

# Perbandingan Nilai Akurasi dari Algoritma ID3 dan Algoritma C4.5 Terhadap

## Beberapa Dataset

*Diajukan Untuk Menyusun Tugas Akhir  
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh:

**KORNELIUS FIRSTDLY**

09121402044

**Jurusan Teknik Informatika**

**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

**2019**

# LEMBAR PENGESAHAN USULAN TUGAS AKHIR

## Perbandingan Nilai Akurasi Dari Algoritma ID3 dan Algoritma C4.5 Pada Beberapa Dataset

Oleh :

KORNELIUS FIRSTDLY  
09121402044

Palembang,

2019

Pembimbing I,



Samsuryadi, S.Si., M.Kom., Ph.D.  
NIP 197102041997021003

Pembimbing II,



Dargus Matthew Saputra, S.T., M.Sc.  
NIP 198505102015041002

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Prunmaha, S.T., M.T.  
NIP 197711012009121004

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 17 Desember 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Kornelius Firstdly  
NIM : 09121402044  
Judul : Perbandingan Nilai Akurasi Algoritma ID3 dan Algoritma C4.5 Terhadap Beberapa Dataset

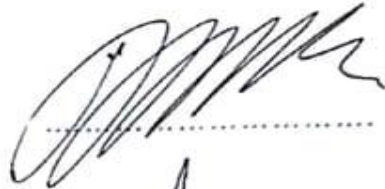
1. Ketua Penguji

Samsurvadi, M.Kom, Ph.D  
NIP. 197102041997021003



2. Sekretaris

Dannv Matthew Saputra, M.Cs  
NIP. 198505102015041002



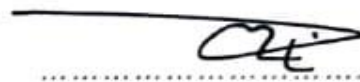
3. Penguji I

Yunita, M.Sc  
NIP. 198306062015042002



4. Penguji II

Osvari Arsalan, M.T  
NIP. 198806282018031001



Mengetahui,  
Ketua Jurusan Teknik Informatika

  
Rifkie Primartha, M.T.  
NIP. 197706042009121004

## HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Kornelius Firstdly  
NIM : 09121402044  
Program Studi : Teknik Informatika (Bilingual)  
Judul Skripsi : Perbandingan Nilai Akurasi Algoritma ID3 dan Algoritma C4.5  
Terhadap Beberapa Dataset  
Hasil Pengecekan Software *iThenticate/Turnitin* : 10 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 16 Januari

2019



Kornelius Firstdly  
NIM. 09121402044

## Motto

- *Tak perlu terang, cukup ada dan tak kunjung padam*  
(pinterest).
- *When you want something, all the universe conspires in helping you to achieve it* (Paulo Coelho, *The Alchemist*).
- *Everyone seems to have a clear idea of how other people should lead their lives, but none about his or her own* (*The Alchemist*).
- *Never stop dreaming and don't forget to wake up* (@normnd.akbr).
- *The secret of life, though, is to fall seven times and get up eight* (*The Alchemist*).
- *Usaha dan Hasil ternyata bersahabat baik* (@box.coklat).

### Kupersembahkan karya tulis ini kepada :

- Tuhan
- Kedua Orangtua saya
- Dosen – dosen pengajar.
- Semua pembimbing termasuk pembimbing akademik, pembimbing penelitian, pembimbing lainnya baik dalam segi pemberkasan maupun motivasi.

