

**PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER
UNTUK MEMPREDIKSI MASA STUDI MAHASISWA**

Studi Kasus: Universitas Amikom Yogyakarta

SKRIPSI



disusun oleh

Irkham Widhi Saputro

15.11.8711

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

**PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER
UNTUK MEMPREDIKSI MASA STUDI MAHASISWA**

Studi Kasus: Universitas Amikom Yogyakarta

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S1
pada Program Studi Informatika



disusun oleh

Irkham Widhi Saputro

15.11.8711

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER UNTUK MEMPREDIKSI MASA STUDI MAHASISWA (Studi Kasus: Universitas Amikom Yogyakarta)

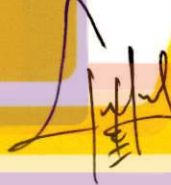
yang dipersiapkan dan disusun oleh

Irkham Widhi Saputro

15.11.8711

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Juni 2018

Dosen Pembimbing,



Bety Wulan Sari, M.Kom.

NIK. 190302254

PENGESAHAN

SKRIPSI

**PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER
UNTUK MEMPREDIKSI MASA STUDI MAHASISWA
(Studi Kasus: Universitas Amikom Yogyakarta)**

yang dipersiapkan dan disusun oleh

Irkham Widhi Saputro

15.11.8711

telah dipertahankan di depan Dewan Penguji
pada tanggal 12 November 2018

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mulia Sulistiyono, M.Kom.
NIK. 190302248

Erni Seniwati, M.Cs.
NIK. 190302231

Bety Wulan Sari, M.Kom.
NIK. 190302254



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 November 2018

DEKAN FAKULTAS ILMU KOMPUTER




Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertanda tangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 16 November 2018



Irkham Widhi Saputro

NIM. 15.11.8711

MOTTO

“Jangan takut untuk membuat kesalahan,
tapi pastikan untuk tidak melakukan kesalahan yang sama dua kali”

– **Akio Morita**

“What does not kill me makes me stronger”

– **Friedrich Nietzsche**

PERSEMBAHAN

Alhamdulillahirrabil' alamin,

Terimakasih ya Allah untuk segala rahmat, nikmat, dan karunia-Mu yang telah mengizinkanku untuk menyelesaikan kuliahku di kampus tercinta Universitas AMIKOM Yogyakarta.

Satu cita telah tercapai, namun perjalanan panjang ini masih belum selesai. Titik ini merupakan awal dari perjuanganku yang lain, perjuangan untuk melakukan hal yang lebih besar. Sebagai ucapan terima kasihku atas capaian ini, kupersembahkan karya kecil ini kepada orang-orang yang spesial.

- Terima kasih untuk Bapak, Ibu, dan Adikku yang telah memberikan doa, semangat, serta nasehat yang teramat sangat saya butuhkan. Kalianlah yang telah menjadi sosok motivasi bagi hidupku.
- Untuk Budhe Cici yang telah memberikan doa, semangat, serta dukungan moril untuk kelancaran kuliahku.
- Buat teman-teman Universitas AMIKOM Yogyakarta, terutama Yoga, dan Alif yang telah menjadi saingan untuk segera menyelesaikan tugas ini, teman-teman kelas S1 Informatika angkatan 2015 khususnya kelas 04, yang telah menjadi sahabat seperjuangan selama menempuh perkuliahan.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah subhanahu wa ta'ala yang telah melimpahkan nikmat dan rahmat-Nya yang melimpah, sehingga penulis dapat menyelesaikan Skripsi yang berjudul "*Penerapan Algoritma Naïve Bayes Classifier untuk Memprediksi Masa Studi Mahasiswa*" dengan tepat waktu.

Makalah ini berisi tentang penerapan teknik *data mining* menggunakan algoritma Naïve Bayes, dari mulai tahap pengumpulan data, hingga tahap implementasi ke dalam program.

Penulis juga ingin menyampaikan rasa terima kasih yang sedalam-dalamnya kepada semua pihak yang telah membantu dalam pembuatan skripsi ini, khususnya kepada:

1. Ibu Krisnawati, S.Si, MT, selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Bapak Sudarmawan, MT, selaku Ketua Program Studi S1 Informatika.
3. Ibu Windha Mega Pradnya D, M.Kom, selaku Sekretaris Program Studi S1 Informatika.
4. Ibu Armadyah Amborowati, S.Kom, M.Eng, selaku Direktur Administrasi Akademik dan Kemahasiswaan yang telah memberikan izin penelitian.
5. Bapak Drs. Asro Nasiri, M.Kom, selaku Direktur Innovation Center Universitas AMIKOM Yogyakarta yang telah memberikan data penelitian.
6. Ibu Bety Wulan Sari, M.Kom, selaku Dosen Pembimbing yang telah membimbing penulis dalam proses pembuatan Skripsi.

