

**TESIS**

**PREDIKSI MASA STUDI MAHASISWA  
MENGUNAKAN KOMBINASI ALGORITMA  
*BAYESIAN NETWORK* DAN *K-NEAREST NEIGHBORS***



**Mariana Windarti**

**No.Mhs : 135302105/PS/MTF**

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

**2016**



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

---

**PENGESAHAN TESIS**

Nama : MARIANA WINDARTI  
Nomor Mahasiswa : 135302105 / PS / MTF  
Konsentrasi : *Soft Computing*  
Judul Tesis : PREDIKSI MASA STUDI MAHASISWA  
MENGUNAKAN KOMBINASI  
ALGORITMA BAYESIAN NETWORK  
DAN K-NEAREST NEIGHBORS

**Nama Pembimbing**

**Tanggal Tanda tangan**

Dr. Alb. Joko Santoso, M.T.

22-3-2016  
.....

Dra. Ernawati, M.T.

22-3-2016  
.....

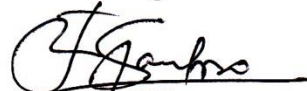




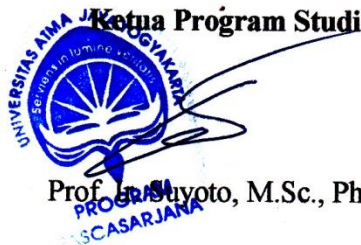
**UNIVERSITAS ATMA JAYA YOGYAKARTA**  
**PROGRAM PASCASARJANA**  
**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

---

**PENGESAHAN TESIS**

Nama : MARIANA WINDARTI  
Nomor Mahasiswa : 135302105/PS/MTF  
Konsentrasi : Soft Computing  
Judul Tesis : Prediksi Masa Studi Mahasiswa Menggunakan  
Kombinasi Algoritma *Bayesian Network* dan *K-Nearest Neighbors*

<b>Nama Penguji</b>	<b>Tanggal</b>	<b>Tanda tangan</b>
Dr. Ir. Alb. Joko Santoso, M.T. (Ketua/Penguji/Pembimbing Utama)	2 Mei 2016 .....	 .....
Dra. Ernawati, M.T. (Anggota/Penguji)	29/4/16 .....	 .....
Prof. Ir. Suyoto, M.Sc., Ph.D. (Anggota/Penguji)	29/4/16 .....	 .....

  
Ketua Program Studi  
Prof. Ir. Suyoto, M.Sc., Ph.D.  
PROGRAM PASCASARJANA

## HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Mariana Windarti

NIM : 135302105/PS/MTF

Konsentrasi : Soft Computing

Judul Tesis : Prediksi Masa Studi Mahasiswa Menggunakan Kombinasi  
*Algoritma Bayesian Network dan K-Nearest Neighbors*

Dengan ini menyatakan bahwa tesis ini

dengan sesungguhnya menyatakan bahwa tesis ini adalah benar-benar karya tulis saya dan bukan merupakan karya orang lain atau hasil bajakan dari karya tulis orang lain. Semua sumber atau informasi di dalam tesis ini yang berasal dari penulis lain baik yang dikutip dan dirujuk telah saya nyatakan dengan benar dengan menyebutkan sumber asli dan dituliskan didalam daftar pustaka.

Yogyakarta, April 2016

Mariana Windarti

## INTISARI

Salah satu faktor yang mempengaruhi kualitas sebuah perguruan tinggi adalah kinerja mahasiswa yang dapat diukur melalui lamanya masa studi. Semakin cepat masa studi mahasiswa maka kinerja mahasiswa semakin baik, begitu juga sebaliknya. Badan Akreditasi Nasional Perguruan Tinggi (BAN PT) bertugas untuk menilai kualitas perguruan tinggi. Dengan penilaian tersebut perguruan tinggi akan lebih meningkatkan mutu dan kualitas guna mencapai suatu visi, misi dan tujuan perguruan tinggi. Perguruan tinggi mempunyai peran penting dalam mengelola dan mengevaluasi kinerja mahasiswa khususnya kelulusan mahasiswa.

Data pada perguruan tinggi belum dimanfaatkan untuk memahami kondisi sebuah perguruan tinggi dalam rangka perbaikan mutu & kualitas. Proses untuk mendapatkan informasi yang berguna dari sejumlah data yang besar sehingga didapat suatu pengetahuan disebut *data mining* (penambangan data). Pada penelitian ini akan dibuat sebuah aplikasi untuk memprediksi masa studi mahasiswa Universitas Widya Dharma (UNWIDHA) Klaten berbasis desktop dengan mengimplementasikan kombinasi algoritma *Bayesian Network* (BN) dan *K-Nearest Neighbors* (KNN).

