

**PERBANDINGAN KINERJA METODE C4.5 DENGAN NAIVE BAYES CLASSIFIER
DALAM KLASIFIKASI JAMUR BERACUN DAN TIDAK BERACUN**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Iqbal Firzal
17.11.1079

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**PERBANDINGAN KINERJA METODE C4.5 DENGAN NAÏVE BAYES
CLASSIFIER DALAM KLASIFIKASI JAMUR
BERACUN DAN TIDAK BERACUN**

yang dipersiapkan dan disusun oleh

Iqbal Firzal

17.11.1079

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 5 Januari 2021

Dosen Pembimbing

Supriatin, M.Kom.

NIK. 190302239

PENGESAHAN
SKRIPSI
PERBANDINGAN KINERJA METODE C4.5 DENGAN NAÏVE BAYES
CLASSIFIER DALAM KLASIFIKASI JAMUR
BERACUN DAN TIDAK BERACUN

yang dipersiapkan dan disusun oleh

Iqbal Firzal

17.11.1079

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Januari 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Sharazita Dyah Anggita, M.Kom.
NIK. 190302285

Wiwi Widayani, M.Kom.
NIK. 190302272

Supriatin, M.Kom.
NIK. 190302239

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 19 Januari 2021

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Semarang, 7 Maret 2021



Iqbal Firzal

NIM 17.11.1079

KATA PENGANTAR

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Panyayang, Kami panjatkan puja dan puji syukur atas kehadiran-Nya, yang telah melimpahkan rahmat, hidayah, dan inayah Nya kepada kami, sehingga penulis dapat menyelesaikan penyusunan skripsi ini, sebagai syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi S1 Informasika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Skripsi ini telah saya susun dengan maksimal dan mendapatkan bantuan dari berbagai pihak sehingga dapat memperlancar pembuatan skripsi ini. Oleh karena itu, dalam kesempatan ini penulis menyampaikan ucapan terima kasih kepada :

Orang tua, keluarga, dan kekasih saya yang telah memberikan dukungan dan semangat. Kemudian untuk dosen pembimbing dan dosen penguji serta dosen-dosen yang saya kenal di Universitas Amikom Yogyakarta yang telah memberikan bimbingan dan arahan kepada saya dalam menjalani studi. Terakhir untuk teman-teman dekat di Yogyakarta, baik yang menjadi teman seperjuangan di kontrakan, di kampus, maupun di organisasi karena sudah memberikan pengalaman tentang arti nilai-nilai kehidupan.

Demikian Skripsi ini, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan bantuan baik moril maupun materil sehingga skripsi ini dapat diselesaikan. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih banyak kekurangan. Hal tersebut dikarenakan keterbatasan ilmu dan pengetahuan yang dimiliki penulis semata. Oleh karena itu penulis mengharapkan kritik dan saran dari pembaca untuk menyempurnakan penulisan skripsi ini. Semoga skripsi ini dapat memberikan manfaat bagi penulis khususnya dan umumnya bagi semua pihak yang membutuhkan. Akhir kata, semoga Allah SWT senantiasa melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal'alamiin.

Semarang, Maret 2021

Penulis

DAFTAR ISI

JUDUL.....	i
DAFTAR ISI.....	vi
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS.....	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Maksud dan Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	3
1.5.1 Bagi Mahasiswa.....	3
1.5.2 Bagi Masyarakat.....	3
1.5.3 Bagi Peneliti Lain.....	3
1.6 Metode Penelitian.....	3
1.6.1 Studi Literatur.....	3
1.6.2 Analisis Permasalahan.....	4
1.6.3 Perancangan Sistem.....	4
1.6.4 Implementasi.....	4
1.6.5 Pengujian.....	4
1.6.6 Penyusunan Laporan.....	4

