

**IMPLEMENTASI DATA MINING UNTUK  
MEMPREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU  
PADA STMIK BINA NUSANTARA JAYA LUBUKLINGGAU**



**TESIS**

**ENDANG ETRİYANTI  
ENTERPRISE IT INFRASTRUCTURE  
172420046**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2  
PROGRAM PASCASARJANA  
UNIVERSITAS BINA DARMA  
PALEMBANG  
2019**

**IMPLEMENTASI DATA MINING UNTUK  
MEMPREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU  
PADA STMIK BINA NUSANTARA JAYA LUBUKLINGGAU**

**Tesis ini diajukan sebagai salah satu syarat  
untuk memperoleh gelar**

**MAGISTER KOMPUTER**



**ENDANG ETRİYANTI  
ENTERPRISE IT INFRASTRUCTURE  
172420046**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2  
PROGRAM PASCASARJANA  
UNIVERSITAS BINA DARMA  
PALEMBANG  
2019**

**Halaman Pengesahan Pembimbing Tesis**

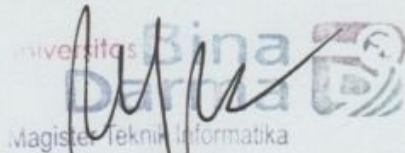
Judul Tesis :       IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI  
                          KELULUSAN MAHASISWA TEPAT WAKTU PADA  
                          STMIK BINA NUSANTARA JAYA LUBUKLINGGAU

Oleh ENDANG ETRIYANTI NIM 172420046 Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 Kosentrasi Enterprise IT Infrastructure, Program Pascasarjana Universitas Bina Darma pada 05 September 2019 dan telah dinyatakan LULUS.

Palembang, 05 September 2019

Mengetahui,

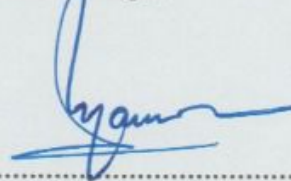
Ketua Program Studi

  
Universitas Bina Darma  
Magister Teknik Informatika

.....  
**Darius Antoni, S.Kom., M.M., Ph.D.**

Tim Pembimbing

Pembimbing I,



.....  
**Dedy Syamsuar, Ph.D.**

Pembimbing II,



.....  
**Yesi Novaria Kunang., S.T., M.Kom.**

**Halaman Pengesahan Penguji Tesis**

Judul Tesis :       IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI  
                          KELULUSAN MAHASISWA TEPAT WAKTU PADA  
                          STMIK BINA NUSANTARA JAYA LUBUKLINGGAU

Oleh ENDANG ETRIYANTI NIM 172420046 Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 Kosentrasi Enterprise Software Development, Program Pascasarjana Universitas Bina Darma pada 05 September 2019 dan dinyatakan LULUS.

Palembang, 05 September 2019

Tim Penguji:

