

**TESIS**

**PENERAPAN *DATA MINING* UNTUK PREDIKSI  
LAMA STUDI MAHASISWA DENGAN  
MENGUNAKAN METODE *NAÏVE BAYES* DAN  
*ADABOOST***



**Jacob Soares**

**No. Mhs: 145302261/PS/MTF**

**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA  
PROGRAM PASCA SARJANA  
UNIVERSITAS ATMA JAYA YOGYAKARTA**

**2017**

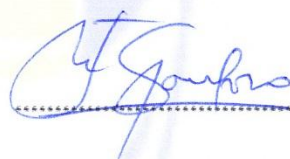



**UNIVERSITAS ATMA JAYA YOGYAKARTA**  
**PROGRAM PASCASARJANA**  
**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

---

**PENGESAHAN DOSEN PEMBIMBING**

Nama : JACOB SOARES  
Nomor Mahasiswa : 145302261/PS/MTF  
Konsentrasi : Soft Computing  
Judul Tesis : Penerapan *Data Mining* Untuk Prediksi Lama Studi Mahasiswa Dengan Menggunakan Algoritma *Naïve Bayes* Dan *Adaboost*

<b>Nama Pembimbing Pertama</b>	<b>Tanggal</b>	<b>Tanda Tangan</b>
Dr. Ir. Alb. Joko Santoso, M.T.	<u>11-09-2017</u>	

<b>Nama Pembimbing Kedua</b>	<b>Tanggal</b>	<b>Tanda Tangan</b>
Prof. Ir. Suyoto, M.Sc., Ph.D.	<u>11-09-2017</u>	





**UNIVERSITAS ATMA JAYA YOGYAKARTA**  
**PROGRAM PASCASARJANA**  
**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

---

**PENGESAHAN TESIS**

Nama : JACOB SOARES  
Nomor Mahasiswa : 145302261/PS/MTF  
Konsentrasi : Soft Computing  
Judul tesis : Penerapan *Data Mining* Untuk Prediksi Lama Studi Mahasiswa Dengan Menggunakan Algoritma *Naïve Bayes* Dan *Adaboost*

**Nama Pembimbing Pertama                      Tanggal                      Tanda Tangan**

Dr. Ir. Alb. Joko Santoso, M.T.

11-09-2017

Prof. Ir. Suyoto, M.Sc., Ph.D.

11-09-2017

Ir. A. Djoko Budiyanto, M.Eng., Ph.D

11-09-2017



**Ketua Program Study**

Prof. Ir. Suyoto, M.Sc., Ph.D.

## PERNYATAAN KEASLIAN KARYA

Saya yang bertanda tangan dibawah ini :

Nama : Jacob Soares

Nomor Mahasiswa : 145302261/PS/MTF

Konsentrasi : Soft Computing

Judul Tesis : Penerapan *Data Mining* Untuk Prediksi Lama Studi

Mahasiswa Dengan Menggunakan Algoritma *Naïve Bayes*  
Dan *Adaboost*

Menyatakan dengan sesungguhnya bahwa pada penulisan tesis ini berdasarkan hasil penelitian, pemikiran dan pemaparan asli dari peneliti sendiri, baik untuk naskah laporan maupun pada kegiatan programming yang tercantum sebagai bagian dari tesis ini. Jika terdapat karya orang lain maka, saya akan mencantumkan *citations* atau sumber yang jelas sesuai yang dikutip.

Demikian pernyataan ini saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh atau sanksi lain berdasarkan peraturan yang berlaku di Universitas Atma Jaya Yogyakarta Program Pascasarjana Program Studi Magister Teknik Informatika. Demikian pernyataan keaslian karya yang saya buat ini dengan keadaan yang sadar dan tanpa paksaan dari pihak manapun.

Yogyakarta, 11 September 2017

Yang membuat pernyataan

**Jacob Soares**

**Nim. 145302261**

## INTISARI

Teknik Prediksi merupakan hal yang patut untuk di perhitungkan dalam mewaspadaai setiap kondisi yang terjadi. Prediksi lama masa studi mahasiswa merupakan langkah penting dalam mempertimbangkan proses pengambilan keputusan sebagai peringatan dini (early warning) terhadap mahasiswa yang berpotensi menyalahi ketentuan lama studinya.

Dili Institute of Technology merupakan salah satu perguruan tinggi swasta di Timor-Leste yang memiliki jumlah mahasiswa yang sangat banyak yang masih terkendala dalam mengontrol lama studi setiap mahasiswanya. Tujuan penelitian ini adalah memanfaatkan teknik data mining dalam memprediksi lama studi mahasiswa pada perguruan tinggi Dili Institute of Technology.

