

**ANALISIS KOMPARASI ALGORITMA
ID3 DAN C4.5 DALAM KLASIFIKASI PENERIMAAN BEASISWA
(STUDI KASUS: BEASISWA SOLOPEDULI TAHUN 2017/2018)**

Skripsi
untuk memenuhi sebagai persyaratan
mencapai derajat Sarjana S-1
Program Studi Teknik Informatika



Disusun oleh:
Nadia Sholihah
16650014

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2020



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-953/Un.02/DST/PP.00.9/04/2020

Tugas Akhir dengan judul : ANALISIS KOMPARASI ALGORITMA ID3 DAN C 4.5 DALAM KLASIFIKASI
PENERIMAAN BEASISWA (STUDI KASUS : BEASISWA SOLOPEDULI TAHUN
2017/2018)

yang dipersiapkan dan disusun oleh:

Nama : NADIA SHOLIAH
Nomor Induk Mahasiswa : 16650014
Telah diujikan pada : Kamis, 02 April 2020
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang/Penguji I

Rahmat Hidayat, S.Kom., M.Cs.
NIP. 19850514 201503 1 002

Penguji II

Dr. Shofwatul 'Uyun, S.T., M.Kom.
NIP. 19820511 200604 2 002

Penguji III

Dr. Agung Fatwanto, S.Si., M.Kom.
NIP. 19770103 200501 1 003

Yogyakarta, 02 April 2020
UIN Sunan Kalijaga
Fakultas Sains dan Teknologi
Dekan



Dr. Murtono, M.Si.
NIP. 19591212 200003 1 001



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi saudara:

Nama : Nadia Sholihah

NIM : 16650014

Judul Skripsi : "Analisis Komparasi Algoritma *ID3* dan *C4.5* Dalam Klasifikasi Penerimaan Beasiswa (Studi Kasus : Beasiswa Solopeduli tahun 2017/2018)"


sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 18 Maret 2020
Pembimbing


Rahmat Hidayat, S.Kom., M.Cs.
NIP. 19850514 201503 1 002

SURAT PERSETUJUAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini :

Nama : Nadia Sholihah

NIM : 16650014

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul “Analisis Komparasi Algoritma *ID3* dan *C4.5* Dalam Klasifikasi Penerimaan Beasiswa (Studi Kasus : Beasiswa Solopeduui tahun 2017/2018)” merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 19 Maret 2020

STATE ISLAMIC UNIVERSITY
SUNAN KALIDIGRA
YOGYAKARTA



Nadia Sholihah
NIM.16650014

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penelitian yang dilakukan dengan judul **Analisis Komparasi Algoritma ID3 dan C4.5 Dalam Klasifikasi Penerimaan Beasiswa (Studi Kasus: Beasiswa Solopeduli)** dapat berjalan dengan baik. Sholawat serta salam semoga tetap tercurahkan kepada junjungan kita Nabi Muhammad SAW.

Tidak lupa penulis juga mengucapkan terima kasih kepada seluruh pihak yang telah membantu dalam penyelesaian skripsi ini. Ucapan terimakasih penulis sampaikan kepada:

1. Bapak Dr. Phil Sahiron, M.A, selaku pelaksana tugas (Plt.) Rektor UIN Sunan Kalijaga Yogyakarta.
2. Bapak Dr. H. Waryono, M.Ag., selaku Wakil Rektor Bidang Kemahasiswaan dan Kerjasama UIN Sunan Kalijaga Yogyakarta.
3. Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Bapak Sumarsono, S.T., M.Kom., selaku Ketua Program Studi S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
5. Bapak Rahmat Hidayat, S.Kom., M.Cs., selaku dosen pembimbing skripsi yang telah sabar membimbing, memberikan arahan, dan motivasi dalam penyusunan skripsi.
6. Bapak Dr. Agung Fatwanto, S.Si., M.Kom., selaku Dosen Pembimbing Akademik.