Penelitian ini bertujuan menggali informasi yang bisa digunakan dari data pelatihan alumni mahasiswa UNWIDHA Klaten untuk memprediksi masa studi mahasiswa berdasarkan IPK 2 semester pertama, nilai ujian nasional (UN), penjurusan sekolah, lulusan sekolah, jalur masuk perguruan tinggi dan hasil tes masuk. Hasil prediksi dikelompokkan kedalam enam kelas yaitu masa studi < 4 tahun, masa studi  $\geq 4$  tahun dan masa studi < 4.5 tahun, masa studi  $\geq 4.5$  tahun dan masa studi < 5 tahun, masa studi  $\geq 5$  tahun dan masa studi < 5.5 tahun, masa studi  $\geq 5.5$  tahun dan masa studi < 6 tahun dan terakhir masa studi  $\geq 6$  tahun. Selain itu, penelitian ini juga untuk menganalisa dan membandingkan kinerja algoritma kombinasi dengan algoritma BN, dan KNN. Hasil dari sistem menunjukkan bahwa algoritma kombinasi memiliki akurasi yang lebih unggul dari pada algoritma BN, dan KNN yaitu sebesar 70 % untuk program studi eksak dan 65 % untuk program studi non eksak. Sedangkan untuk waktu komputasi algoritma kombinasi lebih cepat dari pada algoritma KNN tetapi sedikit lebih lama dari algoritma BN.

**Kata kunci :** *data mining, bayesian network, k-nearest neighbors, masa studi.*

## ABSTRACT

One of the factors that affect the quality of a university is the student's performance can be measured by the length of the study period. The faster the study period, the students of the student's performance is getting better, and vice versa. National Accreditation Board for Higher Education (BAN PT) in charge of assessing the quality of higher education. With these assessments college will further improve the quality and quality in order to achieve a vision, mission and goals of the college. Higher education has an important role in managing and evaluating the performance of students, especially graduate students.

Data on college untapped to understand the conditions of a university in order to improve the quality. The process to obtain useful information from large amounts of data in order to get a knowledge of so-called data mining (data mining). In this study will be made an application to predict future studies students at the University of Widya Dharma (UNWIDHA) Klaten based desktop by implementing a combination of algorithms Bayesian Network (BN) and K-Nearest Neighbors (KNN).

This study aims to explore the information that can be used from the training data of alumni students UNWIDHA Klaten to predict the future of a student's study is based on GPA 2 first half, the national exam (UN), the school majors, graduate school, college entrance lane and the results of the entrance test. The prediction results are grouped into six classes, namely the period of study <4 years of the study period> = 4 years and a study of <4.5 years, the study period> = 4.5 years and a period of study <5 years of the study period> = 5 years and a period of study <5.5 year study period> = 5.5 years and a period of study <6 years and last period of study> = 6 years. In addition, this study is also to analyze and compare the performance of the algorithm algorithm combination with BN, and KNN. The results of the system show that the algorithm has a combination of superior accuracy of the algorithm BN, and KNN which is 70% for inexact course and 65% for non-exact courses. As for the combination algorithm computation time faster than the KNN algorithm, but slightly longer than the BN algorithm.

Keywords: data mining, Bayesian networks, k-nearest neighbors, the study period.

## KATA PENGANTAR

Puji Syukur dan kemuliaan hanya bagi Tuhan Yesus Kristus karena atas kasih karunia dan anugerahNya penulis dapat menyelesaikan pembuatan tesis dengan judul ‘Prediksi Masa Studi Mahasiswa Menggunakan Kombinasi Algoritma *Bayesian Network* dan *K-Nearest Neighbors*’. Dalam menyelesaikan tesis ini, penulis sebagai mahasiswa Program Studi Magister Teknik Informatika, mendapatkan banyak pengalaman dan pengetahuan baru. Penulis juga bisa mengembangkan ilmu dan pengetahuan yang didapatkan dari perkuliahan sehingga bisa digunakan dalam menyelesaikan tesis. Pengalaman tersebut bermanfaat dalam menghadapi dunia kerja setelah selesai menempuh perkuliahan di Universitas Atma Jaya Yogyakarta yang tercinta ini.