Proses klasifikasi terhadap 334 record dataset mahasiswa menggunakan metode naïve bayes dengan nilai akurasi sebesar 97, 60%, dan tingkat eror klasifikasi sebesar 2, 40%. Sedangkan hasil gabungan *Naïve Bayes* dengan *Adaboost* nilai akurasi meningkat menjadi 99, 71% dan terjadi penurunan terhadap eror klasifikasi dengan nilai 0, 29%. Maka dapat disimpulkan bahwa metode *Adaboost* dapat maningkatkan nilai akurasi yang tinggi dan mengurangi tingkat error klasifikasinya. Dari hasil prediksi yang dilakukan oleh peneliti maka, pihak DIT dapat mengetahui jumlah mahasiswa tertentu yang berdasarkan hasil prediksinya dinyatakan berpotensi lulus dengan melampaui ketentuan lama studi yang ditentukan dengan mengambil tindakan preventif secara dini.

**Keywords:** *Data Mining*, *Prediksi*, *Naïve Bayes*, *Adaboost*.



## ABSTRACT

Prediction Technique is a thing worth considering to prepare every situation that will happen. Prediction of student's study period is an important step in considering decision making process as an early warning against students who potentially violate the provisions of their study duration.

Dili Institute of Technology is one of the private universities in Timor-Leste that has a large number of students who are still constrained in controlling the duration of their study. The purpose of this study is to utilize data mining techniques in predicting the duration of student study at Dili Institute of Technology.

The classification process of 334 records of students' dataset record uses Naïve Bayes method with an accuracy of 97, 60%, and error rate classification of 2, 40%. While the combined results of Naïve Bayes with Adaboost accuracy value increased to 99.71% and there was a decrease in classification error with a value of 0, 29%. Then it can be concluded that the Adaboost method can increase the value of high accuracy and reduce the level of classification error. From the prediction results made by the researcher then, the college can know the number of certain students who based on the predictions potentially graduate by exceeding the study duration which is determined by taking preventive action.

**Keywords:** Data Mining, Prediction, Naïve Bayes, Adaboost.

## KATA PENGANTAR

Segala Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmatNya sehingga penulis dapat menyelesaikan tesis dengan judul “Penerapan *Data Mining* Untuk Prediksi Lama Studi Mahasiswa Dengan Menggunakan metode *Naïve Bayes* Dan *Adaboost*” guna memenuhi sebagian persyaratan untuk memperoleh gelar Magister pada Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta. Penulis menyadari kelemahan serta keterbatasan yang ada sehingga dalam menyelesaikan tesis ini penulis memperoleh bantuan dari berbagai pihak, dalam kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Tuhan yang Maha Esa, yang selalu memberikan kesehatan, kekuatan, hikmat akal budi kepandaian sehingga penulis dapat menyelesaikan Tesis ini dengan baik.
2. Dili Institute of Technology (DIT). Yang memberikan kesempatan dan dukungan baik moril maupun materil bagi penulis untuk melanjutkan kuliah program pasca sarjana.
3. Keluarga kecilku tercinta istri dan anak-anakku, Zelia, Zejo dan Prilya yang selalu mendukung serta mendampingi penulis baik susah maupun senang dalam menghadapi berbagai situasi.
4. Ayahanda, ibunda serta kakak dan adik-adikku tercinta, yang selalu mendukung secara moril maupun materil di setiap langkahku.
5. Bapak Prof. Ir. Suyoto, M.Sc., Ph.D. selaku ketua jurusan serta pembimbing kedua yang senantiasa membantu penulis dalam menyelesaikan perkuliahan tepat pada waktunya.
6. Bapak Dr. Ir. Albertus Joko Santoso M.T. selaku pembimbing pertama yang telah meluangkan waktunya dalam membimbing penulis.
7. Bapak Ir. Dr. A. Djoko Budiyanto, M.Eng., Ph.D. selaku penguji yang memberihkan saran dalam melengkapi kekurangan yang dihadapi oleh penulis.