# PERBANDINGAN NILAI AKURASI ALGORITMA ID3 DAN ALGORITMA C4.5 TERHADAP BEBERAPA DATASET

Oleh

Kornelius Firstdly  
09121402044

## ABSTRAK

Algoritma ID3 memiliki keunggulan dalam melakukan klasifikasi serta kemampuan pembelajaran yang baik, sedangkan algoritma C4.5 dalam hal pengolahan data numerik secara kontinu, pengolahan data kategori (diskret), dan menangani nilai atribut yang hilang. Perbedaan kemampuan tersebut menjadi pembuktian akurasi tertinggi bagi masyarakat antara algoritma ID3 dan algoritma C4.5, dengan nilai akurasi yang didapat dari WEKA. Penelitian ini menggunakan 5 macam *dataset*, yaitu *dataset breast-cancer*, *dataset cars*, *dataset vote*, *dataset audiology*, dan *dataset soybeans*. Hasil penelitian menunjukkan bahwa akurasi tertinggi didapatkan oleh algoritma ID3 pada 3 *dataset*, yaitu *dataset breast-cancer* (88.83%), *dataset cars* (96.63%), *dataset vote* (97.99%), sedangkan akurasi tertinggi yang didapat algoritma C4.5 pada 2 *dataset*, yaitu *dataset audiology* (90.66%) dan *dataset soybeans* (97.16%).

**Kata Kunci:** *Data Mining*, *Decision Tree*, Klasifikasi, Algoritma ID3, Algoritma C4.5

# ACCURACY COMPARISON ID3 ALGORITHM AND C4.5 ALGORITHM IN SEVERAL DIFFERENT DATASETS

By

Kornelius Firstdly  
09121402044

## ABSTRACT

ID3 algorithm has advantage in classify and learning skill well, and C4.5 algorithm has advantage in processing numeric data continuous, processing categorical data, and handling missing attribute value. Different skill from two algorithms are being based to prove the best result between ID3 algorithm and C4.5 algorithms. This research used 5 different datasets : breast-cancer dataset, cars dataset, vote dataset, audiology dataset, and soybeans dataset. The result showed that ID3 algorithm accuracy is higher than C4.5 algorithm accuracy in 3 datasets : *dataset breast-cancer (88.83%), dataset cars (96.63%), dataset vote (97.99%)*, and highest accuracy C4.5 algorithm in 2 datasets : *dataset audiology (90.66%) dan dataset soybeans (97.16%)*.

**Keywords** : *Data Mining, Decision Tree, Classification, ID3 Algorithm, C4.5 Algorithm*

## KATA PENGANTAR

Puji Tuhan atas segala nikmat, rahmat, dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir dengan judul “**Perbandingan Nilai Akurasi Algoritma ID3 dan Algoritma C4.5 Terhadap Beberapa Dataset**”. Tugas akhir ini disusun sebagai salah satu syarat kelulusan tingkat sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari bahwa dalam penulisan tugas akhir ini tidak terlepas dari dukungan, bimbingan, kerja sama, dan motivasi dari berbagai pihak. Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian tugas akhir ini, yaitu:

1. Keluarga besar yang selalu mendoakan dan memberikan dukungan baik secara langsung maupun tidak langsung;
2. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya sekaligus sebagai Pembimbing Akademik bagi penulis selama masa perkuliahan;
3. Bapak Samsuryadi, M.Kom, Ph.D dan Bapak Danny Matthew Saputra, M.Cs selaku dosen pembimbing tugas akhir yang telah memberikan bimbingan, ilmu pengetahuan, dukungan, kritik dan saran yang membangun kepada penulis dalam proses penyusunan tugas akhir ini;
4. Ibu Yunita, M.Sc. dan Bapak Osvari Arsalam, M.T. selaku penguji yang telah memberikan kritik dan saran kepada penulis dalam penyempurnaan tugas akhir ini;
5. Mbak Titi beserta penerus tahtanya sebagai admin jurusan Teknik Informatika Bilingual yakni mbak Wiwin yang telah banyak membantu dan membimbing penulis dalam segala urusan administrasi juga dalam hal penjadwalan ujian yang luar biasa sekali;
6. Teman-teman seperjuangan mulai dari jurusan IF, SI serta SK angkatan 2012 hingga 2014 yang telah memberikan motivasi, inspirasi, kerja sama, dan



bantuan selama proses kegiatan akademik berlangsung dimana namanya tidak bisa disebutkan satu-persatu disini;

7. Mereka yang telah menjadi *support* baik itu dalam segi strategi bertemu dengan pembimbing, strategi revisi, pembimbing berkas serta yang sama-sama berjuang dalam memperoleh gelar sarjananya.
8. Seluruh pihak yang turut membantu dalam penyusunan dan penyempurnaan tugas akhir ini yang tidak dapat penulis sebutkan satu-persatu.