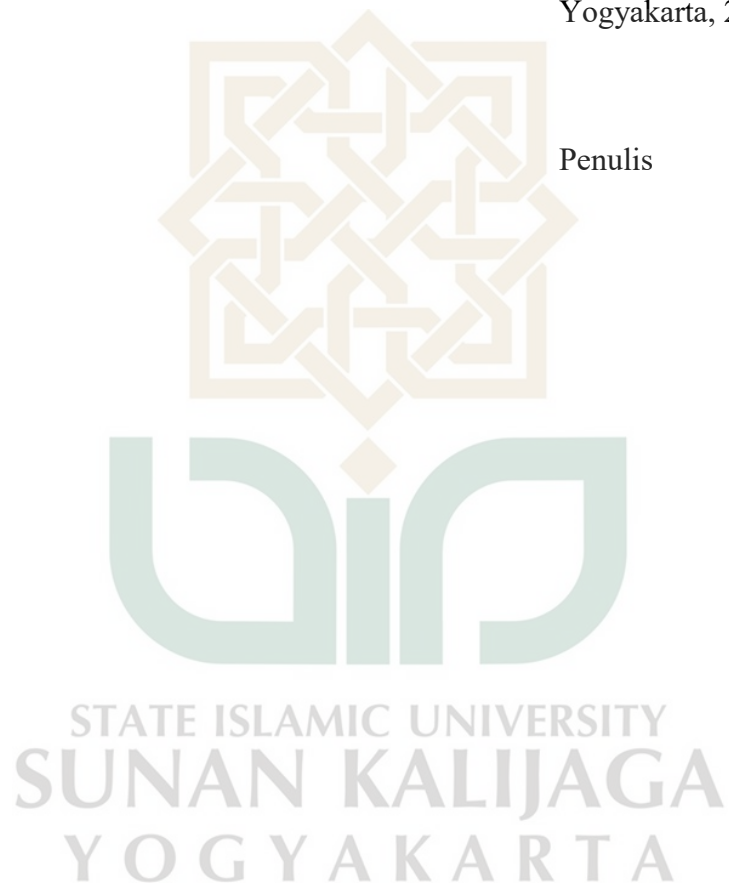
7. Agus Mulyanto, S.Si., M.Kom., Ph.D., Aulia Faqih Rifa'i, M.Kom., Muhammad Didik Rohmad Wahyudi, S.T., M.T., Nurochman, S.Kom., M. Taufiq Nuruzzaman, S.T., Maria Ulfah Siregar, S.Kom. MIT., Ph.D., M.Kom., Dr. Shofwatul 'Uyun, S.T., M.Kom., selaku dosen pengampu mata kuliah program studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
8. Kedua orang tua, Bapak Sugimin S.Pd dan Ibu Fitri Mufidah yang telah memberikan dukungan, semangat dan do'a sehingga penulis dapat selalu berjuang dalam hidup dan berusaha menjadi pribadi yang bermanfaat bagi orang lain.
9. Kedua adik kesayanganku, Amalia Sholihah dan Rafli Maulana Al-Ghifari yang selalu membuat penulis lebih semangat dan bertanggung jawab dalam belajar dan menyelesaikan strata S-1.
10. Pihak Yayasan Solopeduli yang telah membantu dalam kelancaran penulis dalam penyusunan skripsi.
11. Seluruh Keluarga Besar Mbah Mintodiyono dan Mbah Sular yang tak henti-hentinya memberikan do'a dan dukungan.
12. Sahabat-sahabat sejak kecilku Widya Pangestika dan Hanifah Putri M yang memberikan semangat dan tawa canda dalam kehidupan penulis.
13. Sahabat-sahabat gathelsampehalal Sekar Minati, Lina Puspitasari, Ulfa Mulya, dan Yayang Wijaya yang telah memberikan semangat, warna-warni kehidupan dan mungkin sedikit do'a kepada penulis.

14. Teman-teman Trio n Dad Lia, Ulfa, dan Raffi yang membantu dan bekerjasama dengan baik selama berkuliah di Prince of Songkhla, Thailand.
15. Seluruh teman-teman Teknik Informatika angkatan 2016 yang tidak dapat penulis sebutkan satu per satu.
16. Teman-teman sambat Sholihuddin Nur, Devi, Imam, Ricky, Bangun, Garin, Irwan dan Andhy yang selalu menemani dan membuat penulis semangat dalam menyelesaikan tugas akhir.
17. Teman-teman IMAGIRI Yogyakarta yang membuat penulis sadar bahwa manusia bukanlah manusia jika tanpa memberi manfaat kepada sesama.
18. Teman-teman IKPM-Jateng Yogyakarta yang membantu penulis mengetahui arti keorganisasian dan tanggung jawab terhadap amanah.
19. Ajarn-ajarn *Prince of Songkhla University* Kasikrit Damkliang, Numtip Trakulmaykee dan Nittida Elz yang membimbing saya selama program *student mobility credit transfer program* dengan sabar dan mengesankan.
20. Member of Sirarom Thailand yang membuat perjuangan 4,5 bulan belajar di *Prince of Songkhla University* menjadi lebih mudah dan menyenangkan.
21. Kakak-kakak PERMITHA-PSU yang memberikan motivasi kepada penulis untuk menjadi akademisi yang lebih baik dari sekarang.
22. Teman-teman belajar di Thailand P'Aseeyah Mintalaeh, P'Khan Chawanwit Meesang, P'Fandy, dan semua teman-teman baik disana.
23. Teman-teman KKN Kelompok 92 Angkatan 99 yang telah menjadi pengalaman mengesankan dalam belajar bermasyarakat.

24. Serta seluruh pihak yang tidak dapat disebutkan satu persatu dan telah memberikan banyak doa dan dukungan sehingga penelitian ini dapat terselesaikan.

Yogyakarta, 25 Maret 2020

Penulis



HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan kepada kedua orang tua,

Sugimin dan Fitri Mufidah yang memberi kebebasan dengan pemahaman.

Sehingga penulis dapat belajar banyak hal terlepas dari pendidikan yang ada.

Dan kepada kedua adikku tersayang,

*Amalia Sholihah dan Rafla Maulana Alghifari yang memberi kebahagiaan dengan
kerewelan.*



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN MOTTO

“Semua makhluk hebat dalam satu hal, tapi tidak dalam segala hal”

-Spongebob Squarepants-

“Hidup itu adalah seni menggambar tanpa penghapus.”

-John W.Gardner-

“Bekerja keras dan bersikap baiklah. Hal luar biasa akan terjadi.”

