

**PERBANDINGAN ALGORITMA C4.5 DAN *NAÏVE*  
BAYES MENGGUNAKAN METODE KLASIFIKASI  
DALAM PREDIKSI KETERLAMBATAN PEMBAYARAN  
KOS MAKAN DI PONPES NGALAH**



**SKRIPSI**

**Diajukan untuk memenuhi salah satu syarat  
Memperoleh gelar sarjana komputer**

**Oleh :**

**Lailatul Mufidah  
NIM. 201769040017**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS YUDHARTA PASURUAN  
2021**

## PERNYATAAN PENULIS

JUDUL : PERBANDINGAN ALGORITMA C4.5 DAN *NAIVE BAYES* MENGGUNAKAN METODE KLASIFIKASI DALAM PREDIKSI KETERLAMBATAN PEMBAYARAN KOS MAKAN

NAMA : LAILATUL MUFIDAH

NIM : 201769040017

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini karyanya, yang disertai bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.

Pasuruan, 31 Juli 2021



**PERSETUJUAN SKRIPSI**

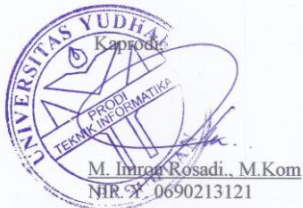
(Lembar ACC)

JUDUL : PERBANDINGAN ALGORITMA C4.5 DAN *NAIVE BAYES* MENGGUNAKAN METODE KLASIFIKASI DALAM PREDIKSI KETERLAMBATAN PEMBAYARAN KOS MAKAN DI PONDOK NGALAH

NAMA : LAILATUL MUFIDAH

NIM : 201769040017

Skripsi ini telah diperiksa dan disetujui  
Pasuruan, 04 Agustus 2021



Pembimbing,

A handwritten signature in black ink, appearing to read "Arif Faizin".

Arif Faizin, M.Kom  
NIP. Y. 0691707002

## PENGESAHAN SKRIPSI

JUDUL : PERBANDINGAN ALGORITMA C4.5 DAN *NAIVE BAYES* MENGGUNAKAN METODE KLASIFIKASI DALAM PREDIKSI KETERLAMBATAN PEMBAYARAN KOS MAKAN DI PONPES NGALAH

NAMA : LAILATUL MUFIDAH  
NIM : 201769040017

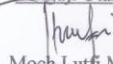
Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal 04 Agustus 2021. Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugrahan gelar Sarjana Komputer (S.Kom)

Pasuruan, 04 Agustus 2021

Pembimbing,

  
Arif Faizah, M.Kom  
NIP. Y. 0691707002

Penguji Utama,

  
Moch Lutfi M.Kom  
NIK. Y. 069170048901



M. Imron Rosadr., M.Kom  
NIK. Y. 0690213121

Penguji Anggota,

  
M. Faishol Amrulloh M.Kom  
NIK. Y. 0691709007



Dekan Fakultas Teknik,

Misbach Munir., S.T., M.T.  
NIK. Y. 0690201015

iv

*“Teruntuk bapak dan ibu saya yang sangat saya cintai,  
Kakak saya yang saya sayangi dan  
seseorang yang selalu memberikan semangat tiada henti,  
Kutujukan skripsi ini untuk kalian :\*”*

## **ABSTRACT**

*Pesantren is an independent religious education institution in terms of materials such as operational costs, facilities in the pesantren require a lot of money, one of which is the provision of food, at the Ngalah Islamic Boarding School the eating facilities for students are carried out using a boarding house where payments are made every month, to improve facilities within the pesantren. , the researcher predicts the late payment of the meal costs.*

*In this case the prediction is made using the classification method with data sourced from the daily board of the pesantren and the data that will be filled in directly by the pesantren students. The researcher used two algorithms as comparison, namely C4.5 and Naive Bayes to find the highest accuracy value.*

*From the calculation of 450 data which is divided into 75% training data and 25% testing data, the highest accuracy is in the C4.5 algorithm of 90%, while the accuracy of the Naive Bayes algorithm is 55%. The final result of the average comparison of the algorithms C4.5 and Naive Bayes using the  $t$ -test method by comparing the accuracy, precision and recall of the two algorithms resulted in an average of 64% for C4.5 and 62% for Naive Bayes.*

***Keywords: Prediction, Classification, C4.5, Naive Bayes***

## ABSTRAK

Pesantren merupakan lembaga pendidikan keagamaan yang mandiri dalam hal materi seperti biaya operasional, fasilitas didalam pesantren membutuhkan biaya yang tidak sedikit salah satunya adalah penyediaan makan, di Pondok Pesantren Ngalah fasilitas makan santri dilakukan menggunakan kos makan yang pembayarannya dilakukan setiap bulan, untuk meningkatkan fasilitas didalam pesantren, peneliti melakukan prediksi terhadap keterlambatan pembayaran kos makan.