1.6.7	Sistematika Penulisan	4
BAB II LANDASAN TEORI.....		6
2.1	Kajian Pustaka.....	6
2.2	Dasar Teori.....	9
2.2.1	Jamur.....	9
2.2.2	<i>Data Set</i>	10
2.2.3	UCI (<i>University of California, Irvine</i>) <i>Machine Learning Repository</i>	10
2.2.4	<i>Mushroom Data Set</i>	11
2.2.5	<i>Data Mining</i>	11
2.2.6	<i>Naïve Bayes Classifier</i>	12
2.2.7	<i>Decision Tree C4.5</i>	13
2.3	Peralatan dan Perancangan Aplikasi Web.....	15
2.3.1	<i>Waterfall Model</i>	15
2.3.2	Aplikasi Web	15
2.3.3	HTML (<i>Hypertext Markup Language</i>).....	15
2.3.4	CSS (<i>Cascading Style Sheets</i>).....	15
2.3.5	Bootstrap.....	16
2.3.6	PHP (<i>Hypertext Preprocessor</i>)	16
2.3.7	MVC (<i>Model-View-Controller</i>).....	16
2.3.8	MySQL Database	16
2.4	Konsep Pemodelan Data	16
2.4.1	UML (<i>Unified Modelling Language</i>).....	16
2.4.2	<i>Use Case Diagram</i>	16
2.4.3	<i>Activity Diagram</i>	17
2.4.4	<i>Sequence Diagram</i>	18

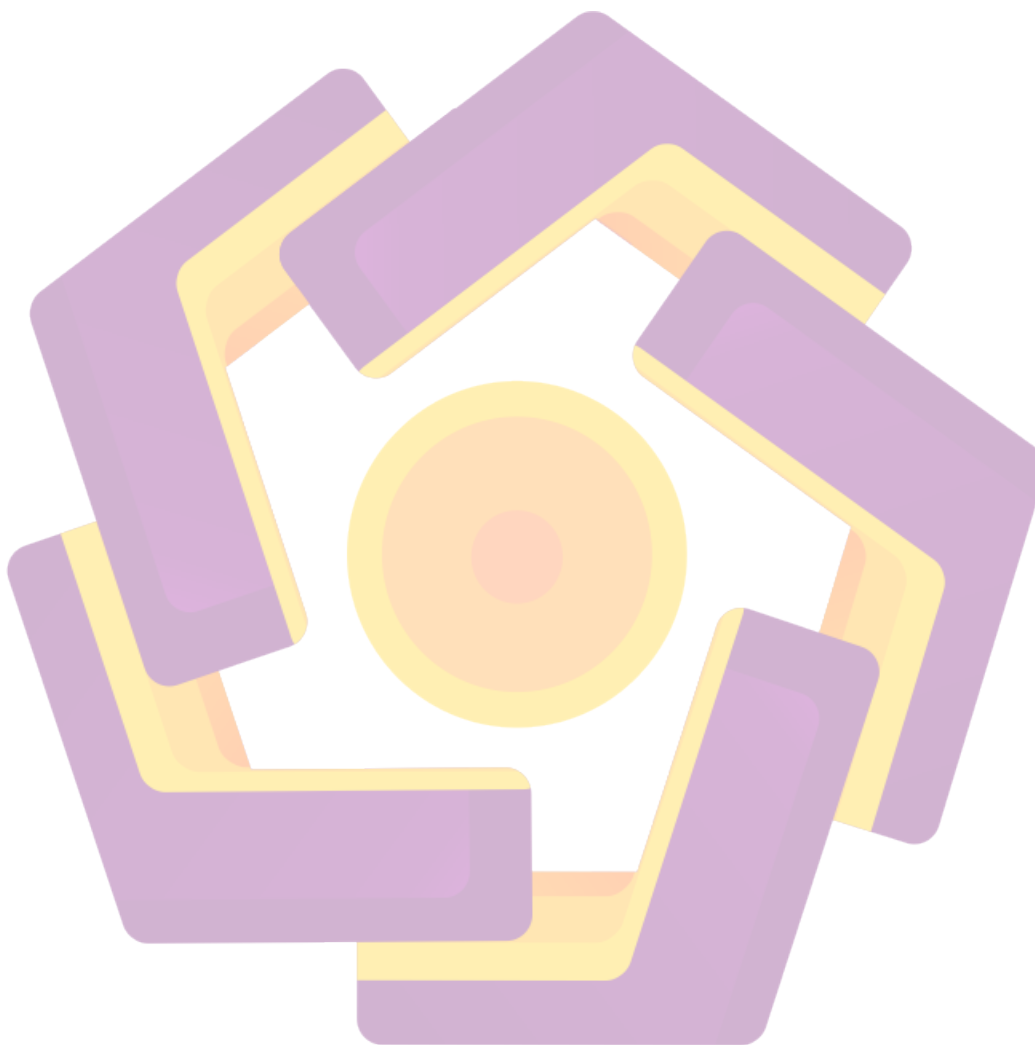
2.4.5	<i>Class Diagram</i>	20
2.4.6	<i>Entity Rational Diagram (ERD)</i>	21
2.5	Pengujian Sistem.....	22
2.5.1	<i>Black Box Testing</i>	22
2.5.2	<i>Confusion Matrix</i>	22
BAB III METODE PENELITIAN		25
3.1	Alur Penelitian.....	25
3.2	Metode Pengumpulan Data.....	25
3.3	Metode Analisis Data.....	26
3.3.1	Pemilihan Data / <i>Data Selection</i>	26
3.3.2	Pemrosesan Awal / <i>Pre Processing</i>	31
3.4	Metode Penambangan Data.....	31
3.4.1	<i>Naïve Bayes Classifier</i>	33
3.4.2	C4.5.....	35
3.5	Metode Pengukuran Nilai Akurasi.....	41
3.6	Metode Pengukuran Durasi Waktu.....	42
3.7	Metode Pengembangan Sistem.....	42
3.8	Perencanaan / <i>Planning</i>	42
3.9	Analisis Kebutuhan / <i>Requirement Analysis</i>	43
3.9.1	Kebutuhan Fungsional.....	43
3.9.2	Kebutuhan Non Fungsional.....	43
3.10	Rancangan Sistem / <i>System Design</i>	44
3.10.1	<i>Use Case Diagram</i>	44
3.10.2	<i>Activity Diagram</i>	45
3.10.3	<i>Sequence Diagram</i>	53