-Conan O'Brien-



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
SURAT PERSETUJUAN SKRIPSI.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR	v
HALAMAN PERSEMBAHAN	ix
HALAMAN MOTTO	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR RUMUS	xvii
INTISARI.....	xviii
ABSTRACT	xix
BAB I.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
1.6. Keaslian Penelitian.....	4
1.7. Sistematika Penulisan.....	4
BAB II.....	6
2.1. Tinjauan Pustaka	6
2.2. Landasan Teori	12
2.2.1. <i>Data Mining</i>	12
2.2.2. Klasifikasi	14
2.2.3. <i>Decision Tree</i> (Pohon Keputusan).....	15

2.2.4.	<i>Iterative Dichotomiser 3 (ID3)</i>	16
2.2.4.1.	Pembentukan Simpul Akar	18
2.2.4.2.	Pembentukan Simpul 1.1	21
2.2.5.	C4.5	24
2.2.5.1.	Pembentukan Simpul Akar	25
2.2.5.2.	Pembentukan Simpul 1.1	27
2.2.5.3.	Pembentukan Simpul 1.1.1	31
2.2.6.	<i>Confusion Matrix</i>	34
2.2.7.	Beasiswa	38
2.2.8.	Yayasan Solopeduli	38
2.2.9.	<i>Python</i>	39
BAB III	41
3.1.	Metode Penelitian.....	41
3.2.	Alur Penelitian.....	41
3.2.1.	Studi Pendahuluan	41
3.2.2.	Pengumpulan Data	42
3.2.3.	<i>Preprocessing Data</i>	42
3.2.4.	Implementasi.....	42
3.2.5.	Pembuatan Laporan	42
3.3.	Perangkat Penelitian.....	43
3.3.1.	Perangkat Keras	43
3.3.2.	Perangkat Lunak	43
BAB IV	44
4.1.	Pengumpulan Data	45
4.2.	<i>Preprocessing Data</i>	45
4.2.1.	Penghilangan Data <i>Null</i>	46
4.2.2.	Pengkategorian dalam Atribut	46
4.3.	Analisis Algoritma	47

4.3.1.	Algoritma <i>Iterative Dichotomiser 3 (ID3)</i>	47
4.3.1.1.	Pembentukan Simpul Akar	48
4.3.2.	Algoritma C4.5	54
4.3.2.1.	Pembentukan Simpul Akar	55
4.3.2.2.	Pembentukan Simpul Keputusan 1.1	58
4.4.	Hasil dengan Perbandingan Data.....	66
4.4.1.	Uji <i>Confusion Matrix</i>	68
4.4.2.	Uji Waktu.....	70
BAB V.....		72
5.1.	Kesimpulan.....	72
5.2.	Saran.....	72
DAFTAR PUSTAKA		74
LAMPIRAN.....		78
CURRICULUM VITAE.....		92

DAFTAR GAMBAR

Gambar 2.1 Tahap-tahap proses KDD (Fayyad et al., 1996)	13
Gambar 2.2 Pohon Keputusan ID3 (simpul akar)	21
Gambar 2.3 Pohon Keputusan ID3 (simpul 1.1 atau akhir)	23
Gambar 2.4 Pohon Keputusan C4.5 (simpul Akar).....	27
Gambar 2.5 Pohon Keputusan C4.5 (simpul 1.1).....	31
Gambar 2.6 Pohon Keputusan C4.5 (simpul 1.1.1 atau akhir).....	34
Gambar 4.1 <i>Flowchart</i> Proses Bisnis	44
Gambar 4.2 Pohon Keputusan <i>ID3</i>	54
Gambar 4.3 Pohon Keputusan C4.5	58
Gambar 4.4 Pohon Keputusan C4.5	66
Gambar 4.5 Pohon Keputusan <i>ID3</i> (4 data)	68
Gambar 4.6 Pohon Keputusan C4.5 (4 data).....	69

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	9
Tabel 2.2 Tinjauan Pustaka (Lanjutan)	10
Tabel 2.3 Tinjauan Pustaka (Lanjutan)	11
Tabel 2.4 Data Studi Kasus Bermain	17
Tabel 2.5 Data Studi Kasus Bermain	18
Tabel 2.6 Jumlah Data Atribut Cuaca	20
Tabel 2.7 Jumlah Data Atribut Kelembaban	23
Tabel 2.8 Jumlah Data Atribut Berangin.....	27
Tabel 2.9 Jumlah Data Atribut Cuaca	30
Tabel 2.10 Jumlah Data Atribut Kelembaban	33
Tabel 2.11 <i>Confusion Matrix</i>	34
Tabel 2.12 Data Prediksi Buah.....	36
Tabel 4.1 Contoh Data.....	45
Tabel 4.2 Keterangan Kategori Data.....	46
Tabel 4.3 Keterangan Kategori Data (Lanjutan).....	47
Tabel 4.4 Contoh Data Setelah <i>Preprocessing</i>	47
Tabel 4.5 Perhitungan Jumlah Data	48
Tabel 4.6 Hasil Perhitungan <i>Entropy</i> dan <i>Information Gain</i>	53
Tabel 4.7 Jumlah Data Atribut BTA	54
Tabel 4.8 Hasil Perhitungan <i>Split Information</i> dan <i>Gain Ratio</i>	57
Tabel 4.9 Jumlah Data Atribut Merokok Terhadap Klasifikasi	58

Tabel 4.10 Jumlah Objek Data dengan Atribut Merokok Bernilai T	59
Tabel 4.11 Hasil Perhitungan <i>Entropy</i> , <i>Information Gain</i> , <i>Split Information</i> , dan <i>Gain Ratio</i>	64
Tabel 4.12 Hasil Perhitungan <i>Entropy</i> , <i>Information Gain</i> , <i>Split Information</i> , dan <i>Gain Ratio</i> (Lanjutan)	65
Tabel 4.13 Jumlah Data Atribut Kendaraan Terhadap Klasifikasi.....	66
Tabel 4.14 Contoh Data Dummy	67
Tabel 4.15 Pembagian Data.....	67
Tabel 4.16 Confusion Matrix dengan Perbandingan Data Uji	70
Tabel 4.17 Uji Waktu dengan Perbandingan Data Uji	71

