

SKRIPSI

**PERBANDINGAN ALGORITMA NAÏVE BAYES DAN ALGRORITMA
C4.5 DALAM PREDIKSI KELULUSAN MAHASISWA PENDIDIKAN
TEKNIK INFORMATIKA DAN KOMPUTER UNIVERSITAS NEGERI
JAKARTA**



NURISA HIKMATUL RAMADHANI

1512617025

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK


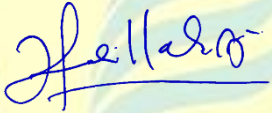
UNIVERSITAS NEGERI JAKARTA

2023

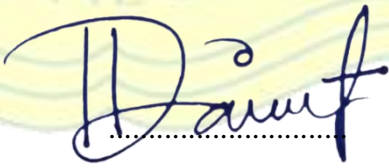


**PERBANDINGAN ALGORITMA NAÏVE BAYES DAN ALGRORITMA
C4.5 DALAM PREDIKSI KELULUSAN MAHASISWA PENDIDIKAN
TEKNIK INFORMATIKA DAN KOMPUTER UNIVERSITAS NEGERI
JAKARTA**

Nurisa Hikmatul Ramadhani, NIM. 1512617025

HALAMAN PENGESAHAN

NAMA DOSEN	TANDA TANGAN	TANGGAL
Dr. Widodo, S.Kom,M.Kom. (Dosen Pembimbing I)		(31 Januari 2023)
Hamidillah Ajie, S.Si., M.T. (Dosen Pembimbing II)		(31 Januari 2023)

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Diat Nurhidayat, S.Pd, M.T.I (Ketua Penguji)		(31 Januari 2023)
Bambang Prasetya Adhi, S.Pd, M.Kom (Penguji I)		(27 Januari 2023)
Murien Nugraheni, S.T., M.Cs. (Penguji II)		(26 Januari 2023)

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Karya tulis skripsi peneliti yang berjudul Perbandingan Algoritma Naïve Bayes dan Algoritma C4.5 dalam Prediksi Kelulusan Mahasiswa Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta merupakan karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain;
2. Karya tulis skripsi peneliti yang berjudul Perbandingan Algoritma Naïve Bayes dan Algoritma C4.5 dalam Prediksi Kelulusan Mahasiswa Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka;
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 25 Januari 2023

Yang membuat pernyataan



Nurisa Hikmatul Ramadhani
No. Reg. 1512617025



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Nurisa Hikmatul Ramadhani
NIM : 1512617025
Fakultas/Prodi : Teknik/Pendidikan Teknik Informatika dan Komputer
Alamat email : nurisahikmatul@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN ALGRORITMA C4.5 DALAM
PREDIKSI KELULUSAN MAHASISWA PENDIDIKAN TEKNIK INFORMATIKA DAN
KOMPUTER UNIVERSITAS NEGERI JAKARTA

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta

Penulis

(Nurisa Hikmatul Ramadhani)

KATA PENGANTAR

Dengan mengucapkan syukur Alhamdulillah segala puji bagi Allah atas karunia dan rahmatnya serta rezeki yang telah dilimpahkan kepada penulis sehingga dapat menyelesaikan skripsi ini. Skripsi dengan judul **“Perbandingan Algoritma Naïve Bayes dan Algoritma C4.5 dalam Prediksi Kelulusan Mahasiswa Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta”** yang disusun sebagai salah satu persyaratan meraih gelar Sarjana Pendidikan.

Penulis menyadari penyusunan skripsi ini masih terdapat banyak kekurangan dan kekeliruan dikarenakan kekurangan dan keterbatasan kemampuan serta pengalaman yang penulis miliki.