7. Seluruh Dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmu yang bermanfaat.
8. Bapak, Ibu, dan Adik tersayang yang selalu memberikan doa, dan dukungannya.
9. Teman-teman S1-IF angkatan 2015 kelas 04 yang telah menjadi keluarga selama proses akademik.
10. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari sepenuhnya bahwa skripsi ini banyak mengandung kekurangan, untuk itu segala kritik, dan saran yang membangun sangat diharapkan. Akhir kata, penulis berharap semoga makalah ini dapat memberikan manfaat bagi pihak yang memerlukan.

Yogyakarta, 16 November 2018

Irkham Widhi Saputro

DAFTAR ISI

JUDUL.....	i
PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN.....	iv
MOTTO.....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	4
1.5 Manfaat Penulisan.....	4
1.6 Metode Penelitian.....	5
1.6.1 Metode Pengumpulan Data.....	5
1.6.2 Tahapan Penelitian.....	6
1.7 Sistematika Penulisan.....	8
BAB II LANDASAN TEORI.....	10
2.1 Tinjauan Pustaka.....	10

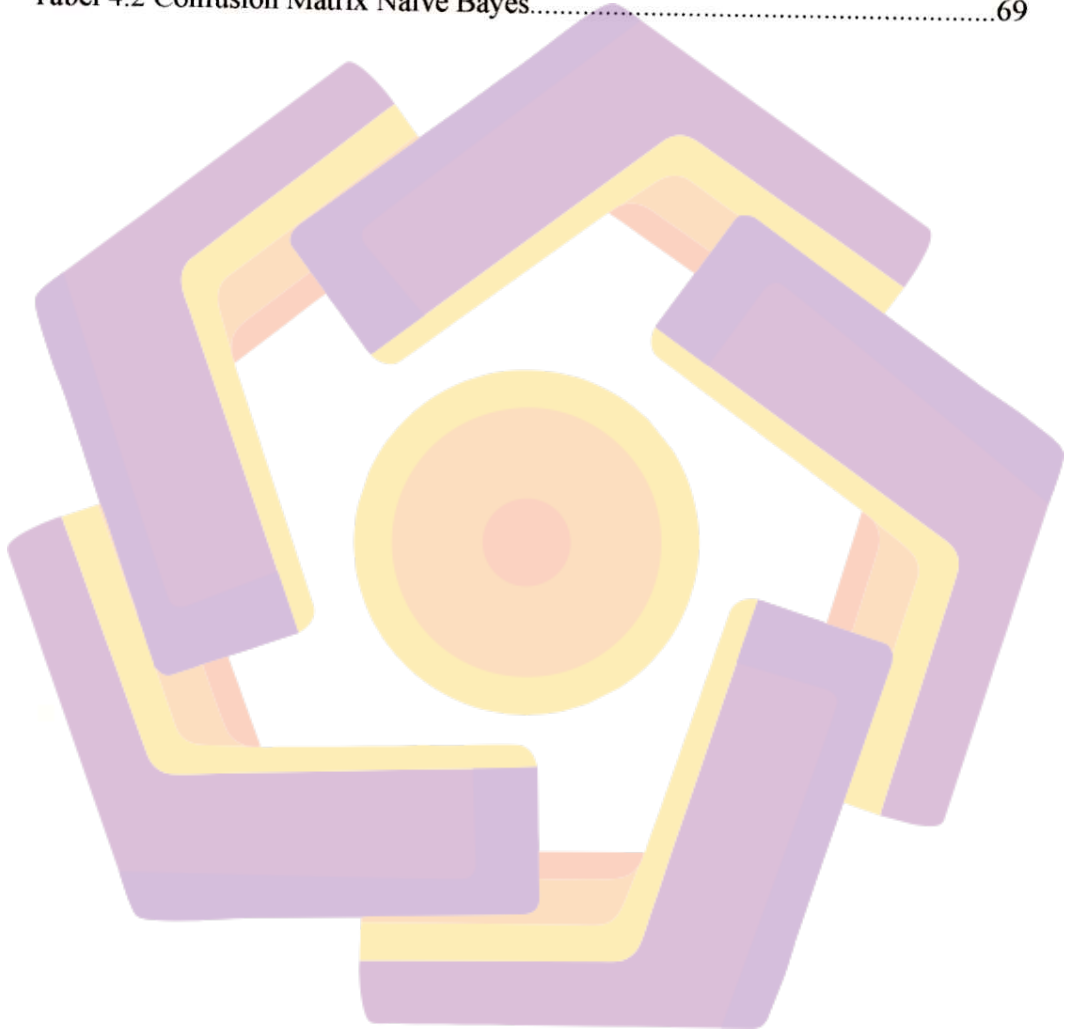
2.2	Dasar Teori.....	12
2.2.1	Pengertian Data Mining.....	12
2.2.2	Teorema Bayes.....	15
2.2.3	Naïve Bayes Classifier.....	16
2.2.4	Multinomial Naïve Bayes Classifier.....	16
2.2.5	Laplace Correction.....	22
2.2.6	One Hot Encoding.....	23
2.2.7	KFold Cross Validation.....	25
2.2.8	Confusion Matrix.....	26
BAB III ANALISIS DAN PERANCANGAN.....		28
3.1	Analisis Masalah.....	28
3.2	Solusi Yang Dipilih.....	29
3.3	Analisis Kebutuhan.....	30
3.3.1	Analisis Kebutuhan Fungsional.....	30
3.3.2	Analisis Kebutuhan Non Fungsional.....	30
3.4	Data Mining dengan Klasifikasi Naïve Bayes.....	31
3.4.1	Data Cleaning.....	31
3.4.2	Data Selection.....	31
3.4.3	Data Transformation.....	33
3.4.4	One Hot Encoding.....	35
3.4.5	Naïve Bayes.....	35
3.4.6	K-Fold Cross Validation.....	45
3.4.7	Confusion Matrix.....	46
3.5	Perancangan Aplikasi.....	47
3.5.1	Diagram Alir Preprocessing.....	47

