Radial Basis Function Network* di Bank “X” Cabang Kediri . Jurnal Sains dan Seni Pomits Vol. 3, No. 2, (2014) ISSN:2337-3539.
- Fais, Sukma Nur, Muhammad Aditya D dan Satria Mulya I. (2014). Klasifikasi Calon Pendoron Darah dengan Metode *Naïve Bayes Classifier*.
- Fajri, Fathorazi Nur. (2018). Perbandingan Sistem Klasifikasi *Naïve Bayes* dan *Decision Tree* Untuk Diagnosa Penyakit Diabetes. Diakses pada tanggal 30 Mei 2018 pukul 14.20 WIB dari https://www.academia.edu/16691034/Perbandingan_Sistem_Klasifikasi_Naive_Bayes_dan_Decision_Tree_Untuk_Diagnosa_Penyakit_Diabetes.

- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi *Data Mining* Model C4.5 dan *Naïve Bayes* Untuk Prediksi Penyakit Diabetes. *Jurnal Techno Nusa Mandiri* Vol. XIII, No. 1 Maret 2016.
- Han, Jiawei and Micheline Kamber. (2006). *Data Mining Concepts and Techniques*, second edition. California: Morgan Kaufman.
- Hosmer, David W. and Stanley Lemeshow. (2000). *Applied Logistik Regression* (2nd ed), Jakarta. John Wiley & Sons, New York.
- Imaslihkah, Sitti, Madu Ratna dan Vita R. (2013). Analisis Regresi Logistik Ordinal terhadap Faktor-faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya. *Jurnal Sains Dan Seni Pomits* Vol. 2, No. 2, (2013)
- Jananto, Arief. (2013). Algoritma *Naive Bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Jurnal Teknologi Informasi DINAMIK* Volume 18, No.1, Januari 2013 : 09-16
- Mauriza, Ahmad Fikri. (2014). Implementasi *Data Mining* Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan *Naïve Bayes*. Skripsi. Program Studi Teknik Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Yogyakarta.
- Melawati, Yuni. (2013). Klasifikasi Keputusan Nasabah Dalam Pengambilan Kredit Menggunakan Model Regresi Logistik Biner Dan Metode *Classification And Regression Trees (Chart)*. Skripsi. Jurusan Pendidikan Matematika Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia.
- Nachrowi, Djalal Nachrowi dan Hardius Usman. (2002). Penggunaan Teknik Ekonometri. Jakarta: PT Raja Grafindo Persada.
- Nugraha, Jaka. (2013). Pengantar Analiis Data Ketgorik. Sleman : CV BUDI UTAMA.

- Paskianti, Kristina. (2011). *Klasifikasi Dokumen Tumbuhan Obat menggunakan Algoritma KNN Fuzzy*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam IPB. Bogor.
- Prasetyo, Eko. (2012). *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI Yogyakarta.
- R, Shyara Taruna dan Saroj Hiranwal. (2013). *Enhanced Naive Bayes Algorithm for Intrusion Detection in Data Mining*. *International Journal of Computer Science and Information Technologies*, Vol. 4, 2013.
- Rajagukguk, Nanci. (2015). *Perbandingan Metode Klasifikasi Regresi Logistik Biner dan Naïve Bayes Pada Status Pengguna KB di Kota Tegal Tahun 2014*. *Jurnal Gaussian*, Volume 4, Nomor 2, Tahun 2015, Halaman 365 – 374.
- Ramandhani, Ridha. (2017). *Metode Bootstrap Aggregating Regresi Logistik Biner Untuk Klasifikasi Kesejahteraan Rumah Tangga di Kota Pati*. *Jurnal Gaussian*, Volume 6, Nomor 1, Tahun 2017, Halaman 121-130.
- Rish, I. 2006. *An empirical study of The Naive Bayes Classifier*. *International Joint Conference on Artificial Intelligence*, California, 41-46.
- Rizki, Agung Priyo. (2016). *Estimasi Parameter Regresi Variabel Dummy Menggunakan Metode Matriks Terboboti*. Skripsi. Jurusan Matematika Fakultas Sains Dan Teknologi Universitas Negeri Maulana Malik Ibrahim Malang.
- Rumaenda, Wella. (2016). *Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner Dan Algoritma C4.5*. *Jurnal Gaussian*, Volume 5, Nomor 2, Tahun 2016, Halaman 299-309.
- Samosir, Riama Oktaviani. (2015). *Perbandingan Klasifikasi Metode Regresi Logistik Biner Dan Radial Basis Function Neural Network Pada Berat Bayi Lahir Rendah*. *Jurnal Gaussian*, Volume 4, Nomor 4, Tahun 2015, Halaman 997-1005.
- Sary, Meydea Pragivta. (2017). *Analisis Pengaruh Ukuran Perusahaan, Debt To Total Asset, Profitabilitas, Sektor Industri, Dan Debt To Total Asset Pada Tindakan*

