

**ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN NAIVE
BAYES PADA LULUSAN TEPAT WAKTU MAHASISWA
DI UNIVERSITAS ISLAM NEGERI RADEN FATAH
PALEMBANG**



TESIS

**MUHAMMAD KAMIL
ENTERPRISE IT INFRASTRUCTURE
182420014**

**PROGRAM STUDI TEKNIK INFORMATIKA-S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2020**

**ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN NAIVE
BAYES PADA LULUSAN TEPAT WAKTU MAHASISWA
DI UNIVERSITAS ISLAM NEGERI RADEN FATAH
PALEMBANG**

Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar

MAGISTER TEKNIK INFORMATIKA



**MUHAMMAD KAMIL
ENTERPRISE IT INFRASTRUCTURE
182420014**

**PROGRAM STUDI TEKNIK INFORMATIKA-S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2020**

Halaman Pengesahan Pembimbing Tesis

Judul Tesis : ANALISIS PERBADINGAN ALGORITMA C4.5 DAN NAÏVE

BAYES PADA LULUSAN TEPAT WAKTU MAHASISWA DI

UNIVERSITAS ISLAM NEGERI RADEN FATAH PALEMBANG

Oleh Muhammad Kamil NIM 182420014 Tesis ini telah disetujui dan disahkan

oleh Penguji Program Studi Teknik Informatika – S2 Konsentrasi Enterprise IT

Infrastructure Pascasarjana Universitas Bina Darma pada 25 Februari 2020 dan

telah dinyatakan LULUS

**Mengetahui,
Program Studi Teknik Informatika – S2
Universitas Bina Darma
Ketua,**


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

Pembimbing

Pembimbing I,


.....
Dr. Widya Cholil, S.Kom., M.IT.

Halaman Pengesahan Penguji Tesis

Judul Tesis : ANALISIS PERBADINGAN ALGORITMA C4.5 DAN NAÏVE

BAYES PADA LULUSAN TEPAT WAKTU MAHASISWA DI

UNIVERSITAS ISLAM NEGERI RADEN FATAH PALEMBANG

Oleh Muhammad Kamil NIM 182420014 Tesis ini telah disetujui dan disahkan oleh Penguji Program Studi Teknik Informatika – S2 Konsentrasi Enterprise IT Infrastructure Pascasarjana Universitas Bina Darma pada 25 Februari 2020 dan telah dinyatakan LULUS

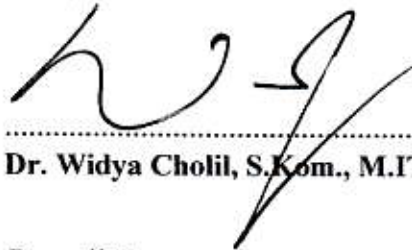
Palembang, 25 Februari 2020
Mengetahui,
Program Pascasarjana
Universitas Bina Darma
Direktur,



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

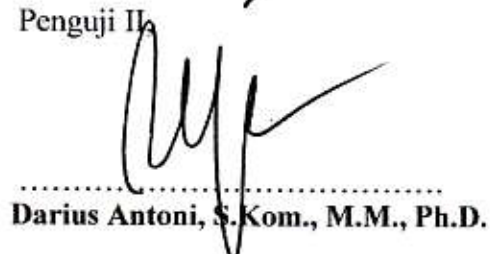
Tim Penguji:

Penguji I,



.....
Dr. Widya Cholil, S.Kom., M.IT.

Penguji II,



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

Penguji III,



.....
A.Haidar Mirza, S.T., M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini

Nama : Muhammad Kamil

NIM : 182420014

Dengan ini menyatakan bahwa :

1. Karya tulis saya (Tesis) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baru (Magister, Sarjana, dan ahli Madya) di Universitas Bina Darma.
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya sendiri dengan arahan tim pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukan ke dalam daftar pustaka.
4. Karena yakin dengan keaslian karya tulis ini, saya menyatakan bersedia tesis/tugas akhir yang saya hasilkan diunggah ke internet.
5. Surat pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagai mana mestinya.

Palembang, 25 Februari 2020

Yang membuat pernyataan,



Muhammad Kamil

NIM : 182420014

ABSTRAK

Universitas Islam Negeri Raden Fatah Palembang Merupakan suatu universitas negeri dipalembang dimana sudah banyak meluluskan mahasiswanya terutama mahasiswa program sarjananya (S1). Pada hal ini Penelitian akan melakukan analisis pada mahasiswa yang lulusan di UIN Raden Fatah Palembang terutama yang lulusan tepat waktu. Pada hal memprediksi kelulusan mahasiswa telah banyak dilakukan dengan memakai metode C4.5 dan metode Naive Bayes. Metode C4.5 merupakan bagian algoritma decision tree yang merubah data menjadi pohon keputusan, kemudian bisa disimpulkan menjadi rule-rule. Sedangkan naïve bayes merupakan metode yang menghitung probabilitas pada tingkat kemunculan data yang satu dengan data yang lainnya. Pada penelitian ini peneliti membandingkan dua metode yaitu metode C4.5 dan Metode Naïve Bayes untuk dipresentasikan dalam kelulusan Mahasiswa. Data yang diambil dari 2 dataset yang sudah diintegrasikan yaitu data mahasiswa dan data alumni tahun 2016 sampai 2019. Proses dilakukan dengan acuan tahapan *knowledge discovery form data* (KDD) dengan menggunakan aplikasi Rapid Miner. Dimana dari hasil pengujian mengukur kinerja kedua metode tersebut menggunakan aplikasi RapidMiner dengan Metode Cross Validation maka diketahui bahwa C4.5 mempunyai nilai akurasi yang lebih bagus yakni sebesar 69,54% sedangkan naïve bayes sebesar 68,38%. Kemudian hasil perbandingan kinerja kedua metode tersebut, metode dengan nilai akurasi terbaik akan bisa diprestasikan untuk pengujian prediksi kelulusan mahasiswa untuk tahun selanjutnya.

