

SKRIPSI

**ANALISIS PERBANDINGAN KINERJA ALGORITMA C4.5 DAN
NAÏVE BAYES UNTUK KLASIFIKASI
PENERIMAAN SISWA BARU**
(Studi Kasus: MTS Modern Al Azhary Ajibarang)



Oleh:

RISA NUR ROHMAH

42419065

PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN

2024

PERSETUJUAN SKRIPSI

JUDUL : ANALISIS PERBANDINGAN KINERJA ALGOTIRMA C4.5
DAN NAÏVE BAYES UNTUK KLASIFIKASI PENERIMAAN
SISWA BARU
(STUDI KASUS: MTS MODERN AL AZHARY AJIBARANG)

NAMA : RISA NUR ROHMAH

NIM : 42419065

Skripsi ini telah diperiksa dan disetujui oleh Dosen Pembimbing tugas akhir guna mencapai Gelar Sarjana Komputer Starta satu pada Fakultas Sains dan Teknologi, Universitas Peradaban.

Bumiayu, 17 November 2023

Pembimbing 1



Tezhar Ravendra.T.P.N, M.Kom

NIDN. 0619019201

Pembimbing 2



Nurul Mega Saraswati, M. Kom

NIDN. 0606069102

Ketua Jurusan



Kharotul Aeni, M.Kom

NIDN. 0618098802

PENGESAHAN SKRIPSI

JUDUL : ANALISIS PERBANDINGAN KINERJA ALGOTIRMA C4.5
DAN NAÏVE BAYES UNTUK KLASIFIKASI PENERIMAAN
SISWA BARU
(STUDI KASUS: MTS MODERN AL AZHARY AJIBARANG)

NAMA : RISA NUR ROHMAH

NIM : 42419065

Skripsi ini telah disetujui dan dipertahankan dihadapan Dewan Penguji. Pada tanggal 03 Januari 2024. Menurut pandangan kami. Skripsi ini memadai dari segi kualitas untuk memperoleh gelar Sarjana Komputer (S.Kom)

Bumiayu, 03 Januari 2024

Nama Penguji

1. Sorikhi, M.Kom
2. Khurotul Aeni, M.Kom
3. Nurul Mega Saraswati, M.Kom
4. Tezhar Rayendra.T.P.N, M.Kom

Tanda Tangan



Dekan Fakultas Sains dan Teknologi

Universitas Peradaban



Dr. apt. Pudiono, SU
NUPN. 9990000424

Ketua Jurusan

Informatika



Khurotul Aeni, M.Kom
NIDN. 0618098802

PERNYATAAN ORISINILITAS

JUDUL : ANALISIS PERBANDINGAN KINERJA ALGOTIRMA C4.5
DAN NAÏVE BAYES UNTUK KLASIFIKASI PENERIMAAN
SISWA BARU
(STUDI KASUS: MTS MODERN AL AZHARY AJIBARANG)

NAMA : RISA NUR ROHMAH

NIM : 42419065

“Saya menyatakan dan bertanggungjawab dengan sebenar-benarnya bahwa skripsi ini adalah hasil karya saya sendiri dan bukan merupakan pengambilan tulisan atau pikiran orang lain yang saya akui sebagai tulisan dan pikiran saya, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa skripsi ini hasil plagiat maka saya bersedia menerima sanksi berupa pencabutan gelar sarjana Komputer beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.

Bumiayu, 17 November 2023

Penulis



Risa Nur Rohmah

NIM. 42419065

KATA PENGANTAR

Bismillahirrahmaanirrahiim

Alhamdulillahilahi Robbil 'aalamin, Segala puji syukur bagi Allah 'azza wa jalla dengan segala rahmat, Nikmat, Hidayah dan Inaya-Nya, sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam semoga tetap tercurahkan dan terlimpahkan kepada Sang Baginda Rasul *Muhammad SAW*, beserta kepada keluarga, para sahabat dan penerus risalahnya, karena atas segala perjuangan beliau selama hidup telah mewariskan ilmu serta penuntun hidup yang mencerahkan umat manusia, semoga kita sebagai penerus risalah beliau, selalu mendapatkan syafa'atnya Amiin.