3.10.4	<i>Class Diagram</i>	56
3.10.5	Rancangan <i>Database</i>	59
3.10.6	Rancangan Tampilan Aplikasi.....	62
BAB IV HASIL DAN PEMBAHASAN		66
4.1	Implementasi / <i>Implementation</i>	66
4.1.1	Implementasi Perangkat Keras	66
4.1.2	Implementasi Perangkat Lunak.....	66
4.1.3	Implementasi <i>Database</i>	66
4.1.4	Implementasi Algoritma	67
4.1.5	Implementasi Tampilan	74
4.2	Integrasi & Uji Coba / <i>Integratation & Testing</i>	76
4.3	Hasil Penambangan Data	80
4.3.1	Aspek Yang diukur	80
4.3.2	Perhitungan Klasifikasi.....	80
4.3.3	Perhitungan Akurasi.....	82
4.3.4	Durasi Waktu Perhitungan.....	84
4.4	Analisa Hasil	85
4.4.1	Analisa Hasil Klasifikasi.....	85
4.4.2	Analisa Hasil Akurasi	85
4.4.3	Analisa Durasi Waktu Perhitungan.....	86
4.4.4	Analisa Terhadap Penelitian Terdahulu Yang Relevan	88
4.5	Kelebihan dan Kekurangan Sistem	89
4.5.1	Kelebihan Sistem	89
4.5.2	Kekurangan Sistem.....	89
BAB V PENUTUP		90