DAFTAR RUMUS

Entropy (1)	16
Information Gain (2)	16
Split Information (3)	24
Gain Ratio (4)	25
Recall (5)	35
Precision (6)	35
Specificity (7)	36
Accuracy (8)	36



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

**Analisis Komparasi Algoritma *ID3* dan *C4.5* dalam Klasifikasi Penerimaan
Basiswa (Studi Kasus : Basiswa Solopeduli)**

Nadia Sholihah

16650014

INTISARI

Beasiswa merupakan bantuan atau penghargaan yang diberikan kepada pelajar dengan ekonomi rendah berupa biaya pendidikan. Berbagai instansi berlomba-lomba dalam membantu mencerdaskan bangsa melalui penyelenggaraan beasiswa, seperti halnya Yayasan Solopeduli. Namun, dalam pelaksanaannya seringkali terjadi salah sasaran atau target. Untuk menangani permasalahan tersebut maka perlu penerapan *data mining*.

Penelitian ini melakukan klasifikasi dengan algoritma *ID3* dan *C4.5* yang keduanya sama-sama algoritma pembangun *decision tree* dan membandingkan akurasi serta efektifitasnya. Jumlah data yang digunakan adalah 512 data tahun ajaran 2017/2018 dan 250 data *dummy* yang dibagi menjadi data latih dan uji. Atribut yang digunakan terdiri dari wilayah, administrasi (Adm), baca tulis qu'ran (BTA), agama, merokok, kendaraan, dan jenjang.

Berdasarkan pengujian menggunakan *confusion matrix*, algoritma *C4.5* lebih baik dibandingkan dengan *ID3* karena memiliki tingkat akurasi yang tinggi. Pada empat pembagian data, akurasi tertinggi masing-masing algoritma terletak pada data uji 70%. Algoritma *ID3* menghasilkan akurasi sebesar 99,06% dan untuk algoritma *C4.5* sebesar 99,81%.

Berdasarkan pengujian waktu eksekusi, algoritma *ID3* lebih efektif dibandingkan *C4.5* karena memiliki waktu terpendek pada data uji 90%. *ID3* memiliki waktu 0.246897 dan *C4.5* memiliki waktu 0.346025.

Kata kunci: *data mining*, klasifikasi, *decision tree*, *ID3*, *C4.5*, *confusion matrix*.

Comparative Analysis of ID3 and C4.5 Algorithms in Scholarship

Acceptance Classification (Case Study: Solopeduli Scholarship)

Nadia Sholihah

16650014

ABSTRACT

Scholarships are aids or awards given to students with low economics in the form of tuition fees. Various agencies are competing to help educate the nation through organizing scholarships, such as the Solopeduli Foundation. However, in its implementation, often, the wrong target or target occurs. To deal with these problems, using data mining is necessarily essential.

This study classifies the ID3, and C4.5 algorithms, both of them are decision tree building algorithms and compare the accuracy and effectiveness. The amount of data used is 512 data for the 2017/2018 school year and 250 dummy data, which is divided into training and test data. The attributes used to consist of the region, administration (ADM), read-write The Quran (BTA), religion, smoking, vehicles, and levels.

Based on testing using a confusion matrix, the C4.5 algorithm is better than ID3 because it has a high degree of accuracy. In the four data shares, the highest accuracy of each algorithm lies in the 70% test data. ID3 algorithm produces an accuracy of 99.06% and for the C4.5 algorithm of 99.81%.

Based on testing the execution time, the ID3 algorithm is more effective than C4.5 because it has the shortest time on the 90% test data. ID3 has a time of 0.246897 and C4.5 has a time of 0.346025.

Keywords: data mining, classification, decision tree, ID3, C4.5, confusion matrix.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Beasiswa merupakan bantuan berupa biaya pendidikan dari suatu instansi yang dalam penyeleksiannya memiliki beberapa syarat diterima. Seperti halnya beasiswa Yayasan Solopeduli yang memiliki beberapa syarat yang terdiri dari syarat umum dan administrasi, syarat umum diantaranya masih tercatat aktif sebagai siswa, beragama islam, bisa membaca Al-Qur'an. Sedangkan untuk syarat administrasi adalah pengisian *form* yang berisi informasi pendaftar, *fotocopy* raport, *fotocopy* (Kartu Tanda Penduduk) KTP orangtua/wali, *fotocopy* Kartu Keluarga (KK), dan *Essay* tentang diri dan keluarga. Penentuan yang berjalan saat ini hanya berdasarkan administrasi dan dilanjutkan survei. Maka terdapat permasalahan yaitu belum ditemukannya formulasi terhadap keputusan pemberian beasiswa baik diterima atau tidak diterima. Hal tersebut menimbulkan banyaknya keobjektifan data dan menimbulkan ketidakadilan. Adanya peluang penyelewengan dan menimbulkan keluhan pendaftar terhadap keadilan pembagian beasiswa dapat menjadi pertimbangan penting dalam perlunya peran teknologi. Keadilan dalam pemberian amanatpun ditegaskan oleh Allah SWT dalam QS.An-Nisaa': 58 yang berbunyi :