Dalam hal ini prediksi dilakukan dengan menggunakan metode klasifikasi dengan data yang bersumber dari pengurus harian pesantren dan data yang akan diisi langsung oleh santri pondok pesantren. Peneliti menggunakan dua lagoritma sebagai pembandingan yaitu C4.5 dan *Naive Bayes* untuk menemukan nilai akurasi yang tertinggi.

Dari perhitungan 450 data yang dibagi menjadi 75% data training dan 25% data testing menghasilkan akurasi tertinggi ada pada algoritma C4.5 sebesar 90%, sedangkan akurasi dari algoritma *Naive Bayes* sebesar 55%. Hasil ahir dari rata-rata perbandingan algoritma C4.5 dan *Naive Bayes* menggunakan metode *t-test* dengan membandingkan akurasi, presisi dan recall kedua algoritma dihasilkan rata-rata 64% untuk C4.5 dan 62% untuk *Naive Bayes*.

**Kata kunci : Prediksi, Klasifikasi, C4.5, *Naive Bayes***

## KATA PENGANTAR

Alhamdulillah rabbi 'alamin, rasa syukur senantiasa penulis panjatkan kepada Allah SWT yang telah memberikan kasih sayang dan keridhoan-Nya sehingga penulis dapat menyelesaikan skripsi ini. Tujuan penulisan skripsi adalah untuk mendapatkan gelar sarjana komputer..

Penulisan skripsi ini tidak lepas dari bantuan berbagai pihak, untuk itu penulis mengucapkan beribu-ribu terimakasih kepada pihak-pihak yang telah membantu :

Murabbi Ruhina Romo KH. Sholeh Bahrudin, selaku Pengasuh Pondok Pesantren Ngalah yang senantiasa membimbing kami dan memberikan barokah do'anya. .

Bapak Dr. H. Kholid Murtadlo, S.E., M.E. selaku Rektor Universitas Yudharta Pasuruan.

Bapak Misbach Munir, S.T., M.T selaku Dekan Fakultas Teknik Universitas Yudharta Pasuruan.

Bapak M. Imron Rosadi, S.Kom., M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Yudharta Pasuruan.

Bapak Arif Faizin, M.Kom, selaku Dosen Pembimbing skripsi.

Santri dan pengurus asrama Pondok Pesantren Ngalah yang telah memberikan bantuan peran dalam memperlancar penelitian dan penulisan skripsi ini. .

Peluk kasih dan beribu sayang penulis sampaikan kepada Ayah, ibu dan kakak tercinta yang senantiasa mendoakan dan tiada henti memberikan semangat dan dukungannya.

Teman-teman Informatika 2017 seperjuangan dan keluarga kecil B.07 yang saling membantu dan memberi semangat.

Seseorang yang tiada henti memberikan dukungan dan semangatnya.

Pasuruan, 26 November 2020

Penulis



## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	<b>i</b>
<b>HALAMAN PERNYATAAN PENULIS</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>HALAMAN PERUNTUKAN</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>ix</b>
<b>DAFTAR TABEL</b> .....	<b>xi</b>
<b>DAFTAR GAMBAR</b> .....	<b>xii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan .....	3
1.4 Manfaat .....	3
1.5 Batasan Masalah .....	3
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>5</b>
2.1 Penelitian Terdahulu .....	5
2.2 Landasan Teori .....	18
2.2.1 Pondok Pesantren Ngalah .....	18
2.2.2 Data Mining .....	19
2.2.3 Metode Klasifikasi .....	21
2.2.4 Algoritma C4.5 .....	21
2.2.5 Algoritma <i>Naive Bayes</i> .....	23
2.2.6 Pengukuran Kinerja .....	24
<b>BAB III METODE PENELITIAN</b> .....	<b>27</b>
3.1 Kerangka Berfikir .....	27
3.2 Metodologi Penelitian .....	28
3.2.1 Studi Literatur dan Pengumpulan Data .....	28

3.2.2 Perancangan dan Pemodelan Sistem .....	28
3.2.3 Data Set .....	29
3.2.4 Data Training .....	29
3.2.5 Data Testing .....	29
3.2.6 Pengolahan Data .....	30
3.2.7 Alat Pendukung .....	33
3.3 Alur Penelitian .....	34
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>35</b>
4.1 Hasil Penelitian .....	35
4.1.1 Perhitungan C4.5 .....	35
4.1.2 Perhitungan <i>Naive Bayes</i> .....	44
4.2 Perhitungan Akurasi Dengan Phyton .....	50
4.2.1 Perhitungan C4.5 .....	50
4.2.2 Perhitungan <i>Naive Bayes</i> .....	53
4.3 Perbandingan Algoritma C4.5 dan <i>Naive Bayes</i> .....	61
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>65</b>
5.1 Kesimpulan .....	65
5.2 Saran .....	65