3.5.2	Proses Klasifikasi Naïve Bayes	49
3.6	Perancangan File Training	50
3.7	Langkah-Langkah Pengembangan Aplikasi.....	51
BAB IV IMPLEMENTASI DAN PEMBAHASAN		55
4.1	Deskripsi Implementasi.....	55
4.2	Implementasi Perancangan Script dan Sistem	55
4.2.1	Implementasi Script Akses File Upload	55
4.2.2	Implementasi Script Query Data Training.....	56
4.2.3	Implementasi Data Selection dan Grouping	56
4.2.4	Implementasi Script Discretization.....	58
4.2.5	Implementasi Script Pembuatan Model	58
4.2.6	Implementasi Script Prediksi Data	59
4.2.7	Implementasi Script Menampilkan Hasil Prediksi	60
4.3	Pembuatan Interface.....	60
4.4	Evaluasi Program.....	64
4.5	Evaluasi Data Target	70
BAB V KESIMPULAN.....		71
5.1	Kesimpulan.....	71
5.2	Saran.....	71
DAFTAR PUSTAKA		72

DAFTAR TABEL

Tabel 2.1 Contoh Golf Dataset	17
Tabel 2.2 Likelihood Outlook.....	18
Tabel 2.3 Likelihood Temperature	19
Tabel 2.4 Likelihood Humidity.....	19
Tabel 2.5 Likelihood Wind.....	20
Tabel 2.6 Probabilitas Kelas Play	20
Tabel 2.7 Contoh Data Prediksi	21
Tabel 2.8 Data Sebelum One Hot Encoding.....	23
Tabel 2.9 Data Sesudah One Hot Encoding.....	24
Tabel 2.10 Confusion Matrix.....	26
Tabel 3.1 Aturan Konversi Nilai.....	34
Tabel 3.2 Sampel Dataset	36
Tabel 3.3 Likelihood Tipe Sekolah.....	39
Tabel 3.4 Likelihood Jenis Kelamin	39
Tabel 3.5 Likelihood Kota Sekolah	40
Tabel 3.6 Likelihood IP Semester 1	40
Tabel 3.7 Likelihood IP Semester 2	41
Tabel 3.8 Likelihood IP Semester 3	42
Tabel 3.9 Likelihood IP Semester 4.....	42
Tabel 3.10 Likelihood IPK	43
Tabel 3.11 Probabilitas Kelas Keterangan Lulus.....	43