إِنَّ اللَّهَ يَأْمُرُكُمْ أَنْ تُؤَدُّوا أَلْأَمَانَاتِ إِلَى أَهْلِهَا وَإِذَا حَكَمْتُمْ بَيْنَ النَّاسِ أَنْ تَحْكُمُوا بِالْعَدْلِ إِنَّ اللَّهَ نِعِمَّا
يَعِظُكُمْ بِهِ فَلْيَبْصِرُوا

“*Sesungguhnya Allah menyuruh kamu menyampaikan amanat kepada yang berhak menerimanya, dan (menyuruh kamu) apabila menetapkan hukum di antara manusia supaya kamu menetapkan dengan adil. Sesungguhnya Allah memberi pengajaran yang sebaik-baiknya kepadamu. Sesungguhnya Allah adalah Maha mendengar lagi Maha Melihat.*” (QS. An-Nisa’: 58)

Dalam mengatasi permasalahan diatas, maka dilakukan analisis penyeleksian pendaftar dengan algoritma *ID3* dan *C4.5* berdasarkan tujuh kriteria yang tiap tahunnya ada.

Terdapat berbagai algoritma untuk klasifikasi, seperti *Naive Bayes*, *K-Means*, *K-Nearest Neighbor*, *Decision Tree*, *Support Vector Machine (SVM)*, dan *Artificial Neural Networks (ANN)*. Dalam algoritma *Decision Tree* terdapat beberapa jenis, yaitu *Decision Tree Iterative Dichotomiser 3 (ID3)*, *Decision Tree C4.5*, *Decision Tree CHAID*, *Decision Tree CART*, dan *Decision Tree C5.0*. Dari beberapa algoritma, dibandingkan akurasi dari *Decision Tree ID3* dan *C4.5*. Pemilihan kedua algoritma dikarenakan data pada kasus ini yang berupa diskrit atau kategorikal, algoritma ini juga memiliki interpretasi konsep yang lebih mudah dipahami dibanding algoritma lain.

Keefektifan pemberian beasiswa terhadap mahasiswa haruslah tepat guna. Komparasi merupakan salah satu metode untuk mengukur dari pendukung keputusan terhadap calon mahasiswa yang ingin mendapatkan beasiswa (Maulana, 2016).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka permasalahan yang ada adalah bagaimana perbandingan akurasi dan keefektifan algoritma *Decision Tree ID3* dan *Decision Tree C4.5* dalam studi kasus Beasiswa Solopeduli.

1.3. Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yang dibahas agar penyusunan dan pembahasan penelitian dapat dilakukan secara terarah dan tercapai sesuai dengan yang diharapkan. Antara lain sebagai berikut:

1. Pemodelan klasifikasi menggunakan data pendaftar Beasiswa Solopeduli pada tahun 2017/2018.
2. Klasifikasi sistem mencakup diterima atau tidak diterima.
3. Klasifikasi dengan penerapan algoritma *Decision Tree ID3* dan *Decision Tree C4.5*.
4. Klasifikasi menggunakan bahasa pemrograman *python*.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah membandingkan algoritma *Decision Tree ID3* dan *C4.5* pada proses penyeleksian penerimaan beasiswa Solopeduli tahun 2017/2018 menggunakan kriteria *Confusion Matrix* dan pengujian waktu eksekusi.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah didapatkannya hasil perbandingan algoritma *Decision Tree ID3* dan *C4.5* pada proses penyeleksian penerimaan beasiswa Solopeduli tahun 2017/2018 menggunakan kriteria *Confusion Matrix* dan pengujian waktu eksekusi.

1.6. Keaslian Penelitian

Penelitian *data mining* dengan analisis komparasi pernah dilakukan, namun analisis komparasi metode *Decision Tree ID3* dan *C4.5* dengan studi kasus Beasiswa Solopeduli belum pernah ditemukan oleh peneliti.

1.7. Sistematika Penulisan

Sebagai gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Penyusunan laporan tugas akhir ini memiliki sistematika penulisan yang diawali dari BAB I dan diakhiri BAB V. Berikut adalah penjelasan pada tiap-tiap bab dalam laporan penelitian ini:

BAB I PENDAHULUAN

Bab pendahuluan berisikan penjelasan mengenai latar belakang dilakukannya penelitian, rumusan masalah penelitian, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab tinjauan pustaka dan landasan teori berisikan mengenai penelitian terdahulu dan teori-teori dasar yang terkait dengan penelitian ini. Teori yang digunakan terdiri dari *data mining*, klasifikasi, *decision tree (ID3 dan C4.5)*, *confusion matrix*, *Python*.

BAB III METODE PENELITIAN

Bab metode penelitian berisi tentang penjelasan mengenai metode ataupun algoritma yang digunakan serta tahapan-tahapan yang dilakukan untuk mencapai tujuan dan kesimpulan tugas akhir.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan membahas analisis data dan hasil dari penelitian yang telah dilakukan.