Proses pelaksanaan tesis yang telah dilakukan tidak terlepas dari bantuan dari berbagai pihak yang sangat membantu keberhasilan penulis selaku pelaksana. Oleh sebab itu dalam kesempatan ini penulis menyampaikan terima kasih kepada semua pihak yang telah membantu penulis baik itu dalam menyelesaikan laporan ini dan juga dalam pelaksanaan tesis:

1. Bapak Prof. Ir. Suyoto, M.Sc., Ph.D., selaku Ketua Program Studi Magister Teknik Informatika sekaligus selaku dosen penguji yang telah menguji penulis dalam sidang tesis dan memberikan masukan yang berguna bagi penulis.
2. Bapak Dr. Ir. Alb. Joko Santoso, M.T., selaku dosen pembimbing I yang dengan sabar telah membimbing dan memberikan masukan yang berguna bagi penulis selama pelaksanaan dan penyusunan laporan tesis.

3. Ibu Dra. Ernawati, M.T., selaku dosen pembimbing II yang membimbing penulis, memberikan waktu dan kesempatan penulis untuk belajar hal-hal baru selama pelaksanaan tesis.
4. Bapak dan ibu dosen Program Studi Magister Teknik Informatika, Program Pascasarjana, Universitas Atma Jaya Yogyakarta. Terima kasih atas segala bimbingan selama ini, sehingga penulis dapat menempuh kuliah di Program Studi Magister Teknik Informatika.
5. Orang tuaku tercinta Bapak Agus Suranto dan Mama Sri Sudarti yang sudah berada di Surga. Terima kasih buat kasih sayang kalian berdua, doa, dukungan dan motivasi selama ini dalam melaksanakan tesis
6. Suamiku tercinta Hendrik Setiawan, yang dengan sabar menemani penulis selama menyelesaikan tesis ini. Terimakasih buat doa, cinta, waktu dan dukungan selama ini.
7. Adikku tersayang Elisabet Marlina dan Alexander Triatmaja. Terimakasih buat dukungan dan doanya selama ini.
8. Keluarga di Pekanbaru, mertuaku Bapak Fx. Sapto Riyadi dan Ibu Susiati. Terimakasih buat kasih sayang, semangat dan doa yang diberikan. Buat adikku Hengky, Lia, Galeh dan ponakanku tersayang Arkano, Terimakasih buat doa dan dukungan selama ini.
9. Teman-teman Magister Teknik Informatika angkatan Januari 2014 khususnya Ma'am Imma, Putri, Yayak dan Lisa. Terima kasih atas bantuan, dukungan, dan kerjasamanya selama ini.



10. Teman-temanku Erna dan Diana. Terimakasih sudah menjadi teman terbaik di akhir-akhir penyusunan tesis ini. Terimakasih buat dukungan dan semangatnya.
11. Semua pihak UNWIDHA yang tidak dapat penulis sebutkan satu-persatu, terimakasih karena telah membantu penulis dalam melakukan penelitian ini.
12. Semua teman-teman yang tidak dapat penulis sebutkan satu per satu yang telah membantu selama ini, terima kasih atas segala bantuan dan dukungan.

Kiranya penulisan laporan tesis ini masih kurang sempurna, oleh karena itu segala kritik dan saran yang membangun sangat penulis harapkan. Semoga apa yang ada di dalam laporan ini bermanfaat untuk pembaca.

Yogyakarta, April 2016

Penulis

## HALAMAN PERSEMBAHAN

Segala Pujian dan Hormat, Kemuliaan dan Kuasa hanya bagi Tuhan Yesus Kristus.

Allah Turut Bekerja dalam Segala Sesuatu untuk  
Mendatangkan Kebaikan Bagi Mereka Yang  
Mengasihi DIA.  
(Roma 8:28)