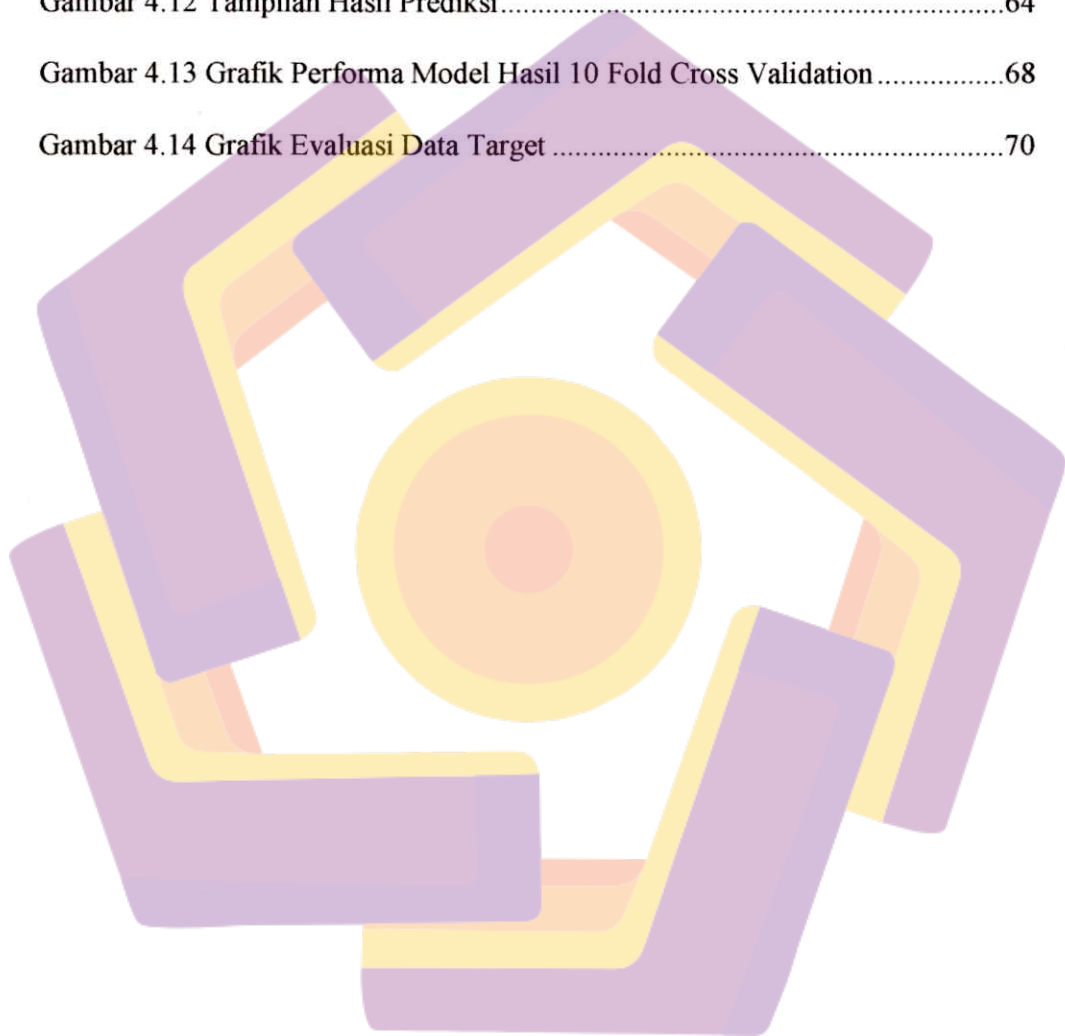
Tabel 3.12 Data Sampel Pertama.....	44
Tabel 3.13 Fold Data Sampel.....	46
Tabel 3.14 Confusion Matrix Data Sampel	47
Tabel 4.1 Hasil 10 Fold Cross Validation.....	65
Tabel 4.2 Confusion Matrix Naïve Bayes.....	69



DAFTAR GAMBAR

Gambar 2.1 Tahapan Data Mining.....	13
Gambar 2.2 Metode Data Mining.....	15
Gambar 2.3 Skema 10 Fold Cross Validation	25
Gambar 3.1 Data Alumni Sebelum Data Selection	32
Gambar 3.2 Data Alumni Sesudah Data Selection.....	32
Gambar 3.3 Data Alumni Sesudah Grouping	33
Gambar 3.4 Data Alumni Hasil Discretization.....	34
Gambar 3.5 Data Alumni Hasil One Hot Encoding	35
Gambar 3.6 Diagram Alir Preprocessing.....	48
Gambar 3.7 Proses Klasifikasi Naïve Bayes.....	50
Gambar 3.8 Bentuk File Training.....	51
Gambar 3.9 Langkah-Langkah Klasifikasi.....	51
Gambar 3.10 Langkah-Langkah Training Data.....	52
Gambar 3.11 Langkah-Langkah Testing Data.....	54
Gambar 4.1 Script Akses File Upload	55
Gambar 4.2 Script Query Data Training.....	56
Gambar 4.3 Script Data Selection.....	57
Gambar 4.4 Script Grouping.....	57
Gambar 4.5 Script Data Discretization	58
Gambar 4.6 Script Pembuatan Model.....	59
Gambar 4.7 Script Prediksi Data	59