Perataan Laba Perusahaan Manufaktur Yang Terdaftar Di Bursa Efek Indonesia Tahun 2011-2015. Publikasi Ilmiah UMS.Fakultas Matematika Jurusan Statistika Universitas Islam Indonesia.

- Siswoyo, Dwi (2007). Ilmu Pendidikan. Yogyakarta: UNY Press.
- Sugiyono. (2007). Metode Penelitian Pendidikan, Pendekatan, Kuantitatif, Kualitatif,dan R&D. Bandung: ALFABETA.
- Supranto, J. (2004). Analisis Multivariat: Arti dan interpretasi. Jakarta: PT. Rineka Cipta.
- Syarli dan Asrul Ashari Mui. (2016). Metode *Naive Bayes* Untuk Prediksi Kelulusan. Jurnal Ilmiah Ilmu Komputer, Vol. 2, No. 1, April 2016.
- Mujiasih, Subekti. (2011). Pemanfaatan *Data Mining* Untuk Prakiraan Cuaca. Jurnal Meteorologi dan Geofisika, Volume12, Nomor 2, September 2011.
- Timur, Dhinda Amalia. (2014). Ketepatan Klasifikasi Keikutsertaan Keluarga Berencana (KB) Menggunakan Analisis Regresi Logistik Biner Dan *Fuzzy K-Nearest Neighbor In Every Class* Di Kabupaten Klaten. Jurnal Gaussian, Volume 3, Nomor 4, Tahun 2014, Halaman 615-62.
- Turban, Efraim. (2005). *Decision Support Systems and Intelligent Systems* Edisi Bahasa Indonesia Jilid 1. Andi: Yogyakarta.
- Undang-Undang Republik Indonesia Nomor 12. (2012). Tentang Pendidikan Tinggi
- Undang Undang Republik Indonesia Nomor 20. (2003). Tentang Sistem Pendidikan Nasional.
- Walpole, Ronald E. (1995). Pengantar Statistik Edisi Ke-4. Jakarta : PT Gramedia.
- Widarjono, Agus. (2010). Analisis Statistika Multivariat Terapan. Yogyakarta: Sekolah Tinggi Ilmu Manajemen YKPN.
- Wijaya , Akhmad Pandhu dan Heru Agus Santoso. (2016). *Naive Bayes Classification* pada Klasifikasi Dokumen Untuk Identifikasi Konten *E-Government*. Publikasi Ilmiah. Jurusan Teknik Informatika, Universitas Dian Nuswantoro Semarang.

LAMPIRAN

Lampiran 1. Data Penelitian

(Data Kelulusan Mahasiswa S1 tahun akademis 2016/2017 Universitas)