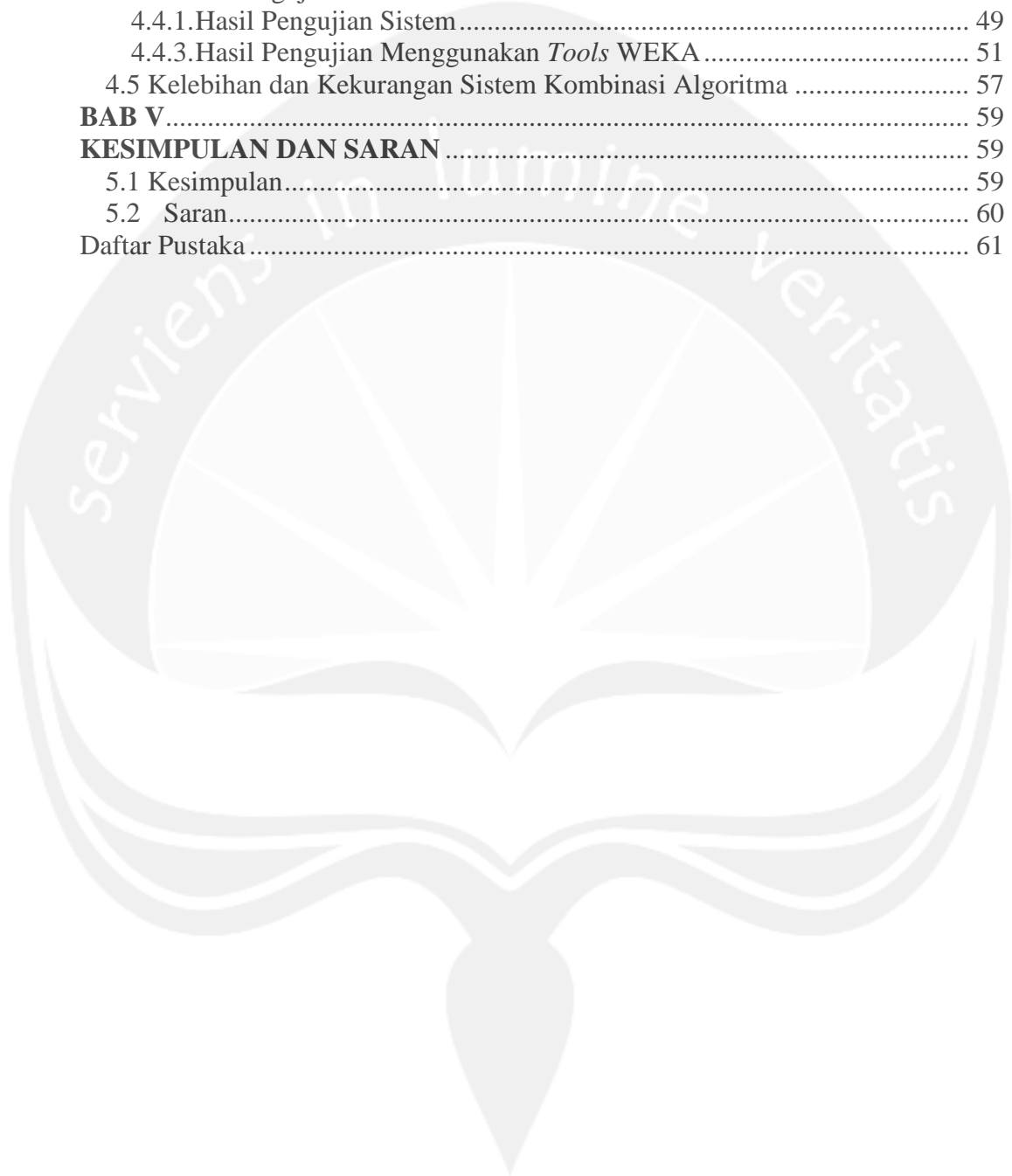
Dan Allah sanggup melimpahkan segala kasih  
karunia kepada kamu, supaya kamu senantiasa  
berkecukupan di dalam segala sesuatu dan malah  
berkelebihan di dalam pelbagai kebajikan  
(2 Korintus 9:8)

Tesis ini terkhusus saya persembahkan buat Mama tercinta yang sudah berada disurga. Karena kasih, cinta, ketulusan, keteladanan dan semangat mama yang membuat saya bisa menyelesaikan penulisan tesis ini, juga buat Bapak yang selalu semangat dan pantang menyerah. Kemudian buat suami saya tercinta dan keluarga saya yang senantiasa mendukung saya dalam susah dan senang.

## DAFTAR ISI

<b>PENGESAHAN TESIS</b> .....	<b>Error! Bookmark not defined.</b>
<b>HALAMAN PERNYATAAN</b> .....	iii
<b>INTISARI</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>HALAMAN PERSEMBAHAN</b> .....	x
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xiii
<b>DAFTAR TABEL</b> .....	xiv
<b>BAB I</b> .....	1
<b>PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah.....	5
1.4 Keaslian Penelitian.....	6
1.5 Manfaat Penelitian .....	6
1.6 Tujuan Penelitian .....	7
1.7 Sistematika Penulisan .....	7
<b>BAB II</b> .....	9
<b>TINJAUAN PUSTAKA DAN LANDASAN TEORI</b> .....	9
2.1 Tinjauan Pustaka .....	9
2.2 Landasan Teori.....	17
2.2.1 <i>Data Mining</i> (Penambangan Data) .....	17
2.2.2 Klasifikasi .....	19
2.2.3 Pengukuran Kinerja Klasifikasi .....	20
2.2.4 <i>Bayesian Network (BN)</i> .....	21
2.2.5 <i>K-Nearest Neighbors (K-NN)</i> .....	23
<b>BAB III</b> .....	24
<b>METODOLOGI PENELITIAN</b> .....	24
3.1 Metode Pengumpulan Data .....	24
3.2 Metode Pengembangan Perangkat Lunak .....	24
3.3 Bahan Penelitian.....	25
3.4 Alat Penelitian.....	25
3.4.1.Kebutuhan Perangkat Keras .....	25
3.4.2 Kebutuhan Perangkat Lunak .....	26
3.5 Tahapan Penelitian .....	26
3.6 Kendala Penelitian .....	29
3.7 Algoritma Kombinasi <i>Bayesian Network</i> dan <i>K-Nearest Neighbors</i> .....	29
<b>BAB IV</b> .....	33
<b>HASIL PENELITIAN DAN PEMBAHASAN</b> .....	33
4.1 Analisis Sistem.....	33
4.1.1 Analisis Kebutuhan Pengguna ( <i>User</i> ).....	33
4.1.2 Analisis Data.....	34
4.1.3 <i>Data Flow Diagram (DFD)</i> .....	35