Gambar 4.8 Script Menampilkan Hasil Prediksi	60
Gambar 4.9 Tampilan Awal pada Sistem	61
Gambar 4.10 Tampilan Model Building	62
Gambar 4.11 Tampilan Halaman Prediksi CSV	63
Gambar 4.12 Tampilan Hasil Prediksi	64
Gambar 4.13 Grafik Performa Model Hasil 10 Fold Cross Validation	68
Gambar 4.14 Grafik Evaluasi Data Target	70



INTISARI

Universitas AMIKOM Yogyakarta merupakan salah satu perguruan tinggi yang memiliki sekitar 1000 mahasiswa baru setiap tahunnya, khususnya pada prodi Informatika. Namun dari jumlah mahasiswa yang masuk tidak sebanding dengan jumlah mahasiswa yang lulus. Hal ini dapat dibuktikan dengan masih banyaknya mahasiswa yang belum dapat lulus tepat waktu. Jumlah mahasiswa yang lulus tepat waktu hanya mencapai kisaran 50% setiap tahunnya. Padahal ketepatan waktu lulus mahasiswa merupakan hal yang sangat penting bagi kriteria akreditasi.

Pada skripsi ini, peneliti mencoba untuk menggunakan teknik *data mining* dengan metode klasifikasi menggunakan 300 data mahasiswa dengan 150 data angkatan 2012, dan 150 data angkatan 2013 pada prodi Informatika yang telah lulus. Dari 300 data tersebut, 144 mahasiswa lulus dengan Tepat Waktu, dan 156 mahasiswa lulus Tidak Tepat Waktu. Data tersebut akan digunakan sebagai data *training*, dan data *testing* dalam proses pembuatan model. Algoritma yang digunakan untuk klasifikasi adalah Naïve Bayes. Variabel penelitian yang digunakan adalah jenis kelamin, tipe sekolah, sekolah kota, IP semester 1, IP semester 2, IP semester 3, IP semester 4, IPK, dan keterangan lulus. Model yang telah dibuat akan diuji menggunakan metode Confusion Matrix dengan K-Fold Cross Validation sebagai metode untuk membagi datanya.

Hasil pengujian yang diperoleh menggunakan metode K-Fold Cross Validation, dan Confusion Matrix pada model algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 68%, nilai presisi sebesar 61,3%, nilai recall sebesar 65,3%, dan nilai f1-score sebesar 61%. Untuk dapat mendapatkan model dengan kinerja yang lebih baik, peneliti menyarankan agar menggunakan data dengan jumlah yang lebih banyak, atau menggunakan variabel penelitian yang berbeda.

Kata Kunci: *data mining*, *Naïve Bayes*, klasifikasi, dan ketepatan waktu lulus

ABSTRACT

AMIKOM Yogyakarta University is one of the colleges that has around 1000 new students every year, especially in the Informatics study program. However, the number of students entering is not proportional to the number of students who graduate. This can be proven by the number of students who have not been able to graduate on time. The number of students who graduate on time only reaches around 50% every year. Even though the timeliness of graduating students is very important for accreditation criteria.

In this thesis, researchers tried to use data mining techniques with the classification method using 300 student data with 150 data for 2012, and 150 data for 2013 in Informatics study programs that had passed. Of the 300 data, 144 students graduated on time, and 156 students graduated not on time. The data will be used as data training, and data testing in the modeling process. The algorithm used for classification is Naïve Bayes. The research variables used were gender, type of school, city school, 1st semester GP, 2nd semester GP, 3rd semester GP, 4th semester P, GPA, and graduation information. The model that has been created will be tested using the Confusion Matrix method with K-Fold Cross Validation as a method to divide the data.

The test results obtained using the K-Fold Cross Validation method, and Confusion Matrix on the Naïve Bayes algorithm model produced an accuracy value of 68%, a precision value of 61.3%, a recall value of 65.3%, and a f1-score value of 61 %. To be able to get a model with better performance, researcher suggest using more amounts of data, or using different research variables.

Keywords: *data mining, Naïve Bayes, classification, and timeliness of graduation*