No	Lama_Studi	JK	IPK	Asal Daerah	Jns_SMA	Jur_SMA	Fakultas	Pekerjaan Ortu
1	Tidak Tepat Waktu	Laki-laki	Memuaskan	Yogyakarta	Swasta	IPA	FTSP	Wiraswasta
2	Tidak Tepat Waktu	Perempuan	Sangat Memuaskan	Yogyakarta	Swasta	IPA	FE	Wiraswasta
3	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Yogyakarta	Swasta	IPA	FTI	Wiraswasta
4	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Yogyakarta	Swasta	IPA	FTI	Wiraswasta
5	Tidak Tepat Waktu	Perempuan	Sangat Memuaskan	Yogyakarta	Swasta	IPA	FE	Wiraswasta
6	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Yogyakarta	Swasta	IPS	FE	Wiraswasta
7	Tidak Tepat Waktu	Perempuan	Sangat Memuaskan	Yogyakarta	Swasta	IPS	FH	Wiraswasta
8	Tidak Tepat Waktu	Perempuan	Sangat Memuaskan	Luar Yogyakarta	Negeri	IPA	FTI	Wiraswasta
9	Tepat Waktu	Perempuan	Memuaskan	Yogyakarta	Negeri	IPA	FK	Wiraswasta
10	Tidak Tepat Waktu	Perempuan	Cumlaude	Luar Yogyakarta	Negeri	IPA	FH	Wiraswasta
11	Tidak Tepat Waktu	Laki-laki	Cumlaude	Yogyakarta	Negeri	IPA	FTI	Wiraswasta
12	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Yogyakarta	Negeri	IPS	FE	Wiraswasta

13	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Luar Yogyakarta	Negeri	IPA	FTI	Wiraswasta
14	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Luar Yogyakarta	Negeri	IPA	FTSP	Wiraswasta
15	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Luar Yogyakarta	Negeri	IPA	FTSP	Wiraswasta
...
...
...
...
...
...
3669	Tidak Tepat Waktu	Perempuan	Cumlaude	Luar Yogyakarta	Swasta	IPA	FPSB	Pegawai Swasta
3670	Tidak Tepat Waktu	Laki-laki	Memuaskan	Luar Yogyakarta	Swasta	IPS	FH	Pegawai Swasta
3671	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Yogyakarta	Swasta	IPA	FTI	Pegawai Swasta
3672	Tidak Tepat Waktu	Laki-laki	Sangat Memuaskan	Luar Yogyakarta	Swasta	IPS	FE	Pegawai Swasta
3673	Tidak Tepat Waktu	Perempuan	Cumlaude	Luar Yogyakarta	Swasta	IPS	FH	Pegawai Swasta

Lampiran 2. Sintaks Regresi Logistik Biner

```

#Regresi Logistik Biner
dataeren=read.delim("clipboard")
str(dataeren)
dataeren$Lama_Studi<-as.factor(dataeren$Lama_Studi)
dataeren$JNS_KELAMIN<-as.factor(dataeren$JNS_KELAMIN)
dataeren$IP_KUM<- as.factor(dataeren$IP_KUM)
dataeren$ASAL_DAERAH<- as.factor(dataeren$ASAL_DAERAH)
dataeren$JENIS_SMA<- as.factor(dataeren$JENIS_SMA)
dataeren$JUR_SMA<- as.factor(dataeren$JUR_SMA)
dataeren$FAKULTAS<- as.factor(dataeren$FAKULTAS)
dataeren$PEKERJAAN_ORTU<-as.factor(dataeren$PEKERJAAN_ORTU)
#Dummy
contrasts(dataeren$IP_KUM)
contrasts(dataeren$JNS_KELAMIN)
contrasts(dataeren$FAKULTAS)
contrasts(dataeren$ASAL_DAERAH)
contrasts(dataeren$JENIS_SMA)
contrasts(dataeren$JUR_SMA)
contrasts(dataeren$PEKERJAAN_ORTU)
str(dataeren) #Untuk melihat struktud data

#Crosstabs setiap variabel x dengan y
xtabs(~Lama_Studi+JNS_KELAMIN, data=dataeren)
xtabs(~Lama_Studi+IP_KUM, data=dataeren)
xtabs(~Lama_Studi+ASAL_DAERAH, data=dataeren)
xtabs(~Lama_Studi+JENIS_SMA, data=dataeren)
xtabs(~Lama_Studi+JUR_SMA, data=dataeren)
xtabs(~Lama_Studi+FAKULTAS, data=dataeren)
xtabs(~Lama_Studi+PEKERJAAN_ORTU, data=dataeren)
#Partisi Data
library (caret)
set.seed(2018)
inTrain1 <- createDataPartition(dataeren$Lama_Studi, p=0.8,
list=FALSE)
training <- dataeren [inTrain1,]
testing <- dataeren [-inTrain1,]
View(training)
#Training Regresi Logistik
training=read.delim("clipboard")
str(training)
training$Lama_Studi <- as.factor(training$Lama_Studi)
training$JNS_KELAMIN <-as.factor(training$JNS_KELAMIN)
training$IP_KUM<- as.factor(training$IP_KUM)
training$ASAL_DAERAH <-as.factor(training$ASAL_DAERAH)
training$JENIS_SMA <- as.factor(training$JENIS_SMA)
training$JUR_SMA <- as.factor(training$JUR_SMA)
training$FAKULTAS <- as.factor(training$FAKULTAS)
training$PEKERJAAN_ORTU <-as.factor(training$PEKERJAAN_ORTU)