Palembang, Januari 2019

Penulis.

## DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
LEMBAR PENGESAHAN TUGAS AKHIR .....	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR .....	iii
HALAMAN PERNYATAAN	
MOTTO DAN PERSEMBAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	ix
DAFTAR GAMBAR .....	xiii
DAFTAR TABEL .....	xvi
DAFTAR RUMUS .....	xvii
<b>BAB I PENDAHULUAN</b>	
1.1 Pendahuluan .....	I-1
1.2 Latar Belakang .....	I-1
1.3 Perumusan Masalah .....	I-3
1.4 Tujuan Penelitian .....	I-3
1.5 Manfaat Penelitian .....	I-3
1.6 Batasan Masalah .....	I-4
1.7 Metodologi Penelitian .....	I-4
1.8 Sistematika Penulisan.....	I-6
1.9 Kesimpulan .....	I-7
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Pendahuluan .....	II-1
2.2 Penelitian Terkait .....	II-1
2.3 Data Mining .....	II-4
2.3.1 Classification.....	II-5
2.3.2 Clustering .....	II-5

2.4	Algoritma Decision Tree .....	II-6
2.5	Algoritma ID3 .....	II-6
2.6	Algoritma C4.5.....	II-8
2.7	WEKA .....	II-9
2.8	Rational Unified Process (RUP) .....	II-10
2.9	Kesimpulan .....	II-16

### **BAB III METODOLOGI PENELITIAN**

3.1	Pendahuluan .....	III-1
3.2	Unit Penelitian .....	III-1
3.3	Metode Pengumpulan Data .....	III-1
3.4	Tahapan Penelitian .....	III-2
3.4.1	Menentukan Ruang Lingkup dan Unit Penelitian .....	III-3
3.4.2	Menemukan Dasar Teori yang Berkaitan dengan Penelitian .....	III-3
3.4.3	Menetapkan Kriteria Pengujian .....	III-3
3.4.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian .....	III-5
3.4.5	Melakukan Pengujian Penelitian .....	III-5
3.5	Metode Pengembangan Perangkat Lunak .....	III-6
3.5.1	Fase Insepsi .....	III-6
3.5.2	Fase Elaborasi.....	III-6
3.5.3	Fase Konstruksi .....	III-7
3.5.4	Fase Transisi .....	III-7
3.6	Penjadwalan Penelitian .....	III-8
3.7	Kesimpulan .....	III-10

## **BAB IV PENGEMBANGAN PERANGKAT LUNAK**

4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Permodelan Bisnis .....	IV-1
4.2.2	Kebutuhan Sistem.....	IV-2
4.3	Analisis dan Desain.....	IV-5
4.3.1	Analisis Masalah .....	IV-5
4.3.2	Analisis Perangkat Lunak.....	IV-6
4.3.3	Analisis Kebutuhan Perangkat Lunak .....	IV-6
4.3.4	Analisis Data .....	IV-6
4.3.5	Desain Perangkat Lunak.....	IV-10
4.4	Diagram Kelas Analisis.....	IV-22
4.5	Diagram Sekuensial .....	IV-26
4.6	Diagram Kelas.....	IV-30
4.7	Desain Antarmuka .....	IV-31
4.8	Implementasi Kelas .....	IV-32
4.9	Implementasi Antarmuka .....	IV-36
4.10	Pengujian Perangkat Lunak.....	IV-38
4.11	Rencana Pengujian.....	IV-39
4.12	Kasus Uji.....	IV-41
4.13	Hasil Pengujian Use Case .....	IV-45
4.13.1	Hasil Pengujian Melakukan Perhitungan dengan Algoritma ID3 .....	IV-45
4.13.2	Hasil Pengujian Melakukan Perhitungan dengan Algoritma C4.5 .....	IV-50
4.14	Kesimpulan .....	IV-60