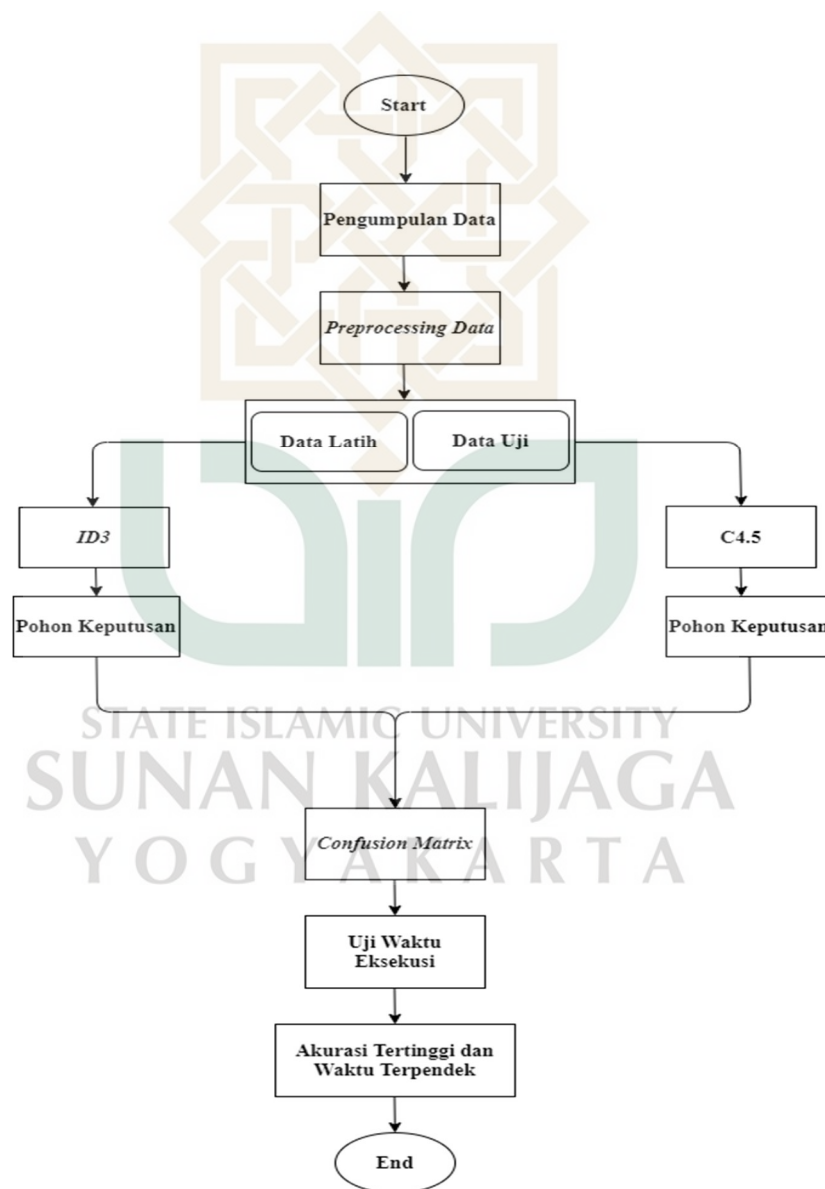
BAB V PENUTUP

Bab penutup berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan. Selanjutnya, kekurangan yang ada pada penelitian dituliskan pada saran untuk pengembangan penelitian di masa yang akan datang.

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini, akan dianalisis secara rinci terkait masing-masing algoritma dan perbandingan keduanya berdasarkan *flowchart* dibawah ini:



Gambar 4.1 *Flowchart* Proses Bisnis

4.1. Pengumpulan Data

Pengumpulan data dilakukan di instansi terkait yaitu Yayasan Solopeduli bagian program beasiswa. Data yang didapat berupa data pendaftar beasiswa pada tahun ajaran 2017/2018 dengan jumlah 512 data dengan klasifikasi 441 peserta diterima dan 71 tidak diterima. Data tersebut berisikan informasi data pendaftar dengan syarat-syarat yang ada, baik tercukupi atau tidak. Dalam penelitian ini, hanya diambil data yang memiliki peran dalam proses klasifikasi. Untuk atribut nama dan sekolah tidak dimasukkan dalam data. Setelah melakukan proses *input*, data perlu dirubah menjadi file .csv.

Tabel 4.1 Contoh Data

Hasil	Wilayah	Adm	BTA	Agama	Merokok	Kendaraan	Jenjang
Ya	Khusus	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Ya	Khusus	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Ya	Khusus	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Tidak	Khusus	Lengkap	Tidak	Islam	Ya	Ya	SMA/K
Ya	Luar Solo	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Ya	Khusus	Lengkap	Baik	Islam	Tidak	Tidak	SMP
Ya	Khusus	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Tidak	Khusus	Lengkap	Tidak	Islam	Ya	Ya	SMP
Ya	Luar Solo	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Ya	Solo	Lengkap	Cukup	Islam	Tidak	Tidak	SD
Tidak	Solo	Tidak	Tidak	Islam	Ya	Tidak	SD

4.2. Preprocessing Data

Sebelum dilakukan analisa menggunakan bahasa pemrograman *python*, data perlu dilakukan *preprocessing* . Proses tersebut meliputi melakukan penghilangan data yang bernilai *null* dan mengkategorikalkan data. Contoh data yang telah dilakukan *preprocessing* dapat dilihat pada tabel 4.2 dan 4.3 dibawah ini.

4.2.1. Penghilangan Data Null

Proses *preprocessing* yang pertama adalah penghilangan data yang kosong atau bernilai *null*. Pertama, dilakukan pengecekan terhadap data. Apakah terdapat data kosong atau tidak. Setelah dilakukan pengecekan, selanjutnya adalah menghilangkan data yang terdeteksi memiliki data kosong.