```

```

contrasts(training$IP_KUM)
contrasts(training$JNS_KELAMIN)
contrasts(training$FAKULTAS)
contrasts(training$ASAL_DAERAH)
contrasts(training$JENIS_SMA)
contrasts(training$JUR_SMA)
contrasts(training$PEKERJAAN_ORTU)
contrasts(testing$IP_KUM)
str(training)
#UJI PARSIAL (Wald)
mylogit=glm(Lama_Studi~IP_KUM+JNS_KELAMIN+JENIS_SMA+ASAL_DAERAH+
            JUR_SMA,
            data=training,
            family="binomial")      #(Var Fakultas dan Pekerjaan
Ortu dikeluarkan, karena tidak signifikan)
summary(mylogit)
#Uji Overall
library(rms)
lrn(mylogit)
#Uji Goodness of fit
library(MKmisc)
HLgof.test(fit = fitted(mylogit), obs = training$Lama_Studi)
#Confussion Matrix
res<-predict(mylogit,training,type="response")
res
head(training)
table(Actualvalues=training$Lama_Studi, Predictedvalue=res>0.5)
#testing data
testing=read.delim("clipboard")
str(testing)
testing$Lama_Studi <- as.factor(testing$Lama_Studi)
testing$JNS_KELAMIN <- as.factor(testing$JNS_KELAMIN)
testing$IP_KUM<- as.factor(testing$IP_KUM)
testing$ASAL_DAERAH <- as.factor(testing$ASAL_DAERAH)
testing$JENIS_SMA <- as.factor(testing$JENIS_SMA)
testing$JUR_SMA <- as.factor(testing$JUR_SMA)
contrasts(testing$IP_KUM)
contrasts(testing$JNS_KELAMIN)
contrasts(testing$FAKULTAS)
contrasts(testing$ASAL_DAERAH)
contrasts(testing$JENIS_SMA)
#Confussion Matrix
restest<-predict(mylogit,testing,type="response")
restest
head(testing)
restest[1]
table(Actualvalues=testing$Lama_Studi, Predictedvalue=restest>0.5)

```

Lampiran 3 Sintaks *Naïve Bayes Classifier*

```

#Naive Bayes
dataeren=read.delim ("clipboard")
dataeren
summary(dataeren)
str(dataeren)
dataeren$Lama_Studi <- as.factor(dataeren$Lama_Studi)
dataeren$JNS_KELAMIN <-as.factor(dataeren$JNS_KELAMIN)
dataeren$IP_KUM<- as.factor(dataeren$IP_KUM)
dataeren$ASAL_DAERAH <-as.factor(dataeren$ASAL_DAERAH)
dataeren$JENIS_SMA <- as.factor(dataeren$JENIS_SMA)
dataeren$JUR_SMA <- as.factor(dataeren$JUR_SMA)
dataeren$FAKULTAS <- as.factor(dataeren$FAKULTAS)
dataeren$PEKERJAAN_ORTU <- as.factor(dataeren$PEKERJAAN_ORTU)
str(dataeren)
#Partisi Data
table (dataeren$Lama_Studi)
library (caret)
set.seed(2018)
inTrain1 <- createDataPartition(dataeren$Lama_Studi, p=0.8,
list=FALSE)
dataTrain1 <- dataeren [inTrain1,]
dataTest1 <- dataeren [-inTrain1,]
View(dataTrain1)
#Checks the row and proportion of arget variable for both
training and testing dataset
nrow(dataTrain1)
nrow(dataTest1)
prop.table(table(dataTrain1$Lama_Studi))
prop.table(table(dataTest1$Lama_Studi))

library(e1071)
library(naivebayes)

#modelbuilding
predictmodel <- naiveBayes(Lama_Studi ~ ., data=dataTrain1)
#check summary of the model
Predictmodel
#check probability setiap variabel
#jns_SMA
prop.table(table(dataTrain1$JENIS_SMA,dataTrain1$Lama_Studi),2)
table(dataTrain1$JENIS_SMA,dataTrain1$Lama_Studi)
#training
predictT <- predict (predictmodel, dataTrain1)
predict
View(predictT)
#print the confusion matrix training
print(confusionMatrix(predictT,dataTrain1$Lama_Studi,
positive ="Tepat Waktu", dnn =c
("Prediction","TRUE")))
```

```
#prediction data testing
predictY <-predict(predictmodel, dataTest1)
predict
#check prediction for the first top 5 rows in the testing data
head(predictY, n=5)
View(predictY)
#print the confusion matrix data testing
print(confusionMatrix(predictY,dataTest1$Lama_Studi,
                      positive = "Tepat Waktu", dnn = c
("Prediction","TRUE")))
```