Mengetahui,

Program Pasca Sarjana

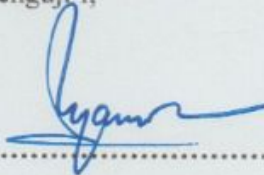
Universitas Bina Darma

Direktur,



.....  
**Dr. Ir. Hj. Hasmawaty AR, M.M., M.T.**

Penguji I,



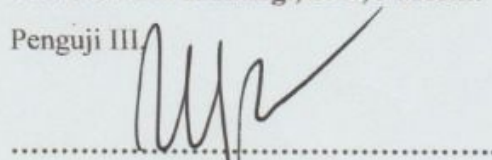
.....  
**Dedy Syamsuar, Ph.D.**

Penguji II,



.....  
**Yesi Novaria Kunang, S.T., M.Kom.**

Penguji III,



.....  
**Darius Antoni, S.Kom., M.M., Ph.D.**

Penguji IV,



.....  
**Tri Basuki Kurniawan, S.T., MEng., Ph.D.**

## ABSTRAK

Lama studi mahasiswa adalah rentang waktu bagi mahasiswa untuk menyelesaikan studinya. Lama studi mencerminkan tingkat pencapaian mahasiswa dalam studinya. Kemudian, rata-rata lama studi akan berpengaruh pada kualitas program studi karena diadopsi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) sebagai salah satu kriteria untuk menentukan tingkat akreditasi. Untuk alasan ini setiap lembaga pendidikan perlu memberikan perhatian serius terhadap masalah lama studi mahasiswa. Tujuan dari penelitian ini adalah untuk memprediksi status kelulusan mahasiswa tepat waktu. Ketidakmampuan mahasiswa untuk menyelesaikan studi tepat waktu dialami oleh sebagian besar Lembaga Pendidikan Tinggi. STMIK Bina Nusantara Jaya Lubuklinggau adalah salah satu perguruan tinggi yang mengalami hal tersebut. Dalam banyak kasus para mahasiswa menyelesaikan studi mereka lebih lama dari rentang waktu yang diharapkan. Akibatnya hal tersebut dapat menghambat mahasiswa baru untuk bergabung dengan lembaga karena kapasitas mahasiswa yang terbatas. Penelitian ini menggunakan teknik *data mining* dalam memprediksi status kelulusan mahasiswa tepat waktu. Pertama, *preprocessing* digunakan untuk mendapatkan dataset yang berkualitas. Kedua, data diproses untuk mendapatkan serangkaian prediksi. Pada langkah ini, dua algoritma *data mining* diterapkan - *Naive Bayes Classifier* dan Algoritma C4.5. Ketiga, hasilnya kemudian divalidasi menggunakan teknik *K-Fold Cross Validation*. Terakhir, *Confusion Matrix* digunakan untuk memvalidasi nilai akurasi hasil prediksi. Hasil penelitian menunjukkan bahwa metode *Naive Bayes Classifier* dapat digunakan untuk memprediksi status kelulusan siswa dengan tingkat akurasi 74,60% sedangkan tingkat akurasi metode Algoritma C4.5 hanya 73,93%. Metode *Naive Bayes Classifier* memiliki nilai akurasi yang lebih besar dibandingkan dengan nilai akurasi metode algoritma C4.5. Maka, metode *Naive Bayes Classifier* direkomendasikan untuk digunakan dalam menyelesaikan masalah prediksi kelulusan mahasiswa pada STMIK Bina Nusantara Jaya Lubuklinggau.

**Kunci:** *Naive Bayes, Algoritma C4.5, Kelulusan Mahasiswa*

## **ABSTRACT**

*A length of study is a duration for a student to complete his/her study. The duration reflects the achievement of a student on his/her study. Then, the average of study duration will contribute to the quality of the study program since it is adopted by national accreditation institution as one of criteria to determine the level of accreditation. For this reason, every higher education institution needs to give serious attention to this matter. The goal of this study is to predict graduation status of students on time. The inability of students to complete their studies on time is faced by most of higher education institution. STMIK Bina Nusantara Jaya Lubuklinggau is one of those which is experienced with this matter. In most cases, the students could complete their studies longer than the expected duration. As a result, it could prevent new students to join the institution since the limited student capacity. This study deploys data mining technique in predicting graduation status of students on time. First, preprocessing is used to obtain a good dataset. Secondly, the data is processed to obtain a set of prediction. In this step, two mining algorithm were applied – Naive bayes classifier and C4.5 algorithm. Thirdly, the result then was validated using K-Fold Cross Validation technique. Finally, Confusion Matrix is deployed to ensure the accuracy of the prediction. The results indicates that the Naive Bayes Classifier method can be used to predict student graduation status with an accuracy rate of 74,60% while the accuracy rate of the C4.5 Algorithm method is only 73.93%. The Naive Bayes Classifier method has a greater accuracy value than the accuracy value of the C4.5 algorithm method. Then, it is recommended to be used in solving the problem of prediction of student graduation at STMIK Bina Nusantara Jaya Lubuklinggau.*

*Key: Naive Bayes, C4.5 Algorithm, Student Graduation*

## MOTTO DAN PERSEMBAHAN

### MOTTO

- ❖ Sesuatu Yang Mudah Akan Menjadi Sulit Untuk Diselesaikan Jika Anda Tidak Mulai Mengerjakannya (Penulis)
- ❖ Raihlah Sesuatu Yang Anda Inginkan, Buktikan Jika Anda Bisa (Penulis)

Semua ini kuperssembahkan untuk

Suamiku tercinta **Bambang Sukoco, S.P.**  
Yang Telah Memberikan Dukungan, Do'a dan Motivasi

Anakku tersayang **Akmal Lathif El Azzam**  
Yang Selalu Memberikan Semangat Untuk Meraih  
Sebuah Cita-cita Dan Impian