**BAB V HASIL DAN ANALISIS PENELITIAN**

5.1 Pendahuluan .....	V-1
5.2 Hasil Pengujian Pencarian Nilai Akurasi dengan Menggunakan Algoritma ID3 dan Algoritma C4.5 .....	V-1
5.3 Analisis Hasil Pengujian .....	V-7
5.4 Kesimpulan.....	V-8

**BAB VI KESIMPULAN DAN SARAN**

6.1 Kesimpulan.....	VI-1
6.2 Saran.....	VI-2

**DAFTAR PUSTAKA**

**LAMPIRAN**

## DAFTAR GAMBAR

	Halaman
2-1. Gambar Struktur Proses RUP .....	II-11
4-1 Sample Data dari Dataset Breast Cancer .....	IV-7
4-2 Sample Data dari Dataset Soybeans .....	IV-7
4-3 Sample Data dari Dataset Vote.....	IV-8
4-4 Sample Data dari Dataset Audiology .....	IV-9
4-5 Sample Data dari Dataset Cars .....	IV-10
4-6 Diagram Use Case Perbandingan Nilai Akurasi Algoritma ID3 dan Algoritma C4.5 Terhadap Beberapa Dataset .....	IV-11
4-7 Activity Diagram Melakukan Perhitungan dengan Algoritma ID3....	IV-20
4-8 Activity Diagram Melakukan Perhitungan dengan Algoritma C4.5 ..	IV-21
4-9 Diagram Kelas Analisis Melakukan Perhitungan dengan Algoritma ID3 .....	IV-23
4-10 Diagram Kelas Analisis Melakukan Perhitungan dengan Algoritma C4.5.....	IV-24
4-11 Diagram Kelas Analisis Memilih Data Input .....	IV-25
4-12 Diagram Sekuensial Melakukan Perhitungan dengan Algoritma ID3	IV-26
4-13 SSD ID3.....	IV-27
4-14 Diagram Sekuensial Melakukan Perhitungan dengan Algoritma C4.5	IV-28
4-15 SSD C4.5 .....	IV-29
4-16 Diagram Kelas Perangkat Lunak .....	IV-30
4-17 Desain Antarmuka .....	IV-31
4-18 Desain Antarmuka Melakukan Perhitungan dengan Algoritma ID3 .....	IV-31
4-19 Desain Antarmuka Melakukan Perhitungan dengan Algoritma C4.5.....	IV-32
4-20 Implementasi Antarmuka Main Page .....	IV-36
4-21 Implementasi Antarmuka Melakukan Perhitungan dengan Algoritma ID3.....	IV-37