5.1 Kesimpulan..... 90

5.2 Saran..... 90

DAFTAR PUSTAKA 91



DAFTAR TABEL

<i>Tabel 2.1 Hasil Pengujian Pertama Perbandingan NBC dan ID Penelitian Lain</i>	6
<i>Tabel 2.2 Hasil Pengujian Kedua Perbandingan NBC dan ID3 Penelitian Lain</i>	6
<i>Tabel 2.3 Hasil Pengujian Ketiga Perbandingan NBC dan ID3 Penelitian Lain</i>	7
<i>Tabel 2.4 Hasil Uji NBC Penelitian Lain</i>	8
<i>Tabel 2.5 Perbandingan Penelitian Terdahulu Dengan Penelitian Ini</i>	8
<i>Tabel 2.6 Notasi Use Case Diagram</i>	17
<i>Tabel 2.7 Notasi Activity Diagram</i>	18
<i>Tabel 2.8 Notasi Sequence Diagram</i>	18
<i>Tabel 2.9 Notasi Class Diagram</i>	21
<i>Tabel 2.10 Notasi ERD</i>	22
<i>Tabel 3.1 Informasi Atribut Dataset Jamur</i>	26
<i>Tabel 3.2 Sampel Data Training Manual</i>	32
<i>Tabel 3.3 Data Testing Manual</i>	32
<i>Tabel 3.4 Probabilitas Atribut Data Testing</i>	34
<i>Tabel 3.5 Perhitungan Node 1</i>	36
<i>Tabel 3.6 Perhitungan Node 1.1</i>	37
<i>Tabel 3.7 Perhitungan Node 1.2</i>	38
<i>Tabel 3.8 Perhitungan Node 1.1.2</i>	38
<i>Tabel 3.9 Perhitungan Node 1.2.2</i>	39
<i>Tabel 3.10 Rancangan Tabel tb_admin</i>	60
<i>Tabel 3.11 Rancangan Tabel tb_atribut</i>	60
<i>Tabel 3.12 Rancangan Tabel tb_nilai</i>	60
<i>Tabel 3.13 Rancangan Tabel tb_dataset</i>	61

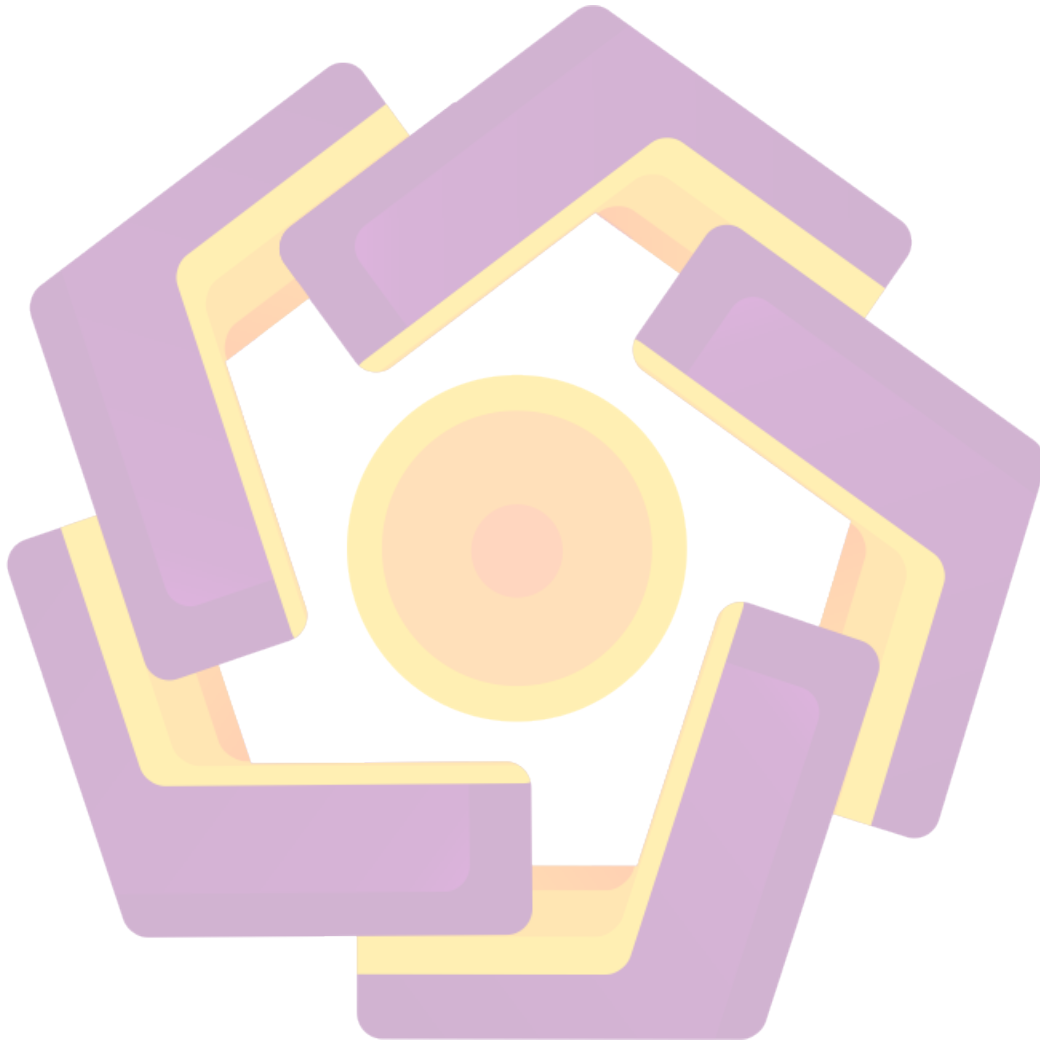
Tabel 3.14 Rancangan Tabel tb_testing 61