Penyelesaian skripsi ini tidak terlepas dari bantuan berbagai pihak. Oleh sebab itu, penulis ingin menyampaikan ucapan terima kasih dan penghargaan setinggi-tingginya kepada yang terhormat:

1. Bapak Dr. Widodo, M.Kom selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer Fakultas Teknik Universitas Negeri Jakarta serta dosen pembimbing I dan Bapak Hamidillah Ajie, S.Si, M.T. selaku dosen pembimbing II yang telah memberikan bimbingan dengan baik sehingga penulis dapat menyelesaikan skripsi ini;
2. Seluruh dosen PTIK yang telah memberikan ilmu selama di perkuliahan;
3. Keluarga saya yaitu Ayah, Ibu, Dhila, Najla yang telah banyak membantu, memberi semangat, memotivasi dan mendoakan penulis;
4. Kak Nadya telah membimbing dan memberikan motivasi agar menyelesaikan skripsi;
5. Sahabat saya, Zahra Tiarani dan Raisha Putri Mellisa yang selalu mengingatkan dan memberi semangat;
6. Teman-teman saya di Angkatan 2017, Angel Karentia, Untari Rachmadani, Chairani Wulandari sebagai teman seperjuangan yang selalu memberi semangat selama kuliah;
7. Sahabat saya, Daus, Bolis, Kiki, Pamela, Anisa, Nisa, Yokhebed, Amoy, Dwi, Windri yang memberi motivasi secara tidak langsung kepada penulis untuk menyelesaikan skripsi;

8. Seluruh pihak yang terlibat secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari keterbatasan kemampuan dan pengetahuan yang dimiliki dalam penyusunan skripsi ini. Oleh karena itu, penulis memohon maaf apabila terdapat kesalahan dalam penyusunan skripsi ini. Penulis berharap agar penyusunan skripsi ini dapat bermanfaat bagi para pembaca dan semua pihak terkait.

Jakarta, 23 Januari 2023



Nurisa Hikmatul Ramadhani



*Mencerdaskan dan
Memartabatkan Bangsa*

ABSTRAK

NURISA HIKMATUL RAMADHANI, Perbandingan Algoritma Naïve Bayes dan Algoritma C4.5 dalam Prediksi Kelulusan Mahasiswa Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta. Program studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta. Dosen Pembimbing Dr. Widodo, M.Kom dan Hamidillah Ajie, S.Si, M.T.

Berdasarkan data yang diperoleh dari tata usaha Program Studi Pendidikan Teknik Informatika dan Komputer didapatkan jumlah persentase mahasiswa yang lulus tepat waktu hanya sebesar 32,72%. Maka dari itu dibutuhkan sebuah prediksi kelulusan mahasiswa dapat digunakan lebih lanjut untuk membantu program studi dalam mengevaluasi dan memperbaiki sistem pembelajaran sehingga program studi dapat menghasilkan lulusan yang berkualitas. Prediksi kelulusan mahasiswa dapat dilakukan menggunakan algoritma Naive Bayes dan algoritma C4.5. Berdasarkan data yang ada maka dilakukan penelitian untuk mencari mekanisme dalam memprediksi kelulusan mahasiswa dengan menggunakan perbandingan algoritma Naïve Bayes dan algoritma C.45. Perbandingan algoritma Naïve Bayes dan algoritma C.45 untuk mencari tahu dari kedua algoritma tersebut mana yang memberikan kinerja terbaik. Data yang digunakan pada penelitian ini yaitu data mahasiswa lulusan tahun 2017-2020. Di mana atribut yang digunakan yaitu jenis kelamin, IP Semester 1-5, IPK Kelulusan, asal daerah, asal sekolah dan jalur masuk Algoritma Naïve Bayes adalah metode yang menghitung probabilitas dari tingkat kemunculan data yang satu terhadap data yang lainnya. Sedangkan algoritma C4.5 adalah satu dari sebagian algoritma dalam metode *decision tree* yang mengubah data menjadi pohon keputusan, kemudian dapat disimpulkan menjadi *rule-rule*. Hasil implementasi dari algoritma Naive Bayes dan C4.5 akan dianalisis menggunakan *confusion matrix*. Berdasarkan hasil pengujian dengan mengukur kinerja kedua metode tersebut menggunakan metode pengujian *Confusion Matrix* dalam *tool* Weka diketahui bahwa algoritma Naive Bayes memiliki akurasi yang lebih baik yakni sebesar 81,2138%, *precision* 83%, *recall* 81%, sedangkan algoritma C4.5 memiliki nilai akurasi sebesar 75,6345%, *precision* 76%, dan *recall* 75%. Berdasarkan hasil perbandingan tingkat akurasi, presisi dan recall dari kedua algoritma, didapatkan hasil yakni Algoritma Naive lebih baik dibandingkan algoritma C4.5 dalam prediksi kelulusan mahasiswa.