4.2.2. Pengkategorian dalam Atribut

Proses *preprocessing* yang kedua adalah mengkategorikan anggota-anggota yang ada dalam atribut. Proses ini dilakukan dengan fitur *find and replace* dalam *microsoft excel*. Hal ini dianggap lebih mempermudah karena proses pengkategorian dapat disesuaikan dengan keinginan. Keterangan terkait kategori suatu nilai dari atribut dapat dilihat pada tabel 4.2 dan 4.3. Dan data yang telah dilakukan 2 proses *preprocessing* atau dapat dikatakan sebagai data bersih adalah seperti pada tabel 4.4.

Tabel 4.2 Keterangan Kategori Data

Atribut	Nilai Awal	Nilai Kategorikal
Hasil	Ya	Y
	Tidak	N
Wilayah	Solo	1
	Luar Solo	2
	Khusus	3
Adm	Lengkap	C
	Tidak Lengkap	I
BTA	Baik	H
	Cukup	M
Agama	Tidak Hadir	A
	Islam	I
Merokok	Ya	T

Tabel 4.3 Keterangan Kategori Data (Lanjutan)

Atribut	Nilai Awal	Nilai Kategorikal
	Tidak	F
Kendaraan	Punya	H
	Tidak	D
Jenang	SD	1

Tabel 4.4 Contoh Data Setelah *Preprocessing*

Hasil	Wilayah	Adm	BTA	Agama	Merokok	Kendaraan	Jenang
Y	3	C	M	I	F	D	1
Y	3	C	M	I	F	D	1
Y	3	C	M	I	F	D	1
N	3	C	A	I	T	H	3
Y	2	C	M	I	F	D	1
Y	3	C	H	I	F	D	2
Y	3	C	M	I	F	D	1
N	3	C	A	I	T	H	2
Y	2	C	M	I	F	D	1
Y	1	C	M	I	F	D	1
N	1	I	A	I	T	D	1

4.3. Analisis Algoritma

4.3.1. Algoritma *Iterative Dichotomiser 3 (ID3)*

Algoritma ID3 merupakan salah satu algoritma pembangun pohon keputusan. Pohon keputusan dibangun dimulai dari simpul akar dengan perhitungan *entropy* dan *information gain*. Pada penelitian ini, 512 data pendaftar digunakan sebagai data latih. Dalam pohon keputusan akan ada simpul akar, simpul keputusan, dan simpul daun. Pembentukan akar hal pertama yang dilakukan.

4.3.1.1. Pembentukan Simpul Akar

Sebelum hal tersebut, dilakukan perhitungan jumlah objek data dengan klasifikasinya berdasarkan atribut wilayah, administrasi, BTA (Baca Tulis Al-Qur'an), agama, merokok, kendaraan, dan jenjang. Tabel dibawah 4.5 menggambarkan hasil perhitungan jumlah data.

Tabel 4.5 Perhitungan Jumlah Data

Atribut	Jumlah Kasus		Hasil Klasifikasi	
			Y	N
Total	512		441	71
Wilayah	1	139	120	19
	2	30	29	1
	3	343	292	51
Adm	C	501	441	60
	I	11	0	11
BTA	H	301	301	0
	M	140	140	0
	A	71	0	71
Agama	I	512	512	0
Merokok	T	94	23	71
	F	418	418	0
Kendaraan	H	147	76	71
	D	365	365	0
Jenjang	1	160	140	20
	2	217	174	43
	3	135	127	8

Hal pertama adalah melakukan perhitungan *entropy* total dan dilanjutkan dengan *entropy* masing-masing atribut.

$$\begin{aligned}
 \text{Entropy (Total)} &= \left(\left(\frac{-441}{512} \right) \log_2 \frac{441}{512} \right) + \left(\left(\frac{-71}{512} \right) \log_2 \frac{71}{512} \right) \\
 &= (-0,86132 (-0,21537)) + (-0,13867 (-2,85027)) \\
 &= (0,18550) + (0,39525)
 \end{aligned}$$

$$= 0,58075$$

Menghitung *entropy* atribut wilayah dengan nilai 1, 2, dan 3.

$$\begin{aligned} \text{Entropy (Wilayah, 1)} &= \left(\left(\frac{-120}{139} \right) \log_2 \frac{120}{139} \right) + \left(\left(\frac{-19}{139} \right) \log_2 \frac{19}{139} \right) \\ &= (-0,86331 (-0,21205)) + (-0,13669 (-2,87102)) \\ &= (0,18306) + (0,39244) \\ &= 0,57550 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Wilayah, 2)} &= \left(\left(\frac{-29}{30} \right) \log_2 \frac{29}{30} \right) + \left(\left(\frac{-1}{30} \right) \log_2 \frac{1}{30} \right) \\ &= (-0,96667 (-0,04890)) + (-0,03333 (-4,90689)) \\ &= (0,04728) + (0,16356) \\ &= 0,21084 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Wilayah, 3)} &= \left(\left(\frac{-292}{343} \right) \log_2 \frac{292}{343} \right) + \left(\left(\frac{-51}{343} \right) \log_2 \frac{51}{343} \right) \\ &= (-0,85131 (-0,23224)) + (-0,14869 (-2,74964)) \\ &= (0,19770) + (0,40884) \\ &= 0,60654 \end{aligned}$$

Menghitung *entropy* atribut administrasi dengan nilai C dan I.