Kata Kunci : C4.5, Naïve Bayes, Aplikasi Rapid Miner, Cross Validation, Kelulusan Mahasiswa

ABSTRACT

Raden Fatah State Islamic University of Palembang It is a State University Dipalembang which has been a lot of students have graduated his students, especially the undergraduate program (S1). In this case, research will conduct an analysis on students who graduate in UIN Raden Fatah Palembang especially those who are graduates on time. In terms of predicting the graduation of students has been done much by using method C 4.5 and Naïve Bayes method. Method C 4.5 is a part of the decision tree algorithm that converts data into a decision tree, then it can be concluded into a rule-rule. While naïve Bayes is a method that calculates probability at the level of occurrence of data that one with the other data. In this research the researcher compares two methods namely the C 4.5 method and the Naïve Bayes method to be presented in student graduation. Data extracted from 2 datasets that have been integrated are data of students and alumni data from 2016 to 2019 years. Process is done by the stage Reference Knowledge Discovery form data (KDD) using Rapid Miner application. Where from the test results measuring the performance of both methods using RapidMiner application with Cross Validation method then known under C 4.5 have a better accuracy value of 69,54% while naïve Bayes of 68,38%. Then the results of the performance comparison of both methods, the method with the best accuracy value will be to be donated for the test of student graduation predictions for the next year

Keywords: *C 4.5, Naïve Bayes, Rapid Miner application, Cross Validation, student graduation*

MOTTO DAN PERSEMBAHAN

MOTTO

**“Hai orang-orang beriman, Jadikanlah sabar dan sholatmu
Sebagai penolongmu, sesungguhnya
Allah SWT beserta orang-orang yang sabar”
(Al-Baqarah :153)**

“Keberhasilan adalah sebuah proses. Niatmu adalah awal keberhasilan. Peluh keringatmu adalah penyedapnya. Tetesan air matamu adalah pewarnanya. Doamu dan doa orang-orang disekitarmu adalah bara api yang mematangkannya. Kegagalan di setiap langkahmu adalah pengawetnya. Akan dari itu, bersabarlah, Karena Allah SWT selalu menyertai orang-orang yang penuh kesabaran dalam proses menuju keberhasilan.”(Muhammad Kamil)

PERSEMBAHAN

Tesis ini penulis persembahkan untuk:

- ✓ Allah SWT sebagai wujud rasa syukur atas ilmu yang Allah SWT Berikan kepadaku**
- ✓ Kepada kedua orang tua tercinta, ayah dan ibuku dalam ketulusan dari hati yang selalu mendoakan, memberikan nasehat dan dukungan**
- ✓ Teman-teman Magister Teknik Informatika**

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan tesis dengan judul **“Analisis Perbandingan Algoritma C4.5 dan Naïve Bayes pada Lulusan Tepat Waktu Mahasiswa di UIN Raden Fatah Palembang”**.

Tesis ini diajukan untuk memenuhi salah satu syarat untuk memperoleh gelar Magister Komputer Pada Universitas Bina Darma. Dalam Penulisan tesis ini lepas dari hambatan dan kesulitan, namun berkat bimbingan, bantuan, nasihat dan saran serta kerjasama dari berbagai pihak khususnya pembimbing, segala hambatan tersebut akhirnya dapat diatasi dengan baik.

Dalam penulisan tesis ini ternyata tidak lepas dari kekurangan baik dalam aspek kualitas maupun aspek kualitas dari materi penelitian yang disajikan. Semua ini didasarkan dari keterbatasan yang dimiliki penulis.

Penulis menyadari bahwa tesis ini jauh dari sempurna sehingga penulis membutuhkan kritik dan saran yang bersifat membangun untuk kemajuan pendidikan di masa yang akan datang. Selanjutnya dalam penulisan tesis ini penulis banyak diberi bantuan dari berbagai pihak.

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

1. Prof. Dr. Sunda Ariana, M.Pd., M.M. Selaku Rektor Universitas Bina Darma Palembang
2. Dr. Ir. Hasmawaty AR, M.M., M.T. Selaku Direktur Pascasarjana Universitas Bina Darma.
3. Darius Antoni, S.Kom., M.M., Ph.D. Selaku Ketua Program Studi Magister Teknik Informatika.