Tabel 4.1 Hasil Black Box Testing..... 76

Tabel 4.2 Hasil Uji Klasifikasi Oleh Sistem 81

Tabel 4.3 Hasil Akurasi Algoritma 84

Tabel 4.4 Hasil Durasi Waktu Perhitungan 85



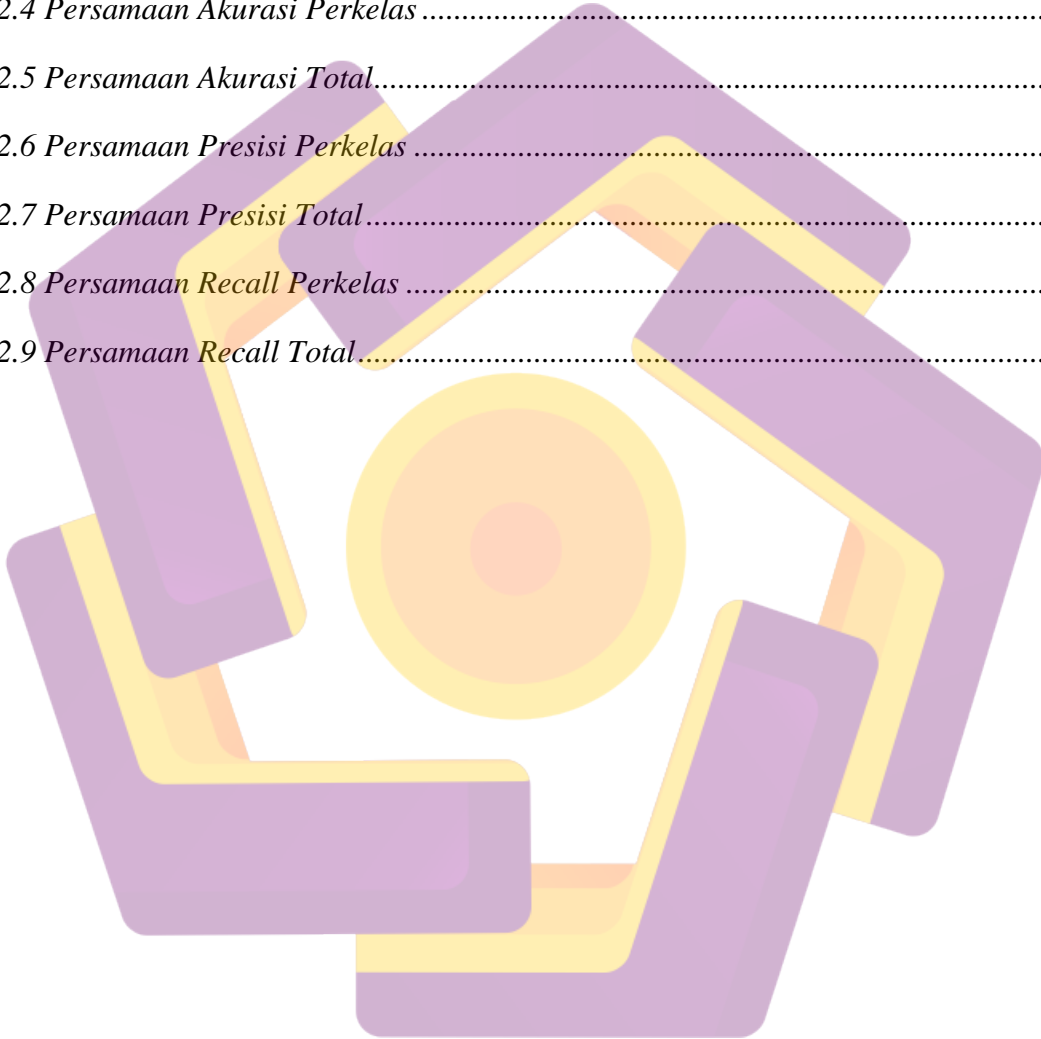
DAFTAR GAMBAR

<i>Gambar 3.1 Alur Penelitian.....</i>	25
<i>Gambar 3.2 Pohon Keputusan C4.5.....</i>	40
<i>Gambar 3.3 Tahapan Waterfall.....</i>	42
<i>Gambar 3.4 Rancangan Use Case Diagram</i>	44
<i>Gambar 3.5 Rancangan Activity Diagram Proses Login</i>	45
<i>Gambar 3.6 Rancangan Activity Diagram Proses Ganti Password.....</i>	46
<i>Gambar 3.7 Rancangan Activity Diagram Proses Mengelola Atribut.....</i>	47
<i>Gambar 3.8 Rancangan Activity Diagram Proses Mengelola Nilai Atribut</i>	48
<i>Gambar 3.9 Rancangan Activity Diagram Proses Mengelola Dataset.....</i>	49
<i>Gambar 3.10 Rancangan Activity Diagram Proses Mengelola Data Testing.....</i>	50
<i>Gambar 3.11 Rancangan Activity Diagram Proses Mengelola Proses Konsultasi</i>	51
<i>Gambar 3.12 Rancangan Activity Diagram Proses Perhitungan Akurasi.....</i>	52
<i>Gambar 3.13 Rancangan Sequence Diagram Login</i>	53
<i>Gambar 3.14 Rancangan Sequence Diagram Ganti Password.....</i>	53
<i>Gambar 3.15 Rancangan Sequence Diagram Pengelolaan Atribut.....</i>	54
<i>Gambar 3.16 Rancangan Sequence Diagram Pengelolaan Nilai Atribut.....</i>	54
<i>Gambar 3.17 Rancangan Sequence Diagram Pengelolaan Dataset.....</i>	55
<i>Gambar 3.18 Rancangan Sequence Diagram Pengelolaan Data Testing</i>	55