Alhamdulillah, dengan segala ikhtiar dan do'a, penulis dapat menyelesaikan penyusunan skripsi dengan judul: Analisis Kinerja Algoritma C4.5 dan Naïve Bayes untuk Klasifikasi Penerimaan Siswa baru untuk di jadikan syarat memperoleh gelar Sarjana Komputer di Fakultas Sains dan Teknologi Universitas Peradaban Bumiayu. Dalam penyusunan skripsi ini tentu tidak akan selesai tanpa adanya bantuan dan kerjasama dari berbagai pihak. Oleh karena itu, melalui kesempatan ini penulis selayaknya menyampaikan terima kasih sebesar-besarnya kepada:

1. Bapak Dr. Muh. Kadarisma, S.H., M.Si. selaku Rektor Universitas Peradaban Bumiayu
2. Bapak Dr. apt. Pudjono, S.U. selaku Dekan Fakultas Sains dan Teknologi Universitas Peradaban Bumiayu
3. Ibu Khurotul Aeni, M.Kom. selaku Kepala Program Studi Informatika
4. Bapak Tezhar Rayendra TPN, M.Kom. selaku Dosen Pembimbing Skripsi yang sabar dan memberikan arahan supaya hasil dari karya ilmiah ini maksimal.
5. Ibu Nurul Mega Saraswati, M.Kom. selaku Dosen Pembimbing Skripsi yang selalu sabar dalam memberikan arahan supaya hasil dari karya ilmiah ini maksimal serta selalu mengingatkan tentang penyusunan skripsi.
6. Seluruh Dosen Informatika Universitas Peradaban Bumiayu.
7. Kedua orangtua saya, Bapak Ritanto dan Ibu Tarminah yang selalu memberikan do'a dan dukungan penuh, mereka adalah alasan saya untuk tetap bertahan sejauh ini.

8. Kakak dan adik saya, Venny Anggraeni dan David Tri Pamungkas, yang selalu memberikan do'a dan semangat kepada saya yang selalu mengingatkan apa yang dimulai harus diselesaikan.
9. Bapak dan Ibu Guru MTs Modern Al Azhary Ajibarang yang telah membantu dan mendukung dalam proses penelitian.
10. Teman-teman program studi S-1 Informatika angkatan 2019, terimakasih atas dukungan dan kebersamaan selama penulis menempuh studi di program studi S-1 Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban Bumiayu.
11. Untuk grup idol EXO, NCT, SEVENTEEN, STRAY KIDS yang memberikan motivasi dan semangat dari karya-karyanya.
12. Tidak lupa untuk semua pihak yang memberikan penulis dukungan yang tidak dapat disebutkan satu persatu. Terimakasih banyak.

Penulis menyadari bahwa masih banyak kekurangan dari skripsi ini. Oleh karena itu, penulis memohon maaf atas segala kekurangan yang terjadi selama proses penyusunan skripsi ini.

Kepada semua yang telah memberikan ilmu dan pengalaman kepada penulis, semoga ilmu yang diberikan dapat bermanfaat bagi penulis sampai pada masa yang akan datang, semoga Allah SWT membalas kebaikan semuanya. Dengan selesainya skripsi ini penulis berharap semoga dapat bermanfaat bagi penulis dan pembaca khususnya. Aamiin

Bumiayu, 17 November 2023

Penulis



Risa Nur Rohmah

NIM. 42419065

DAFTAR ISI

SKRIPSI.....	i
PENERIMAAN SISWA BARU	i
PERSETUJUAN SKRIPSI.....	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN ORISINILITAS	iv
KATA PENGANTAR.....	v
DAFTAR ISI	v
DAFTAR GAMBAR.....	i
DAFTAR TABEL	
ABSTRAK	x
<i>ABSTRACT</i>	
	xii
	BAB I
PENDAHULUAN.....	1 1.1
Latar Belakang.....	1 1.2
Rumusan Masalah.....	4 1.3
Tujuan Penelitian.....	5 1.4
Manfaat Penelitian.....	5 1.5
Batasan Masalah	5 1.6
Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	
2.1 Penelitian Terkait.....	
2.2 Landasan Teori	1
2.2.1. Analisis	13
2.2.2. Perbandingan.....	13
2.2.3. Klasifikasi	13
2.2.4. Data Mining	13
2.2.5. Algoritma C4.5	14
2.2.6. Algoritma <i>Naïve Bayes</i>	17
2.2.7. Penerimaan Peserta Didik Baru	18
2.2.8. Confusion Matrix	18
2.2.9. Kurva ROC	19
2.2.10. Python	19
2.2.11. Google Colab	20
2.2.12. PHP	21

2.2.13. Xampp.....	21
2.2.14. MySQL	22
2.3 Kerangka Pemikiran	23
BAB III METODOLOGI PENELITIAN	25
3.1 Tahapan Penelitian	25
3.2 Pengumpulan Data.....	27
3.3 Pengolahan Data Awal	27
3.4 Metode yang Diusulkan.....	28
3.5 Experimen dan pengujian model	30
3.6 Evaluasi dan Validasi	30
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Identifikasi Masalah	31
4.2 Pengumpulan Data.....	31
4.3 Pengolahan Data Awal	32
4.4 Metode yang diusulkan.....	33
4.5 Eksperimen dan pengujian model.....	38
4.6 Evaluasi dan Validasi	41
BAB V PENUTUP.....	45
5.1 Kesimpulan.....	45
5.2 Saran	45
DAFTAR PUSTAKA.....	45