4. Dr. Widya Cholil, S.Kom., M.I.T. Selaku Pembimbing yang telah memberikan bimbingan dan arahan dalam penulisan tesis ini.
5. Kedua orang tua, atas semua nasehat, kasih sayang, kesabaran dalam mendidik, motivasi, semangat dan doa serta seluruh pengorbanan jiwa dan raga yang takkan pernah terbalaskan.
6. Pihak sekretariat pascasarjana Universitas Bina Darma Palembang yang telah memberikan bimbingan pelayanan dengan baik.
7. Berbagai pihak yang telah memberikan bantuan dorongan serta berbagai pengalaman pada proses penyusunan tesis ini.

Penulis mengharapkan para pembaca dapat memahami dan memaklumi jika ada kesalahan dalam penulisan pada pengungkapan keilmuan bidang ilmu komputer. Akhir katan, penulis mengucapkan banyak terima kasih dan semoga tesis ini dapat bermanfaat bagi kita semua.

Palembang, 25 Februari 2020

Penulis

Muhammad Kamil

NIM : 182420014

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN PEMBIMBING TESIS	ii
HALAMAN PENGESAHAN PENGUJI TESIS	iii
LEMBAR PERNYATAAN	iv
ABSTRAK	v
MOTTO DAN HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	2
1.3. Rumusan Masalah	2
1.4. Ruang Lingkup Penelitian	3
1.5. Tujuan Penelitian	3
1.6. Manfaat Penelitian	4
1.7. Susunan dan Struktur Tesis	4
BAB II KAJIAN PUSTAKA	6
2.1 Data	6
2.2 Sistem Basis Data	8
2.3 Data Mining	8
2.3.1. Tugas - Tugas Data Mining	8
2.3.2. Tahap – Tahap Data Mining	10
2.4 Klasifikasi	12
2.5 Algoritma C4.5	13

2.6 Naive Bayes.....	15
2.7 Pengukuran Kinerja Klasifikator	18
2.8 RapidMiner.....	19
2.9 Penelitian Terdahulu	20
2.10 Kerangka Pemikiran.....	21
BAB III METODOLOGI PENELITIAN	23
3.1 Metode Penelitian	23
3.2 Analisis Data Penelitian	24
3.3 Tahapan Penelitian.....	24
BAB IV ANALISIS DATA	27
4.1 Data Penelitian.....	27
4.2 Pengelohan Data Mining.....	27
BAB V HASIL DAN PEMBAHASAN	36
5.1 Implementasi Metode C4.5 dengan Aplikasi RapidMiner.....	35
5.2 Implementasi Metode Naive Bayes dengan Aplikasi RapidMiner .	37
5.3 Hasil Implementasi Klasifikasi Metode C4.5.....	38
5.4 Hasil Implementasi Klasifikasi Metode Naive Bayes	42
5.5 Perbandingan Tingkat Akurasi Metode Algoritma C4.5 dan Naive Bayes	43
BAB VI PENUTUP	45
6.1. Kesimpulan.....	45
6.2. Saran	46
DAFTAR PUSTAKA	
DAFTAR RIWAYAT HIDUP	
LAMPIRAN	

DAFTAR TABEL

Tabel 2.1 Contoh data yang digunakan dalam penelitian	6
Tabel 2.2 Pengkatagorian Atribut IPK.....	7
Tabel 2.3 Pengkatagorian Masa Studi	8
Tabel 2.4 Confusion Matrix	18
Tabel 5.1 Hasil Perhitungan Nilai Akurasi dari C4.5	39
Tabel 5.2 Hasil Perhitungan Nilai Akurasi dari Naive Bayes.....	42
Tabel 5.3 Perbandingan Metode C4.5 dan Naive Bayes	43

DAFTAR GAMBAR

Gambar 2.1 Tahap-tahap Data Mining	11
Gambar 2.2 Kerangka Pemikiran	21
Gambar 3.1 Model Tahapan Penelitian	25
Gambar 4.1 Data Alumni 2016-2019	28
Gambar 4.2 Data Alumni yang sudah cleaning.....	28
Gambar 4.3 Data Mahasiswa.....	29
Gambar 4.4 Data Alumni Mahasiswa.....	30
Gambar 4.5 Data Alumni Mahasiswa Testing	31
Gambar 4.6 Tahapan Preprocessing	32
Gambar 4.7 <i>discretize by binning</i>	33
Gambar 5.1 Proses Metode C4.5	35
Gambar 5.2 RapidMiner gambaran Clasification Decision Tree	36
Gambar 5.3 Proses Metode Naive Bayes	37
Gambar 5.4 RapidMiner Model Sub Proses Clasification Naive Bayes	38
Gambar 5.5 Implementasi Data Set pada RapidMiner	39
Gambar 4.6 Pohon Keputusan dengan Metode C4.5.....	41

DAFTAR LAMPIRAN

1. SK Pembimbing.....
2. Surat Izin Penelitian.....
3. Jurnal Tesis.....
4. Plagiat Turnitin Jurnal.....
5. Plagiat Turnitin Tesis.....
4. Sertifikat Seminar dan International Program.....
5. Lembar Perbaikan Tesis.....
6. Lembar Konsultasi.....