<i>Gambar 3.19 Rancangan Sequence Diagram Konsultasi</i>	56
<i>Gambar 3.20 Rancangan Sequence Diagram Akurasi.....</i>	56
<i>Gambar 3.21 Rancangan Class Diagram NBC.....</i>	57
<i>Gambar 3.22 Rancangan Class Diagram C4.5.....</i>	58
<i>Gambar 3.23 Rancangan ERD Chen.....</i>	59

<i>Gambar 3.24 Relasi Antar Tabel Database.....</i>	<i>59</i>
<i>Gambar 3.25 Rancangan Tampilan Tata Letak Dasar</i>	<i>62</i>
<i>Gambar 3.26 Rancangan Tampilan Halaman Konsultasi.....</i>	<i>63</i>
<i>Gambar 3.27 Rancangan Tampilan Halaman Login</i>	<i>63</i>
<i>Gambar 3.28 Rancangan Tampilan Halaman Pengelola Atribut dan Nilai</i>	<i>64</i>
<i>Gambar 3.29 Rancangan Tampilan Halaman Pengelola Dataset dan Data Testing</i>	<i>65</i>
<i>Gambar 3.30 Rancangan Tampilan Halaman Akurasi</i>	<i>65</i>
<i>Gambar 4.1 Implementasi Database NBC.....</i>	<i>67</i>
<i>Gambar 4.2 Implementasi Datase C4.5.....</i>	<i>67</i>
<i>Gambar 4.3 Implementasi Tampilan Halaman Konsultasi</i>	<i>74</i>
<i>Gambar 4.4 Implementasi Tampilan Halaman Login</i>	<i>74</i>
<i>Gambar 4.5 Implementasi Tampilan Halaman Atribut dan Nilai</i>	<i>75</i>
<i>Gambar 4.6 Implementasi Tampilan Halaman Dataset dan Data Testing</i>	<i>75</i>
<i>Gambar 4.7 Implementasi Tampilan Halaman Akurasi.....</i>	<i>76</i>
<i>Gambar 4.8 Uji Klasifikasi NBC Oleh Sistem</i>	<i>81</i>
<i>Gambar 4.9 Uji Klasifikasi C4.5 Oleh Sistem</i>	<i>81</i>
<i>Gambar 4.10 Uji Akurasi NBC Oleh Sistem.....</i>	<i>82</i>
<i>Gambar 4.11 Uji Akurasi C4.5 Oleh Sistem.....</i>	<i>83</i>