Kedua Orang Tua, **Ayah dan Ibu** yang Selalu Memberikan Do'a  
dan Dukungan

**Keluarga Bzarku**, Yang Telah Memberikan Dukungan dan Motivasi

Terkhusus Untuk **Teman-teman Angkatan MTI 17**  
Terima Kasih atas kebersamaan yang telah kita lalui  
dengan penuh suka cita

# KATA PENGANTAR



Alhamdulillah, atas segala nikmat yang diberikan oleh Allah SWT yang selalu memberikan berkah, rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu Pada STMIK Bina Nusantara Jaya Lubuklinggau.

Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Komputer pada Universitas Bina Darma Palembang. Dalam penulisan tesis ini penulis telah berusaha semaksimal mungkin memberikan dan menyajikan yang terbaik. Tetapi penulis juga menyadari bahwa tesis ini masih jauh dari sempurna, hal ini dikarenakan terbatasnya pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun untuk kesempurnaan tesis ini.

Pada kesempatan ini, tidak lupa penulis mengucapkan terimah kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasihat, dan pemikiran dalam menyelesaikan tesis ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd. selaku Rektor Universitas Bina Darma Palembang.
2. Dr. Ir. Hj. Hasmawaty AR, M.M., M.T. Selaku Direktur Pascasarjana Universitas Bina Darma Palembang.



3. Darius Antoni, S.Kom., M.M., Ph.D. Selaku Ketua Program Studi Magister Teknik Informatika Universitas Bina Darma Palembang.
4. Dedy Syamsuar, Ph.D. Selaku pembimbing I yang telah memberikan bimbingan dan arahan dalam penulisan tesis ini.
5. Yesi Novaria Kunang, S.T., M.Kom. Selaku pembimbing II yang telah memberikan bimbingan dan arahan dalam penulisan tesis ini.
6. Pihak Sekretariat Pascasarjana Universitas Bina Darma Palembang yang telah memberikan bimbingan pelayanan dengan baik.

Palembang, 05 September 2019

Penulis

**Endang Etriyanti**

172420046

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN PEMBIMBING .....	ii
HALAMAN PENGESAHAN PENGUJI .....	iii
SURAT PERNYATAAN.....	iv
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
MOTTO DAN PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xiv
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Rumusan Permasalahan .....	4
1.5 Tujuan Penelitian .....	4
1.6 Manfaat Penelitian .....	4
1.7 Ruang Lingkup Penelitian.....	5
1.8 Susunan dan Struktur Tesis.....	5
<b>BAB II KAJIAN PUSTAKA</b> .....	<b>7</b>
2.1 Data Mining .....	7
2.1.1 Konsep Klasifikasi .....	9
2.1.2 Evaluasi Model Klasifikasi .....	14
2.1.3 RapidMiner .....	17

2.2 Prediksi.....	18
2.3 Kriteria Kelulusan Mahasiswa .....	18
2.4 Penelitian Terdahulu .....	23
2.5 Kerangka Pemikiran.....	30
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>32</b>
3.1 Desain Penelitian.....	33
3.2 Jadwal Penelitian.....	33
3.3 Metode Penelitian Yang Digunakan .....	34
3.4 Teknik Pengumpulan Data.....	34
3.5 Teknik Pengambilan Sampel.....	36
3.6 Teknik Analisa Data.....	36
3.6.1 Tahap Penelitian.....	36
3.6.2 Instrument Penelitian .....	37
<b>BAB IV ANALISA DATA.....</b>	<b>39</b>
4.1. Tahapan Analisa Data .....	39
4.2. Pengumpulan Data .....	40
4.2.1. Pengolahan Data Awal.....	42
4.2.2. Hasil Pengolahan Data .....	45
<b>BAB V HASIL DAN PEMBAHASAN .....</b>	<b>49</b>
5.1. Implementasi Dengan RapidMiner .....	49
5.1.1. Implementasi Model <i>Naive Bayes Classifier</i> .....	49
5.1.2. Implementasi Model Algoritma C4.5 .....	55
5.2. Perbandingan Hasil Implementasi.....	63
<b>BAB VI PENUTUP .....</b>	<b>65</b>
6.1. Kesimpulan .....	65
6.2. Saran.....	66
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR TABEL