4-22	Implementasi Antarmuka Melakukan Perhitungan dengan Algoritma C4.5 .....	IV-38
4-23	Form Input Data Training Algoritma ID3 .....	IV-46
4-24	Alert Data Training Berhasil Masuk ke Algoritma ID3 .....	IV-46
4-25	Decision Tree yang Tercipta Pada Fase Training Algoritma ID3 .....	IV-47
4-26	Rules yang Dihasilkan Pada Fase Training Algoritma ID3 .....	IV-47
4-27	Form Input Data Testing Algoritma ID3 .....	IV-48
4-28	Alert Data Testing Berhasil Masuk ke Algoritma ID3 .....	IV-48
4-29	Prediksi yang Dihasilkan Pada Fase Testing Algoritma ID3 .....	IV-49
4-30	Akurasi yang Dihasilkan Pada Fase Testing Algoritma ID3 .....	IV-49
4-31	Tombol Done yang Mengembalikan Form Algoritma ID3 ke Form Main Page .....	IV-50
4-32	Form Input Data Training Algoritma C4.5 .....	IV-51
4-33	Alert Data Training Berhasil Masuk ke Algoritma C4.5 .....	IV-51
4-34	Decision Tree yang Tercipta Pada Fase Training Algoritma C4.5 .....	IV-52
4-35	Rules yang Dihasilkan Pada Fase Training Algoritma C4.5 .....	IV-52
4-36	Form Input Data Testing Algoritma C4.5 .....	IV-53
4-37	Alert Data Testing Berhasil Masuk ke Algoritma C4.5 .....	IV-53
4-38	Prediksi yang Dihasilkan Pada Fase Testing Algoritma C4.5 .....	IV-54
4-39	Akurasi yang Dihasilkan Pada Fase Testing Algoritma C4.5 .....	IV-54
4-40	Tombol Done yang Mengembalikan Form Algoritma C4.5 ke Form Main Page .....	IV-55
4-41	Perubahan File Type CSV ke Type ARFF Pada Dataset Audiology dan Dataset Cars .....	IV-55
4-42	Open File Pada Perubahan File Type CSV ke Type ARFF Pada Dataset Audiology dan Dataset Cars .....	IV-56
4-43	File CSV Pada Perubahan File Type CSV ke Type ARFF Pada Dataset Audiology dan Dataset Cars .....	IV-56
4-44	Save File ARFF Pada Perubahan File Type CSV ke Type ARFF Pada Dataset Audiology dan Dataset Cars .....	IV-57
4-45	Pilih Explorer Untuk Memulai Pengujian WEKA .....	IV-57

4-46	Open File ARFF Untuk Mulai Pengujian WEKA.....	IV-58
4-47	File ARFF Untuk Mulai Pengujian WEKA .....	IV-58
4-48	Penggunaan Fitur ReplaceMissingValue Untuk Mulai Pengujian WEKA .....	IV-59
4-49	Pemilihan Pengujian ID3/J48 (C4.5 WEKA) Untuk Mulai Pengujian WEKA .....	IV-59
4-50	Pengaturan Untuk Mulai Pengujian WEKA dan Hasil Akurasi WEKA .....	IV-60



## DAFTAR TABEL

	Halaman
3-1. WBS Perbandingan Nilai Akurasi ID3 dan Algoritma C4.5 Terhadap Beberapa Dataset .....	III-9
4-1 Kebutuhan Fungsional Perangkat Lunak.....	IV-3
4-2 Kebutuhan Non Fungsional Perangkat Lunak.....	IV-4
4-3 Definisi Aktor .....	IV-12
4-4 Definisi Use Case .....	IV-12
4-5 Skenario Use Case Melakukan Perhitungan dengan Algoritma ID3 .	IV-13
4-6 Skenario Use Case Melakukan Perhitungan dengan Algoritma C4.5	IV-15
4-7 Skenario Use Case Memilih Data Input .....	IV-17
4-8 Tabel Implementasi Kelas .....	IV-33
4-9 Rencana Pengujian Use Case Melakukan Perhitungan dengan Algoritma ID3.....	IV-39
4-10 Rencana Pengujian Use Case Melakukan Perhitungan dengan Algoritma C4.5 .....	IV-40
4-11 Pengujian Use case Melakukan Perhitungan dengan Algoritma ID3.	IV-41
4-12 Pengujian Use case Melakukan Perhitungan dengan Algoritma C4.5	IV-43
5-1 Tabel Hasil Pengujian Pada Dataset Audiology .....	V-2
5-2 Tabel Hasil Pengujian Pada Dataset Breast-cancer .....	V-3
5-3 Tabel Hasil Pengujian Pada Dataset Cars.....	V-4
5-4 Tabel Hasil Pengujian Pada Dataset Soybeans.....	V-5
5-5 Tabel Hasil Pengujian Pada Dataset Vote .....	V-6
5-6 Tabel Hasil Pengujian Perbandingan Program Penelitian dengan Aplikasi WEKA.....	V-7