Lampiran 4. Output Regresi Logistik Biner

1. Uji Overall

```
> library(rms)
> lrm(mylogit)
Logistic Regression Model

lrm(formula = mylogit)

              Model Likelihood      Discrimination      Rank Discrim.
              Ratio Test              Indexes              Indexes
obs          2939  LR chi2      310.03  R2          0.164  C          0.733
0            2411  d.f.          9      g          1.138  Dxy         0.465
1            528  Pr(> chi2) <0.0001  gr         3.121  gamma      0.483
max |deriv| 4e-10                                     gp         0.138  tau-a      0.137
                                      Brier       0.131

              Coef   S.E.   Wald Z  Pr(>|Z|)
Intercept   -3.9938  0.6052  -6.60  <0.0001
IP_KUM=1     0.6809  0.6511   1.05  0.2956
IP_KUM=2     1.8090  0.5909   3.06  0.0022
IP_KUM=3     2.7020  0.5911   4.57  <0.0001
JNS_KELAMIN=1 0.7959  0.1114   7.14  <0.0001
JENIS_SMA=1  -0.2179  0.2028  -1.07  0.2827
JENIS_SMA=2   0.4474  0.1503   2.98  0.0029
ASAL_DAERAH=1 0.2464  0.1161   2.12  0.0338
JUR_SMA=1    -0.0884  0.1972  -0.45  0.6540
JUR_SMA=2    -0.6733  0.1903  -3.54  0.0004
```

2. Uji Parsial

```
> summary(mylogit)

Call:
glm(formula = Lama_Studi ~ IP_KUM + JNS_KELAMIN + JENIS_SMA +
     ASAL_DAERAH + JUR_SMA, family = "binomial", data = training)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2626  -0.6413  -0.4618  -0.2393   2.8228

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.99378    0.60515  -6.600  4.12e-11 ***
IP_KUM1       0.68092    0.65102   1.046  0.295600
IP_KUM2       1.80901    0.59087   3.062  0.002202 **
IP_KUM3       2.70201    0.59103   4.572  4.84e-06 ***
JNS_KELAMIN1  0.79594    0.11145   7.142  9.21e-13 ***
JENIS_SMA1   -0.21788    0.20281  -1.074  0.282684
JENIS_SMA2    0.44743    0.15026   2.978  0.002904 **
ASAL_DAERAH1  0.24636    0.11610   2.122  0.033835 *
JUR_SMA1     -0.08836    0.19715  -0.448  0.654012
JUR_SMA2     -0.67334    0.19028  -3.539  0.000402 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2767.8  on 2938  degrees of freedom
Residual deviance: 2457.7  on 2929  degrees of freedom
AIC: 2477.7

Number of Fisher Scoring iterations: 6
```

3. Uji Goodness of fit (Kecocokan Model)

```
Hosmer-Lemeshow C statistic

data: fitted(mylogit) and training$Lama_Studi
X-squared = 7.8141, df = 8, p-value = 0.4518
```