4.2 Perancangan Basis Data .....	40
4.3 Implementasi Sistem .....	42
4.3.1 Implementasi Menggunakan Kombinasi Algoritma.....	43
4.4 Hasil Pengujian .....	49
4.4.1. Hasil Pengujian Sistem .....	49
4.4.3. Hasil Pengujian Menggunakan <i>Tools</i> WEKA.....	51
4.5 Kelebihan dan Kekurangan Sistem Kombinasi Algoritma .....	57
<b>BAB V</b> .....	59
<b>KESIMPULAN DAN SARAN</b> .....	59
5.1 Kesimpulan.....	59
5.2 Saran.....	60
Daftar Pustaka .....	61



## DAFTAR GAMBAR

Gambar 2.1 Tahapan dalam Penambangan Data ( Han et al., 2012).....	18
Gambar 3.1 Tahapan Penelitian .....	28
Gambar 3.2 Flowchart Algoritma Kombinasi <i>Bayesian Network</i> dan KNN.....	32
Gambar 4.1 DFD Level 0 (Diagram Konteks).....	36
Gambar 4.2 DFD Level 1 .....	38
Gambar 4.3 DFD Level 2 Kelola <i>User</i> .....	39
Gambar 4.4 DFD Level 2 Kelola Mahasiswa .....	40
Gambar 4.5 Struktur <i>Bayesian Network</i> .....	43
Gambar 4.6 Tampilan Halaman <i>Import Data File .xls</i> .....	45
Gambar 4.7 Tampilan Halaman Prediksi Algoritma BN (Eksak).....	46
Gambar 4.8 Tampilan Halaman Prediksi Algoritma BN (Non Eksak).....	47
Gambar 4.9 Tampilan Halaman Prediksi Algoritma KNN (Eksak) .....	48
Gambar 4.10 Tampilan Halaman Prediksi Algoritma Kombinasi (Eksak).....	48
Gambar 4.11 Hasil Pengujian 40 Data Uji <i>Bayesian Network</i> pada WEKA.....	53
Gambar 4.12 Hasil Pengujian 40 Data Uji KNN pada WEKA.....	53
Gambar 4.13 Hasil Prediksi 40 Data Uji dengan BN.....	54
Gambar 4.14 Hasil Prediksi 40 Data Uji dengan KNN .....	55

## DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian.....	13
Tabel 2.2 Matriks konfusi untuk klasifikasi dua kelas (Prasetyo, 2012) .....	20
Tabel 4.1 Dataset Program Studi TI (Eksak) .....	34
Tabel 4.2 Dataset Program Studi Manajemen (Non Eksak) .....	35
Tabel 4.3 Struktur Tabel Tbl_User .....	41
Tabel 4.4 Struktur Tabel Tbl_Mahasiswa.....	41
Tabel 4.5 Hasil Pengujian Berdasarkan Akurasi (Eksak) .....	50
Tabel 4.6 Hasil Pengujian Berdasarkan Akurasi (Non Eksak) .....	50
Tabel 4.7 Hasil Pengujian Berdasarkan Waktu Komputasi (Eksak) .....	50
Tabel 4.8 Hasil Pengujian Berdasarkan Waktu Komputasi (Non Eksak).....	50
Tabel 4.9 Hasil Pengujian WEKA Prodi Eksak.....	52
Tabel 4.10 Hasil Pengujian WEKA Prodi Non Eksak.....	56
Tabel 4.11 Hasil Perbandingan Akurasi Atribut Numerik dan Kategori.....	57