## DAFTAR RUMUS

	Halaman
II-1. Perumusan <i>Information Gain</i> .....	II-7
II-2. Perumusan <i>Entropy</i> .....	II-7
II-3. Perumusan <i>Gain Ratio</i> .....	II-8
II-4. Perumusan <i>SplitInfo</i> .....	II-9

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab I memberikan penjelasan umum mengenai keseluruhan penelitian. Antara lain membahas latar belakang masalah penelitian, perumusan masalah penelitian, tujuan penelitian dan manfaat penelitian. Pendahuluan dimulai dengan penjelasan mengenai pentingnya pemilihan data serta cara pengolahan data yang tepat dan penerapan algoritma klasifikasi pada proses pengolahan data yang menjadi latar belakang dari penelitian ini.

### **1.2 Latar Belakang Masalah**

Data memiliki ciri khas masing-masing. Terdapat data yang bila dikelompokkan akan didapat sebuah pola, kesimpulan tersendiri, namun ketika salah mengolahnya akan memberikan hasil yang kurang tepat. Ketepatan dalam mengolah data harus diperhatikan oleh orang yang mengolah data tersebut, dikarenakan pengolahan data sangat berpengaruh kepada hasil yang akan didapat, maka dari itu, pengolahan data tidak bisa dipandang sebelah mata dengan hanya dipikir sebuah pengurutan saja.

Beberapa model pengolahan data yang telah dikembangkan dapat membantu dalam pengolahan data, yaitu algoritma IDS, ID3, C4.5, CHAID, dan CART (Defiyanti *et al.*, 2010). Peneliti memilih 2 model pengolahan data, yaitu

algoritma ID3 dan algoritma C4.5 sebagai metode dalam penelitian karena terdapat perbedaan pernyataan dari Motie *et al* (2012) dan Defiyanti *et al* (2010).

Algoritma ID3 atau *Iterative Dichotomizer 3* merupakan sebuah algoritma yang berasal dari algoritma *decision tree*, yaitu algoritma yang biasanya digunakan orang banyak untuk memperoleh informasi yang berhubungan dengan pengambilan suatu keputusan dengan metode klasifikasi (Peng *et al.*, 2009).

Algoritma ID3 memiliki kecepatan dalam pengklasifikasian dan kemampuan pembelajaran yang sangat baik, serta pemodelan yang sederhana (Defiyanti *et al.*, 2010). Algoritma ID3 juga cocok jika mengolah data dengan atribut pengklasifikasian sedikit (Yuxun *et al.*, 2010).

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang banyak dipakai dan disukai karena memiliki kelebihan-kelebihan, yaitu dapat mengolah data numerik (*kontinyu*) dan kategori (diskret), serta dapat menangani nilai atribut yang hilang, dan menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat antara algoritma *Neural Network* (NN), algoritma *Support Vector Machine* (SVM), dan algoritma *Naïve Bayesian Classifier* (NB) (Youn *et al.*, 2006).

Oleh karena adanya perbedaan pernyataan dari Motie *et al* (2012) dan Defiyanti *et al* (2010), maka peneliti akan melakukan penelitian untuk membuktikan kinerja yang lebih baik antara algoritma ID3 dan algoritma C4.5.

### 1.3 Perumusan Masalah

Penelitian ini membuktikan tingkat akurasi dari algoritma ID3 dan algoritma C4.5 berdasarkan pada penelitian yang dilakukan Motie *et al* (2012) bahwa tingkat akurasi algoritma C4.5 lebih tinggi daripada algoritma ID3. Namun, berdasarkan Defiyanti *et al* (2010) bahwa kinerja algoritma ID3 lebih tinggi daripada algoritma C4.5. Oleh karena itu, penelitian ini bertujuan untuk membuktikan tingkat akurasi terbaik dari algoritma ID3 dan algoritma C4.5.