4. *Confusion Marix Data Training*

```
> table(Actualvalues=training$Lama_Studi, Predictedvalue=res>0.5)
      Predictedvalue
Actualvalues FALSE TRUE
0          2394    17
1           504    24
```

5. *Confusion Matrix Data Testing*

```
> table(Actualvalues=testing$Lama_Studi, Predictedvalue=restest>0.5)
      Predictedvalue
Actualvalues  FALSE TRUE
tepat waktu    125    7
tidak tepat waktu 597    5
```

Lampiran 5. Output Naïve Bayes Classifier

1. Jumlah Lama Studi Mahasiswa untuk Data Kelulusan Mahasiswa Tahu akademik 2016/2017

```
> table (dataeren$Lama_studi)

Tepat waktu Tidak Tepat waktu
      660           3013
```

2. Proporsi Data *Training* dan Data *Testing*

```
> prop.table(table(dataTrain1$Lama_studi))

Tepat waktu Tidak Tepat waktu
  0.1796529    0.8203471

> prop.table(table(dataTest1$Lama_studi))

Tepat waktu Tidak Tepat waktu
  0.1798365    0.8201635
```

3. *A Priori Probabilities*

```
A-priori probabilities:
Y
  Tepat waktu Tidak Tepat waktu
    0.1796529    0.8203471

Conditional probabilities:
          JNS_KELAMIN
Y          Laki-laki Perempuan
Tepat waktu  0.3030303  0.6969697
Tidak Tepat waktu 0.5636665  0.4363335

          IP_KUM
Y          Cukup Cumlaude Memuaskan Sangat Memuaskan
Tepat waktu  0.005681818  0.611742424  0.024621212    0.357954545
Tidak Tepat waktu 0.057237661  0.303608461  0.122355869    0.516798009

          ASAL_DAERAH
Y          Luar Yogyakarta Yogyakarta
Tepat waktu  0.7140152  0.2859848
Tidak Tepat waktu 0.7714641  0.2285359

          JENIS_SMA
Y          Lain-lain Negeri Swasta
Tepat waktu  0.17613636  0.72537879  0.09848485
Tidak Tepat waktu 0.20157611  0.62131895  0.17710494

          JUR_SMA
Y          IPA IPS LAIN-LAIN
Tepat waktu  0.5587121  0.3314394  0.1098485
Tidak Tepat waktu 0.6117793  0.2903360  0.0978847

          FAKULTAS
Y          FE FH FIAI FK FMIPA FPSB FTI FTSP
Tepat waktu  0.814393939  0.001893939  0.005681818  0.013257576  0.123106061  0.028409091  0.013257576  0.000000000
Tidak Tepat waktu 0.159270012  0.115719618  0.066777271  0.046038988  0.137702198  0.104106180  0.232268768  0.138116964

          PEKERJAAN_ORTU
Y          Pegawai Swasta Pensiun Petani PNS Tidak Bekerja wiraswasta
Tepat waktu  0.236742424  0.200757576  0.026515152  0.424242424  0.000000000  0.111742424
Tidak Tepat waktu 0.173786810  0.147656574  0.027789299  0.453753629  0.007880547  0.189133140
```


4. *Confusion Matrix* untuk Data Training

```
> #print the confusion matrix training
> print(confusionMatrix(predictT,dataTrain1$Lama_Studi,
+           positive ="Tepat waktu", dnn =c ("Prediction","TRUE")))
Confusion Matrix and Statistics

          TRUE
Prediction  Tepat waktu Tidak Tepat waktu
Tepat waktu      317          125
Tidak Tepat waktu  211          2286

Accuracy : 0.8857
```

5. *Confusion Matrix* untuk Data Testing

```
> view(predictY)
> #print the confusion matrix
> print(confusionMatrix(predictY,dataTest1$Lama_Studi,
+           positive ="Tepat waktu", dnn =c ("Prediction","TRUE")))
Confusion Matrix and Statistics

          TRUE
Prediction  Tepat waktu Tidak Tepat waktu
Tepat waktu      82          30
Tidak Tepat waktu  50          572

Accuracy : 0.891
```