8. Seluruh Pengajar dan staff Dili Institute of Technology (DIT) yang selalu mendukung penulis dalam meraih gelar Magister Teknik Informatika.
9. Seluruh Pengajar dan staff administrasi Program Studi Magister Teknik Informatika Universitas Atmajaya Yogyakarta.
10. Teman-teman Magister Teknik Informatika angkatan september 2014 dan september 2015 Universitas Atma Jaya Yogyakarta Jaime, Apolinario, Komang, Vasio, Ebe dan semua yang tidak sempat saya sebutkan satu-satu. Terima kasih untuk bantuan yang telah diberikan dan kerbersamaan yang telah terjalin luar biasa selama ini.
11. Teman-Temanku Gil, Ana, Ongko, Edio Offy dan semua yang tidak sempat saya sebutkan satu-satu. Terima kasih untuk bantuan yang telah diberikan dan kerbersamaan yang telah terjalin luar biasa selama ini.

Penulis menyadari bahwa tesis ini masih banyak kekurangan baik isi maupun susunannya. Semoga tesis ini dapat bermanfaat tidak hanya bagi penulis juga bagi para pembaca.

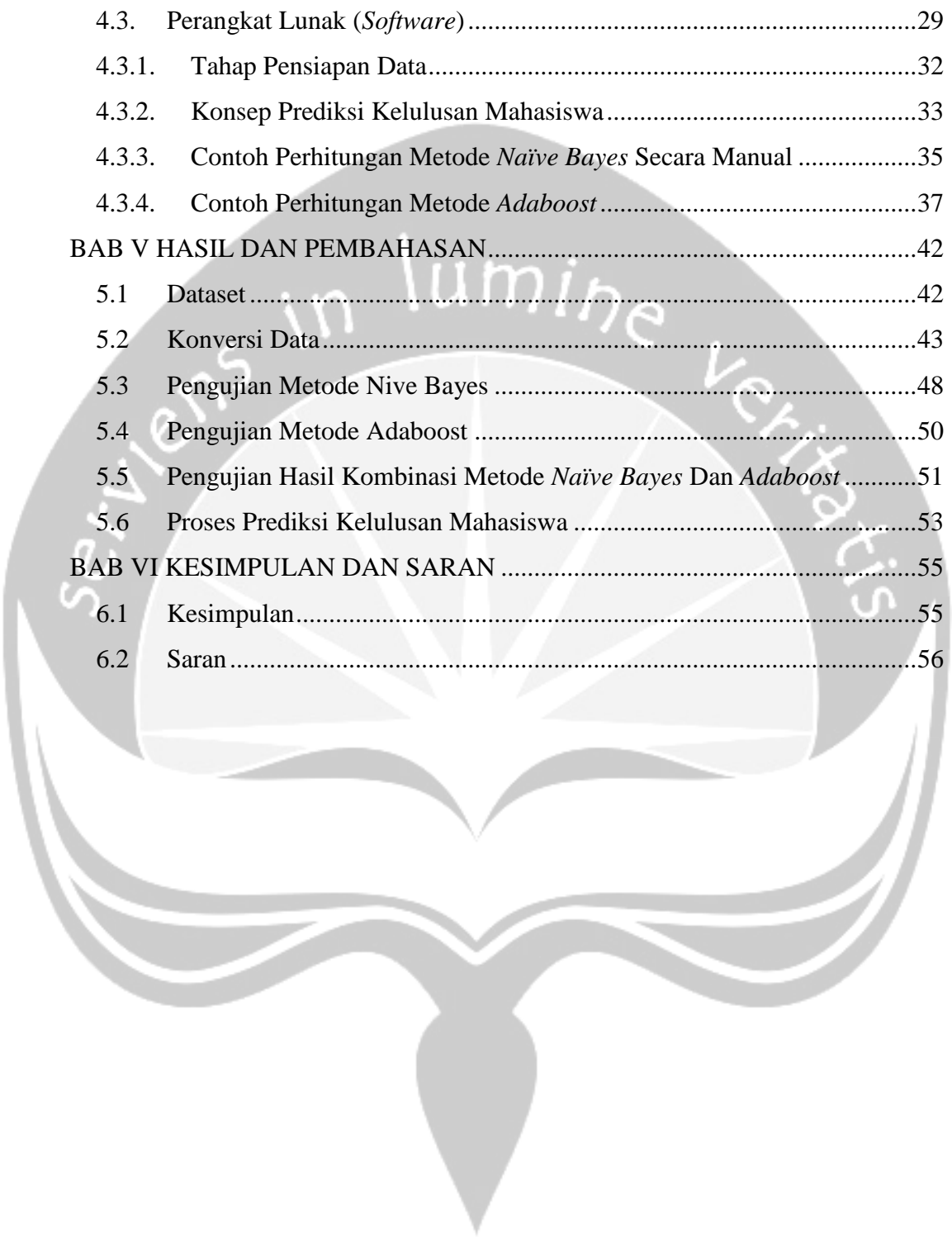
Yogyakarta, 12 September 2017

Penulis



## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN DOSEN PEMBIMBING .....	ii
HALAMAN PENGESAHAN TESIS .....	iii
PERNYATAAN KEASLIAN KARYA .....	iv
INTISARI.....	v
ABSTRACT.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu.....	7
2.1.1 Penelitian Yang Terkait Dengan Metode <i>Naïve Bayes</i> .....	7
2.1.2 Penelitian Yang Terkait Dengan Metode <i>Adaboost</i> .....	10
BAB III LANDASAN TEORI.....	14
3.1 Data Mining.....	14
3.1.1 Fungsi Data Mining .....	15
3.1.2 Arsitektur Data Mining.....	15
BAB IV METODOLOGI PENELITIAN .....	27
4.1. Metodologi Penelitian .....	27
4.2. Perangkat Keras ( <i>Hardware</i> ).....	28



4.3.	Perangkat Lunak ( <i>Software</i> ) .....	29
4.3.1.	Tahap Pensiapan Data.....	32
4.3.2.	Konsep Prediksi Kelulusan Mahasiswa.....	33
4.3.3.	Contoh Perhitungan Metode <i>Naïve Bayes</i> Secara Manual .....	35
4.3.4.	Contoh Perhitungan Metode <i>Adaboost</i> .....	37