DAFTAR GAMBAR

Gambar 2. 1 Decision Tree (Pohon Keputusan)	17
Gambar 2. 2 Google Colab	20
Gambar 2. 3 Xampp	21
Gambar 2. 4 Kerangka Pemikiran	23
Gambar 3. 1 Tahapan Penelitian	25
Gambar 3. 2 Dataset calon siswa baru	27
Gambar 3. 3 Metode yang diusulkan	29
Gambar 4. 1 Dataset calon siswa baru	32
Gambar 4. 2 pohon keputusan.....	35
Gambar 4. 3 hasil akurasi algoritma C4.5.....	36
Gambar 4. 4 hasil akurasi algoritma Naïve Bayes	38
Gambar 4. 5 Halaman Sign In.....	38
Gambar 4. 6 Halaman Utama/ Dashboard	39
Gambar 4. 7 Halaman Tahun Ajaran	39
Gambar 4. 8 Halaman data prediksi	40
Gambar 4. 9 Tampilan data prediksi	40
Gambar 4. 10 Halaman Data Pengguna	41
Gambar 4. 11 Halaman Ganti Password	41

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	11
Tabel 2. 2 Confusion Matrix	18
Tabel 3. 1 Dataset yang telah ditransformasi	28
Tabel 4. 1 dataset yang sudah dilakukan preproceasing	32
Tabel 4. 2 Jumlah Tiap Atribut	33
Tabel 4. 3 Probabilitas kelas	37
Tabel 4. 4 Hasil Perhitungan	37
Tabel 4. 5 Hasil Evaluasi Model Menggunakan Data Testing.....	42
Tabel 4. 6 Hasil Confusion matrix C4.5	42
Tabel 4. 7 Hasil Confusion matrix Naïve Bayes.....	43
Tabel 4. 8 Hasil Uji Coba Program.....	44

ABSTRAK

Penerimaan peserta didik baru merupakan proses seleksi akademis calon siswa pada jenjang tertentu dengan aturan yang telah ditentukan. Menentukan siswa dan siswi yang bermutu dan berkualitas, maka dibutuhkan instrumen penilaian yang tepat dan akurat sehingga nantinya akan memperoleh siswa dan siswi sesuai dengan harapan sekolah. Dalam proses rekap penilaian masih menggunakan aplikasi *Microsoft Excel*. Hal ini rentan menyebabkan data penerimaan calon siswa baru tidak tepat sehingga membuat panitia kewalahan seperti mengurutkan perangkungan yang nilainya sama dan proses pengolahannya membutuhkan waktu yang lama. *Data mining* adalah gabungan sejumlah disiplin ilmu komputer yang mendefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database systems*. Algoritma yang digunakan adalah algoritma *C4.5* dan algoritma *Naïve Bayes*. Penelitian ini bertujuan untuk mengetahui tingkat akurasi dari dua algoritma metode klasifikasi terhadap penerimaan peserta didik baru. *Dataset* yang digunakan sebanyak 187 data. Hasil klasifikasi menunjukkan bahwa algoritma *C4.5* memiliki nilai akurasi sebesar 97,3%. Sedangkan algoritma *Naïve Bayes* mendapatkan nilai akurasi sebesar 92,1%. Sehingga dapat disimpulkan bahwa algoritma *C4.5* dan algoritma *Naïve Bayes* mendapatkan hasil yang sangat baik, tetapi algoritma *C4.5* memperoleh hasil yang lebih baik dibandingkan algoritma *Naïve Bayes*.

kata kunci: penerimaan peserta didik baru, *data mining*, *c4.5*, *naïve bayes*

ABSTRACT

New student screening is a process of academic selection of prospective students at a certain level with predetermined rules. Determining quality and qualified students, it requires a precise and accurate assessment instrument so that later it will get students and students in accordance with school expectations. In the process of recapitulating the assessment still uses the Microsoft Excel application. This is prone to causing the data on the acceptance of new student candidates to be inaccurate so that it overwhelms the committee such as sorting rankings that have the same value and the processing process takes a long time. Data mining is a combination of a number of computer science disciplines that define as the process of discovering new patterns from very large data sets, including methods that are slices of artificial intelligence, machine learning, statistics, and database systems. The algorithms used are the C4.5 algorithm and the Naïve Bayes algorithm. This study aims to determine the accuracy level of the two classification method algorithms on the admission of new students. The dataset used is 187 data. The classification results show that the C4.5 algorithm has an accuracy value of 97.3%. While the Naïve Bayes algorithm gets an accuracy value of 92.1%. So it can be concluded that the C4.5 algorithm and the Naïve Bayes algorithm get very good results, but the C4.5 algorithm gets better results than the Naïve Bayes algorithm.

keywords: new student screening *data mining*, *c4.5*, *naïve bayes*