DAFTAR RUMUS

<i>Rumus 2.1 Persamaan NBC.....</i>	13
<i>Rumus 2.2 Persamaan Gain C4.5.....</i>	14
<i>Rumus 2.3 Persamaan Entropy C4.5.....</i>	14
<i>Rumus 2.4 Persamaan Akurasi Perkelas</i>	23
<i>Rumus 2.5 Persamaan Akurasi Total.....</i>	23
<i>Rumus 2.6 Persamaan Presisi Perkelas</i>	24
<i>Rumus 2.7 Persamaan Presisi Total.....</i>	24
<i>Rumus 2.8 Persamaan Recall Perkelas</i>	24
<i>Rumus 2.9 Persamaan Recall Total.....</i>	24



INTISARI

Jenis jamur sangat beragam dan cukup banyak dimanfaatkan meskipun terdapat beberapa jamur yang beracun. Sehingga terdapat salah satu hasil penelitian klasifikasi jenis jamur dalam buku berjudul *The Audubon Society Field Guide to North American Mushrooms* yang kemudian dibuat ke dalam bentuk *dataset* untuk digunakan dalam penelitian *data mining*. Karena perkembangan penelitian mengenai *data mining* membuat algoritma *data mining* banyak ditemukan, diantaranya adalah *Naïve Bayes Classifier* dan C4.5. Keduanya dikenal memiliki akurasi yang baik dan banyak diimplementasikan. Maka perlu diketahui algoritma yang lebih baik diantara keduanya dalam kasus klasifikasi data jamur beracun dan tidak beracun.

Penelitian ini menggunakan metode *library research* atau penelitian kepustakaan untuk mengetahui berbagai informasi yang diperlukan dalam penelitian dan pengumpulan data yang digunakan. Adapun data yang digunakan adalah *Mushroom Dataset* yang diunduh dari *UCI Machine Learning Repository*.

Tiga aspek pengukuran yang digunakan sebagai parameter penentu algoritma yang lebih baik yaitu hasil klasifikasi, akurasi, dan total durasi waktu proses. Berdasarkan uji perbandingan yang dilakukan didapatkan hasil klasifikasi yang benar dari kedua algoritma, dengan hasil akurasi *Naïve Bayes Classifier* 99,718% dan C4.5 100%, serta total durasi waktu *Naïve Bayes Classifier* selama 182,4266338 detik, dan C4.5 selama 104,291364 detik. Sehingga dapat diketahui bahwa algoritma C4.5 lebih baik dibandingkan dengan algoritma *Naïve Bayes Classifier* dalam klasifikasi data jamur beracun dan tidak beracun.

Kata kunci : Perbandingan, *Naïve Bayes Classifier*, C4.5, Dataset Jamur, Data Mining

ABSTRACT

Types of mushrooms are very diverse and quite widely used even though there are some mushrooms that are poisonous. So that there is one of the results of research on the classification of mushroom species in a book entitled 'The Audubon Society Field Guide to North American Mushrooms' which is then made into a dataset for use in data mining research. Due to the development of research on data mining, many data mining algorithms have been found, including the Naïve Bayes Classifier and C4.5 that are known to have good accuracy and widely implemented. So it is necessary to know which algorithm is better between them in classification of poisonous and edible mushroom data.

This research uses library research method to find out a variety of information needed in the research and data collection used. The data used is the 'Mushroom Dataset' which is downloaded from the UCI Machine Learning Repository.

Three measurement aspects are used as parameters to determine which algorithm is better, namely the classification result, accuracy, and total processing time duration. Based on the comparison test conducted, the correct classification results were obtained from the two algorithms, with the accuracy of the Naïve Bayes Classifier 99.718 %% and C4.5 100%, and the total duration of the Naïve Bayes Classifier time of 182.4266338 seconds, and C4.5 for 104, 291364 seconds. So it can be seen that the C4.5 algorithm is better than the Naïve Bayes Classifier algorithm in the classification of poisonous and edible mushroom data.

Keyword : Comparison, Naïve Bayes Classifier, C4.5, Mushroom Dataset, Data Mining