BAB V HASIL DAN PEMBAHASAN.....		42
5.1	Dataset .....	42
5.2	Konversi Data.....	43
5.3	Pengujian Metode Nive Bayes .....	48
5.4	Pengujian Metode Adaboost .....	50
5.5	Pengujian Hasil Kombinasi Metode <i>Naïve Bayes</i> Dan <i>Adaboost</i> .....	51
5.6	Proses Prediksi Kelulusan Mahasiswa .....	53
BAB VI KESIMPULAN DAN SARAN .....		55
6.1	Kesimpulan.....	55
6.2	Saran .....	56

## DAFTAR TABEL

Tabel 2.1: Tabel perbandingan Penelitian.....	24
Table 3.1: <i>Confusion Matrix</i> .....	39
Table 4.1: Contoh Data Training Untuk Perhitungan Metode <i>Adaboost</i> .....	50
Tabel 4.2: Bobot Awal .....	51
Tabel 4.3: Tingkat error pada $T = 2, 5$ .....	51
Tabel 4.4: Tingkat error pada $T = 5, 5$ .....	51
Tabel 4.5: Tingkat error pada $T = 8, 5$ .....	52
Tabel 4.14. Update bobot pre Norm klasifikasi pertama .....	52
Tabel 4.6: Update bobot baru untuk $k = 2$ .....	53
Tabel 4.7: Update bobot klasifikasi kedua.....	54
Tabel 4.8: Update bobot klasifikasi ketiga.....	54
Tabel 5.1: Data Alumni angkatan 2010-2014.....	56
Tabel 5.2: Pendapatan orang tua .....	58
Tabel 5.3: Asal Sekolah .....	58
Tabel 5.4: Nilai Tes Seleksi atau tes masuk.....	59
Tabel 5.5: Konversi IPK .....	60
Table 5.6: Data yang siap diming .....	60
Table 5.7: Confusion Matrix model metode <i>Nive Bayes</i> .....	63
Table 5.8: Confusion Matrix Implementasi Algoritma <i>Adaboost</i> .....	65
Table 5.9: Confusion Matrix Hasil Kombinasi Metode NB Dan <i>Adaboost</i> ....	67

## DAFTAR GAMBAR

Gambar 3.1: Arsitektur sistem <i>Data Mining</i> .....	29
Gambar 3.2: Data mining sebagai tahapan dalam proses KDD.....	31
Gambar 4.1: Tahapan Penelitian.....	45
Gambar 4.2: Konsep prediksi kelulusan mahasiswa.....	46
Gambar 5.1: Klasifikasi data kelulusan menggunakan metode <i>Naïve Bayes</i> ..	62
Gambar 5.2: Pengujian Metode <i>Adaboost</i> .....	64
Gambar 5.3: Kombinasi metode <i>Naïve Bayes</i> dan <i>Adaboost</i> .....	66
Gambar 5.4: Hasil prediksi kelulusan.....	68