Kata Kunci: prediksi kelulusan, perbandingan, algoritma c4.5, algoritma naïve bayes

ABSTRACT

NURISA HIKMATUL RAMADHANI, Comparison of the Naïve Bayes Algorithm and the C4.5 Algorithm in the Prediction of Graduation in Informatics and Computer Engineering Education Students, Jakarta State University. Informatics and Computer Engineering Education Study Program, Faculty of Engineering, Jakarta State University. Supervisor Dr. Widodo, M.Kom and Hamidillah Ajie, S.Si, M.T.

Based on data obtained from the administration of the Informatics and Computer Engineering Education Study Program, it was found that the percentage of students who graduated on time was only 32.72%. Therefore a prediction of student graduation is needed which can be used further to assist study programs in evaluating and improving the learning system so that study programs can produce quality graduates. Prediction of student graduation can be done using the Naive Bayes algorithm and the C4.5 algorithm. Based on the existing data, research was conducted to find a mechanism for predicting student graduation by using a comparison of the Naïve Bayes algorithm and the C.45 algorithm. Comparison of the Naïve Bayes algorithm and the C.45 algorithm to find out which of the two algorithms provides the best performance. The data used in this study is data on graduate students in 2017-2020. Where the attributes used are gender, Semester 1-5 GPA, Graduation GPA, region of origin, school of origin and route of entry. The Naïve Bayes Algorithm is a method that calculates the probability of the occurrence rate of one data relative to other data. While the C4.5 algorithm is one of several algorithms in the method *decision tree* which turns the data into a decision tree, then it can be concluded to be *rule-rule*. The results of the implementation of the Naive Bayes and C4.5 algorithms will be analyzed using *confusion matrix*. Based on the test results by measuring the performance of the two methods using the test method *confusion matrix* in tool Weka is known that the Naive Bayes algorithm has a better accuracy of 81.2138%, *precision* 83%, *recall* 81%, while the C4.5 algorithm has an accuracy value of 75.6345%, *precision* 76%, and *recall* 75%. Based on the results of a comparison of the levels of accuracy, precision and recall of the two algorithms, the result is that the Naive Algorithm is better than the C4.5 algorithm in predicting student graduation.

Keywords: Graduation Prediction, Comparison, C4.5 Algorithm, Naïve Bayes Algorithm