### 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk :

1. Membandingkan nilai akurasi program penelitian dengan nilai akurasi WEKA pada algoritma ID3 dan algoritma C4.5.
2. Mengetahui nilai akurasi yang tertinggi dari algoritma ID3 dan algoritma C4.5 dalam pengolahan beberapa *dataset*.

### 1.5 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut :

1. Memudahkan bagi pemula atau orang awam yang ingin mempelajari algoritma ID3 dan algoritma C4.5.
2. Perangkat lunak dapat digunakan untuk membantu dalam menentukan perhitungan algoritma ID3 dan algoritma C4.5 terhadap beberapa *dataset*.
3. Menunjukkan tingkat akurasi tertinggi antara algoritma ID3 dan algoritma C4.5 berdasarkan pengujian data di beberapa *dataset* yang ada.

## 1.6 Batasan Masalah

Penelitian ini terdapat beberapa batasan masalah, yaitu sebagai berikut :

1. Penelitian ini hanya terfokus kepada tingkat akurasi untuk perbandingan algoritma ID3 dan algoritma C4.5.
2. Penelitian ini hanya menggunakan 5 *dataset* yang diambil dari WEKA dan UCI Dataset, yaitu dataset 1 (breast-cancer) sebanyak 286 data, dataset 2 (audiology) sebanyak 200 data, dataset 3 (cars) sebanyak 1728 data, dataset 4 (soybean) sebanyak 683 data, dan dataset 5 (vote) sebanyak 435 data.

## 1.7 Metodologi Penelitian

Tahapan-tahapan yang akan dilakukan pada pelaksanaan skripsi ini adalah sebagai berikut :

### 1. Studi Literatur

Studi literatur dilakukan dengan mengumpulkan bahan-bahan referensi baik dari buku, artikel, paper, jurnal, makalah, maupun situs-situs internet. Studi literatur yang dilakukan terkait dengan perbandingan algoritma ID3 dan algoritma C4.5.

### 2. Analisis Permasalahan

Pada tahap ini dilakukan analisis terhadap hasil studi literatur untuk mengetahui dan mendapatkan pemahaman mengenai cara kerja algoritma ID3 dan algoritma C4.5 yang diujikan pada beberapa dataset untuk mendapatkan nilai akurasi tertinggi diantara kedua algoritma tersebut.

### 3. Perancangan Sistem

Pada tahap ini dilakukan perancangan arsitektur, perancangan data dan perancangan antarmuka. Proses perancangan dilakukan berdasarkan hasil analisis studi literatur yang telah didapatkan.

### 4. Implementasi Sistem

Pada tahap ini dilakukan proses implementasi pengkodean program dalam aplikasi komputer menggunakan bahasa pemrograman yang telah dipilih yang sesuai dengan analisis dan perancangan yang sudah dilakukan.

### 5. Pengujian

Pada tahap ini dilakukan proses pengujian dan percobaan terhadap sistem sesuai dengan kebutuhan yang ditentukan sebelumnya serta memastikan program yang dibuat berjalan seperti yang diharapkan.

### 6. Penyusunan Laporan

Pada tahap ini dilakukan dokumentasi hasil analisis dan implementasi dari algoritma ID3 dan algoritma C4.5 terhadap pencarian akurasi tertinggi dari beberapa dataset, serta membuat kesimpulan dan saran pada hasil akhir penelitian.

## **1.8 Sistematika Penulisan**

Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut:

### **1. BAB I Pendahuluan**

Bab ke-1 berisi tentang penjelasan latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, metodologi penelitian, dan sistematika penulisan.

### **2. BAB II Tinjauan Pustaka**

Bab ke-2 berisi landasan dasar teori yang akan digunakan dalam melakukan analisis, perancangan, dan implementasi tugas akhir yang dilakukan pada bab-bab selanjutnya.

### **3. BAB III Metodologi Penelitian**

Pada Bab-3 berisi tentang tahapan yang akan dilaksanakan pada penelitian ini. Setiap rencana tahapan penelitian akan dijelaskan secara rinci dengan mengacu pada suatu kerangka kerja. Pada akhir Bab-3 berisi perancangan manajemen proyek pada pelaksanaan penelitian.