## DAFTAR PUSTAKA

## LAMPIRAN-LAMPIRAN

## DAFTAR TABEL

	Halaman
Tabel 2.1 Penelitian Terdahulu .....	13
Tabel 2.2 Model Perhitungan <i>Confusion Matrix</i> .....	24
Tabel 3.1 Kriteria Prediksi Keterlambatan Pembayaran Kos Makan .....	30
Tabel 4.1 Data Taining .....	35
Tabel 4.2 Penentuan Entropy .....	37
Tabel 4.3 Perhitungan Atribut Formal Record Tinggi .....	40
Tabel 4.4 Perhitungan Atribut Besar SPP Record 50000 .....	41
Tabel 4.5 Probabilitas Kelas .....	44
Tabel 4.6 Probabilitas Atribut Informal .....	45
Tabel 4.7 Probabilitas Atribut Formal .....	45
Tabel 4.8 Probabilitas Atribut Alamat .....	46
Tabel 4.9 Probabilitas Atribut Pekerjaan .....	46
Tabel 4.10 Probabilitas Atribut Besar SPP .....	46
Tabel 4.11 Probabilitas Atribut Prnghasilan .....	47
Tabel 4.12 Probabilitas Atribut Jumlah Keluarga .....	47
Tabel 4.13 Kelas Prediksi .....	48
Tabel 4.14 Hasil Prediksi .....	49
Tabel 4.15 Perbandingan Algoritma C4.5 Dan Naïve Bayes ..	61
Tabel 4.16 Hasil Perhitungan T-Test .....	62

## DAFTAR GAMBAR

	Halaman
Gambar 2.1 Metode yang diusulkan Istiana dan Waspada., 2019 .....	5
Gambar 2.2 Metode yang diusulkan Apandi et all., 2019 .....	6
Gambar 2.3 Metode yang diusulkan Gerhana et al., 2019 .....	7
Gambar 2.4 Metode yang diusulkan Ginting et al., 2020 .....	8
Gambar 2.5 Metode yang diusulkan Isa, 2017 .....	9
Gambar 2.6 Metode yang diusulkan Muqorobin et al., 2019 ....	10
Gambar 2.7 metode yang diusulkan T.H.Apandi et al., 2019 ...	11
Gambar 2.8 Metode yang diusulkan Kurniawan, 2018 .....	12
Gambar 2.9 Metode yang diusulkan Saputra et al., 2020 .....	12
Gambar 2.10 Metode yang diusulkan Fauziah et al., 2020 .....	13
Gambar 2.11 Langkah Data Mining .....	20
Gambar 3.1 Skema Kerangka Berpikir.....	27
Gambar 3.2 Use Case Diagram .....	28
Gambar 3.12 Diagram Alur Penelitian .....	34
Gambar 4.1 Pohon Keputusan .....	42
Gambar 4.2 Import Data .....	50
Gambar 4.3 Instalasi Chefboost .....	51
Gambar 4.4 Import Chefboost dan Pandas Library .....	51
Gambar 4.5 Konfigurasi Algoritma C4.5 .....	52
Gambar 4.6 Rule Algoritma C4.5 .....	52
Gambar 4.7 Hasil Perhitungan Akurasi .....	53
Gambar 4.8 Memasukkan Data Dari Drive .....	53
Gambar 4.9 Memasukan Library .....	54
Gambar 4.10 Membaca Data Dari Drive .....	55
Gambar 4.11Perintah Memisahkan Atribut Dengan Label .....	57
Gambar 4.12Perintah Menjalankan Training.....	58
Gambar 4.13Menerapkan Algoritma <i>Naive Bayes</i> .....	58
Gambar 4.14Proses Perhitungan Data Testing .....	59
Gambar 4.15Menampilakn Probabilitas .....	59
Gambar 4.16Menghitung Akurasi <i>Naive Bayes</i> .....	60
Gambar 4.17Menampilkan Hasil Prediksi .....	60
Gambar 4.18Menampilkan Hasil Akurasi, Presisi, dan Recall ..	61

## **DAFTAR LAMPIRAN**

- Lampiran 1. Data Santri
- Lampiran 2. Daftar Riwayat Hidup
- Lampiran 3. Kartu Seminar
- Lampiran 4. Lembar Bimbingan
- Lampiran 5. Surat Permohonan Izin Riset
- Lampiran 6. Surat Perizinan Riset