Memartabatkan Bangsa

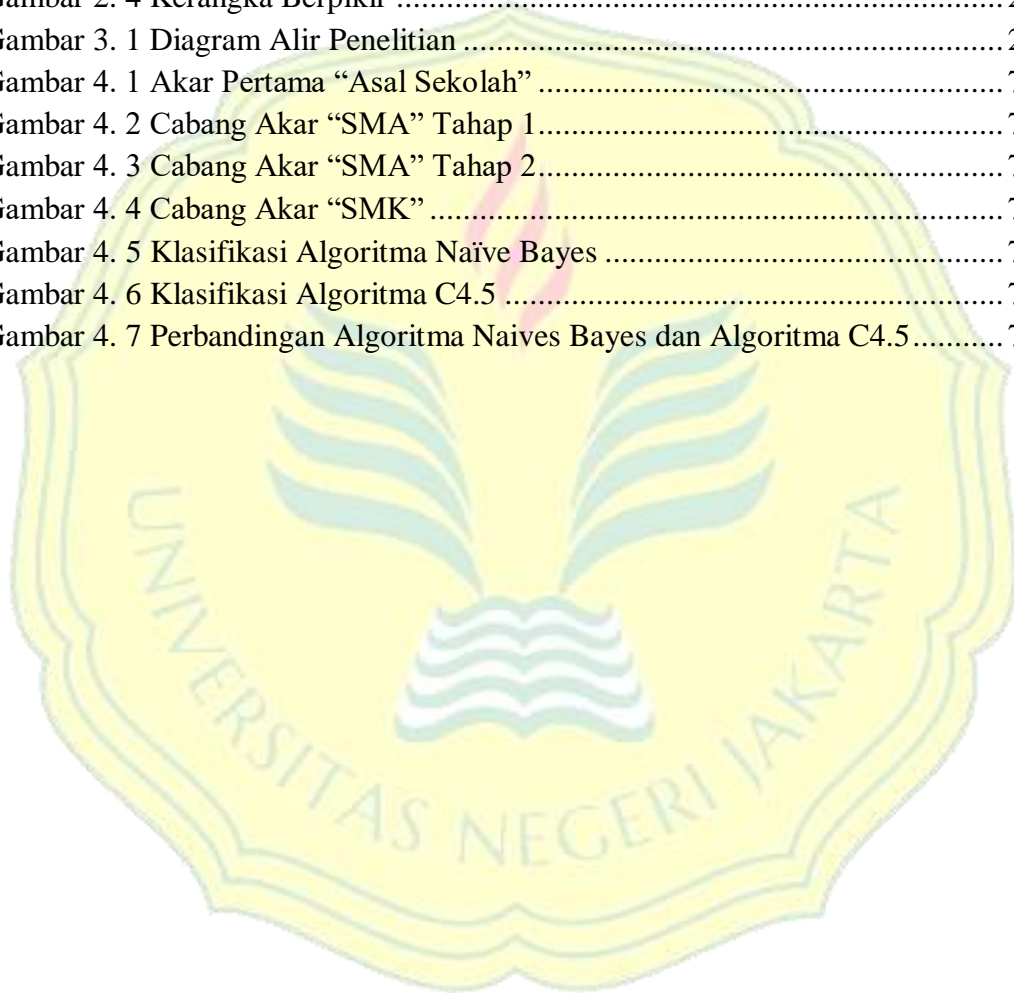
DAFTAR ISI

LEMBAR PERNYATAAN	ii
KATA PENGANTAR	iii
ABSTRAK	v
ABSTRACT	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Batasan Masalah	5
1.4 Rumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Kegunaan Penelitian	6
1.6.1 Kegunaan Teoritis	6
1.6.1 Kegunaan Praktis	6
BAB II KERANGKA TEORITIK	7
2.1 Landasan Teori	7
2.1.1 <i>Data mining</i>	7
2.1.2 Klasifikasi	9
2.1.3 Definisi dan Konsep Prediksi	10
2.1.4 Algoritma Naïve Bayes	11
2.1.5 Algoritma C4.5	14
2.1.6 WEKA	17
2.1.7 Pengujian Akurasi	18
2.2 Penelitian yang Relevan	19
2.3 Kerangka Berpikir	25
BAB III METODOLOGI PENELITIAN	27
3.1 Tempat dan Waktu Penelitian	27
3.2 Alat dan Bahan Penelitian	27
3.2.1 Alat Penelitian	27

3.2.2 Bahan Penelitian	28
3.3 Proses Penelitian.....	28
3.3.1 Teknik Pengumpulan Data	28
3.3.2 Mengolah Data.....	30
3.3.3 Membersihkan Data	30
3.3.4 Transformasi Data.....	30
3.3.5 Membuat <i>Dataset</i>	30
3.3.6 Implementasi dan Pengujian Data	31
3.3.7 Analisis Data.....	34
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	36
4.1 Mengolah Data	36
4.2 Membersihkan Data.....	36
4.3 Transformasi Data	39
4.4 Membuat Dataset.....	41
4.5 Penerapan Algoritma Naïve Bayes.....	43
4.6 Penerapan Algoritma C4.5.....	54
4.7 Pengujian dengan Aplikasi WEKA	74
4.7.1 Hasil Klasifikasi Algoritma Naïve Bayes menggunakan WEKA	74
4.7.2 Hasil Klasifikasi Algoritma C4.5 menggunakan WEKA.....	76
4.8 Analisis Data Penelitian.....	78
4.9 Pembahasan.....	78
4.10 Aplikasi Hasil Penelitian.....	79
BAB V KESIMPULAN DAN SARAN.....	80
5.1 Kesimpulan.....	80
5.2 Saran	80
DAFTAR PUSTAKA	81
LAMPIRAN	85
TENTANG PENULIS	168