### **4. BAB IV Pengembangan Perangkat Lunak**

Bab ke-4 membahas mengenai perancangan dan lingkungan implementasi algoritma ID3, implementasi algoritma C45, pengujian terhadap dataset yang memiliki data yang berbeda, hasil pengujian berupa akurasi algoritma ID3, dan akurasi algoritma C45.



## **5. BAB V Hasil dan Analisis Penelitian**

Bab ke-5 berisi hasil pengujian yang berdasarkan tahapan yang telah direncanakan sebelumnya. Analisis yang diberikan oleh penelitian akan menjadi dasar dari kesimpulan yang akan diambil dalam penelitian ini.

## **6. BAB VI Kesimpulan dan Saran**

Bab ke-6 berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam implementasi algoritma ID3 dan algoritma C4.5 dalam pencarian akurasi tertinggi dari beberapa dataset, serta sebagai pembuktian tingkat akurasi tertinggi antara algoritma ID3 dan algoritma C4.5 dari setiap dataset.

### **1.9 Kesimpulan**

Penelitian ini akan berfokus untuk melihat ketepatan keputusan yang dihasilkan pada pengujian data dari beberapa dataset dengan menggunakan beberapa metode klasifikasi. Metode – metode klasifikasi yang digunakan antara lain algoritma ID3 dan algoritma C45. Penelitian ini ditujukan untuk membuktikan algoritma yang terbaik untuk menangani banyak data yang berbeda dari beberapa dataset dengan melihat akurasi yang dihasilkan oleh kedua algoritma yang digunakan.

**DAFTAR PUSTAKA**

- Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. (2013). Predicting Students' Performance Using Id3 and C4.5 Classification Algorithms. *International Journal of Data Mining and Knowledge Management Process*, 3(5), 39–52. <https://doi.org/10.5121/ijdkp.2013.3504>
- Amalia, H. & Evicienna (2013). Algoritma C45 Untuk Prediksi Hasil Pemilihan Legislatif DPRD DKI Jakarta.
- Azwanti, N. (2018). Algoritma C45 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu). *Simetris*, Vol 9, No 1. ISSN : 2252-4983.
- Defiyanti, S., & Pardede, C. D. . (2010). Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi Spam-Mail.
- Han, J. (2006). Data mining : concepts and techniques. *Journal of Chemical Information and Modeling*, 101–103. <https://doi.org/10.1017/CBO9781107415324.004>
- Hernandez, M., Hildrum, K., Jain, P., Wagle, R., Alexe, B., Krishnamurthy, R., ... Venkatramani, C. (2013). Constructing Consumer Profiles from Social Media Data.
- Hoyyi, A., Inspiranti, D., Tyasti, A. E. (2015). Algoritma Iterative Dichotomiser (ID3) Untuk Mengidentifikasi Data Rekam Medis. *Jurnal Gaussian*, Vol 4, No 2, 237-246.

- Khoonsari, P. E., & Motie, A. (2012). A Comparison of Efficiency and Robustness of ID3 and C4 . 5 Algorithms Using Dynamic Test and Training Data Sets, *2*(5), 540–543.
- Kruchten, P. (2003). *The Rational Unified Process: An Introduction*. Addison-Wesley Longman Publishing Co., Inc, Boston, USA.
- Mustapha, N., Nuruddin, A. A. B., Sitanggang, I. M., Yaakob, R. (2011). An Extended ID3 Decision Tree Algorithm for Spatial Data, *6*.
- Peng, W., Chen, J., & Zhou, H. (2009). An Implementation of IDE3 Decision Tree Learning Algorithm. *Machine Learning, 1*, 1–20.
- Youn, S., & McLeod, D. (2006). Spam email classification using an adaptive ontology. *Journal of Software, 2*(3), 43–55.  
<https://doi.org/10.4304/jsw.2.3.43-55>