Tabel	Halaman
Tabel 2.1.2.1. Ukuran Evaluasi Model Klasifikasi	<b>Error! Bookmark not defined.</b>
Tabel 2.1.3.1. Kriteria Kelulusan Mahasiswa .....	<b>Error! Bookmark not defined.</b>
Tabel 2.1.3.1 Penelitian Terdahulu .....	<b>Error! Bookmark not defined.</b>
Tabel 2.1.3.1. Jadwal Penelitian.....	<b>Error! Bookmark not defined.</b>
Tabel 3.6.2.1 Atribut Data Sebelum Pra Pemrosesan Data	<b>Error! Bookmark not defined.</b>
Tabel 4.2.2.1 Atribut Data Setelah Pra Pemrosesan Data	<b>Error! Bookmark not defined.</b>
Tabel 4.2.2.2. Rekapitulasi data mahasiswa lulus tepat waktu dan terlambat berdasarkan angkatan dan jenis kelamin.....	<b>Error! Bookmark not defined.</b>
Tabel 4.2.2.3. Rekapitulasi data rata-rata indeks prestasi (IP) mahasiswa pada semester 1,2,3 dan 4 berdasarkan angkatan .....	<b>Error! Bookmark not defined.</b>
Tabel 4.2.2.4. Rekapitulasi data rata-rata indeks prestasi kumulatif (IPK) mahasiswa pada semester 4 berdasarkan angkatan	<b>Error! Bookmark not defined.</b>
Tabel 5.1.1.1 Tuple Positif dan Tuple Negatif.....	<b>Error! Bookmark not defined.</b>
Tabel 5.1.2.1. Perbandingan Hasil Implementasi Metode Naive Bayes Classifier dan Algoritma C4.5 .....	<b>Error! Bookmark not defined.</b>

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1-1 Confusion Matrix .....	<b>Error! Bookmark not defined.</b>
Gambar 2.5-1. Kerangka Pemikiran.....	<b>Error! Bookmark not defined.</b>
Gambar 3.1-1 Desain Penelitian .....	<b>Error! Bookmark not defined.</b>
Gambar 4.2-1. Contoh Data Mentah Atribut IP-S1 s.d IP-S4 dan IPK-S4....	<b>Error! Bookmark not defined.</b>
Gambar 4.2-2. Proses Mapping Atribut IP-S1 s.d IP-S4 dan IPK-S4.....	<b>Error! Bookmark not defined.</b>
Gambar 4.2-3 Hasil Mapping Atribut IP-S1 s.d IP-S4 dan IPK-S4.....	<b>Error! Bookmark not defined.</b>
Gambar 5.1-1 Penetapan Label Kelas Tujuan.....	<b>Error! Bookmark not defined.</b>
Gambar 5.1-2. Pemilihan Operator Yang Digunakan	<b>Error! Bookmark not defined.</b>
Gambar 5.1-3. Proses Training dan Testing Metode Naive Bayes Classifier	<b>Error! Bookmark not defined.</b>
Gambar 5.1-4. Hasil Implementasi Metode Naive Bayes Classifier .....	<b>Error! Bookmark not defined.</b>
Gambar 5.1-5. Tingkat Akurasi Metode Naive Bayes Classifier	<b>Error! Bookmark not defined.</b>
Gambar 5.1-6. Penetapan Label Kelas Tujuan.....	<b>Error! Bookmark not defined.</b>

Gambar 5.1-7. Pemilihan Operator Yang Digunakan **Error! Bookmark not defined.**

Gambar 5.1-8. Proses Training dan Testing Metode Algoritma C4.5 ..... **Error! Bookmark not defined.**

Gambar 5.1-9. Hasil Implementasi Metode Algoritma C4.5 **Error! Bookmark not defined.**

Gambar 5.1-10. Tingkat Akurasi Metode Algoritma C4.5 **Error! Bookmark not defined.**

Gambar 5.1-11. Pohon Keputusan Berdasarkan Informatif Gain **Error! Bookmark not defined.**

## DAFTAR LAMPIRAN

1. SK. Pembimbing
2. Surat Izin Penelitian
3. Daftar Wawancara
4. Jurnal
5. Sertifikat Seminar
6. Lembar Perbaikan Tesis
7. Data Set Mahasiswa Tahun Angkatan 2013 dan 2014