DAFTAR GAMBAR

Gambar 2.1 Proses <i>Knowledge Discovery in Database (KDD)</i>	8
Gambar 2.2 Alur Metode Naïve Bayes (Saleh, 2015)	12
Gambar 2. 3 Aplikasi WEKA.....	17
Gambar 2. 4 Kerangka Berpikir	26
Gambar 3. 1 Diagram Alir Penelitian	29
Gambar 4. 1 Akar Pertama “Asal Sekolah”	70
Gambar 4. 2 Cabang Akar “SMA” Tahap 1.....	71
Gambar 4. 3 Cabang Akar “SMA” Tahap 2.....	72
Gambar 4. 4 Cabang Akar “SMK”	73
Gambar 4. 5 Klasifikasi Algoritma Naïve Bayes	75
Gambar 4. 6 Klasifikasi Algoritma C4.5	77
Gambar 4. 7 Perbandingan Algoritma Naives Bayes dan Algoritma C4.5.....	79



*Mencerdaskan dan
Memartabatkan Bangsa*

DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian	22
Tabel 3.1 Alat Penelitian.....	27
Tabel 3. 2 Tabel Atribut <i>Dataset</i>	32
Tabel 3.3 Tabel Kelas Data	32
Tabel 3.4 Tabel <i>Confusion Matrix</i>	34
Tabel 4. 1 <i>Data Training</i>	37
Tabel 4. 2 Hasil Transformasi Data	40
Tabel 4. 3 Rekapitan Hasil Perhitungan Peluang Tiap Atribut Kategorial	44
Tabel 4. 4 <i>Data Testing</i>	46
Tabel 4. 5 Hasil Prediksi Kelulusan Pada <i>Data Testing</i>	50
Tabel 4. 6 <i>Confusion Matrix</i>	53
Tabel 4. 7 <i>Entropy</i> Total Kasus	54
Tabel 4. 8 Hasil Perhitungan 1	57
Tabel 4. 9 Hasil Perhitungan <i>Node</i> Cabang SMA	61
Tabel 4. 10 Hasil Perhitungan Akhir Cabang SMA	64
Tabel 4. 11 Hasil Perhitungan <i>Node</i> Cabang SMK	68
Tabel 4. 12 Hasil Perhitungan Akhir Cabang SMK	69
Tabel 4. 13 Hasil Pengujian <i>Stratified Cross-validation</i>	74
Tabel 4. 14 <i>Confusion matrix</i> algoritma Naïve Bayes	74
Tabel 4. 15 <i>Detailed Accuracy</i> Algoritma Naive Bayes	75
Tabel 4. 16 Hasil Pengujian <i>Stratified Cross-validation</i>	76
Tabel 4. 17 <i>Confusion matrix</i> algoritma C4.5	76
Tabel 4. 18 <i>Detailed Accuracy</i> Algoritma C4.5	77

Mencerdaskan dan
Memartabatkan Bangsa

DAFTAR LAMPIRAN

LAMPIRAN. 1 Rekap Nilai <i>Mean</i> dan Standar Deviasi Klasifikasi “Tepat Waktu”	85
LAMPIRAN. 2 Rekap Nilai <i>Mean</i> dan Standar Deviasi Klasifikasi “Tidak Tepat Waktu”	93
LAMPIRAN. 3 Perhitungan <i>Entropy</i> Atribut. = SMA.....	101
LAMPIRAN. 4 Perhitungan <i>Entropy</i> Atribut. = SMK.....	145
LAMPIRAN. 5 <i>Dataset</i> Prediksi Kelulusan Mahasiswa Pendidikan Teknik Informatika Universitas Negeri Jakarta 2017-2020.....	165



*Mencerdaskan dan
Memartabatkan Bangsa*