$$\begin{aligned} \text{Entropy (Adm, C)} &= \left(\left(\frac{-441}{501} \right) \log_2 \frac{441}{501} \right) + \left(\left(\frac{-60}{501} \right) \log_2 \frac{60}{501} \right) \\ &= (-0,88024 (-0,18403)) + (-0,11976 (-3,06178)) \\ &= (0,16199) + (0,36668) \\ &= 0,52867 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Adm, I)} &= \left(\left(\frac{-0}{11} \right) \log_2 \frac{0}{11} \right) + \left(\left(\frac{-11}{11} \right) \log_2 \frac{11}{11} \right) \\ &= 0 \end{aligned}$$

Menghitung *entropy* atribut Baca Tulis Al-Qur'an (BTA) dengan nilai H, M, A.

$$\begin{aligned} \text{Entropy (BTA, H)} &= \left(\left(\frac{-301}{301} \right) \log_2 \frac{301}{301} \right) + \left(\left(\frac{-0}{301} \right) \log_2 \frac{0}{301} \right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy (BTA, M)} &= \left(\left(\frac{-140}{140} \right) \log_2 \frac{140}{140} \right) + \left(\left(\frac{-0}{140} \right) \log_2 \frac{0}{140} \right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy (BTA, M)} &= \left(\left(\frac{-0}{71} \right) \log_2 \frac{0}{71} \right) + \left(\left(\frac{-71}{71} \right) \log_2 \frac{71}{71} \right) \\ &= 0 \end{aligned}$$

Menghitung *entropy* atribut agama dengan satu nilai yaitu I.

$$\begin{aligned} \text{Entropy (Agama, I)} &= \left(\left(\frac{-512}{512} \right) \log_2 \frac{512}{512} \right) + \left(\left(\frac{-0}{512} \right) \log_2 \frac{0}{512} \right) \\ &= 0 \end{aligned}$$

Menghitung *entropy* atribut merokok dengan nilai T dan F.

$$\begin{aligned} \text{Entropy (Merokok, T)} &= \left(\left(\frac{-23}{94} \right) \log_2 \frac{23}{94} \right) + \left(\left(\frac{-71}{94} \right) \log_2 \frac{71}{94} \right) \\ &= (-0,24468 (-2,03103)) + (-0,755319 (-0,40484)) \\ &= (0,49695) + (0,305785) \\ &= 0,80274 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Merokok, F)} &= \left(\left(\frac{-418}{418} \right) \log_2 \frac{418}{418} \right) + \left(\left(\frac{-0}{418} \right) \log_2 \frac{0}{418} \right) \\ &= 0 \end{aligned}$$

Menghitung *entropy* atribut kendaraan dengan nilai H dan D.

$$\begin{aligned} \text{Entropy (Kendaraan, H)} &= \left(\left(\frac{-76}{147} \right) \log_2 \frac{76}{147} \right) + \left(\left(\frac{-71}{147} \right) \log_2 \frac{71}{147} \right) \\ &= (-0,51701 (-0,95174)) + (-0,48299 (-1,04993)) \\ &= (0,49206) + (0,50711) \\ &= 0,99917 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Kendaraan, D)} &= \left(\left(\frac{-365}{365} \right) \log_2 \frac{365}{365} \right) + \left(\left(\frac{-0}{365} \right) \log_2 \frac{0}{365} \right) \\ &= 0 \end{aligned}$$

Menghitung entropy jenjang dengan tiga nilai yaitu 1, 2, dan 3.

$$\begin{aligned} \text{Entropy (Jenjang, 1)} &= \left(\left(\frac{-140}{160} \right) \log_2 \frac{140}{160} \right) + \left(\left(\frac{-20}{160} \right) \log_2 \frac{20}{160} \right) \\ &= (-0,87500 (-0,19265)) + (-0,12500 (-3)) \\ &= (0,16856) + (0,37500) \\ &= 0,54356 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Jenjang, 2)} &= \left(\left(\frac{-174}{217} \right) \log_2 \frac{174}{217} \right) + \left(\left(\frac{-43}{217} \right) \log_2 \frac{43}{217} \right) \\ &= (-0,80184 (-0,31861)) + (-0,19816 (-2,33529)) \\ &= (0,25547) + (0,46275) \\ &= 0,71823 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Jenjang, 3)} &= \left(\left(\frac{-127}{135} \right) \log_2 \frac{127}{135} \right) + \left(\left(\frac{-8}{135} \right) \log_2 \frac{8}{135} \right) \\ &= (-0,94074 (-0,088130910278665)) + (-0,05926 (-4,07682)) \\ &= (0,08291) + (0,24159) \\ &= 0,32450 \end{aligned}$$

Setelah *entropy* total dan masing-masing atribut dengan nilai-nilainya didapatkan, maka langkah selanjutnya adalah menghitung *information gain* yang nantinya akan dijadikan acuan dalam menentukan simpul.

$$\begin{aligned} \text{Gain (Wilayah)} &= 0,58075 - \left(\left(\frac{139}{512} \right) 0,57550 + \left(\frac{30}{512} \right) 0,21084 + \left(\frac{343}{512} \right) 0,60654 \right) \\ &= 0,58075 - 0,57493 \\ &= 0,00582 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Administrasi)} &= 0,58075 - \left(\left(\frac{501}{512} \right) 0,52867 + \left(\frac{11}{512} \right) 0 \right) \\
 &= 0,58075 - 0,51731 \\
 &= 0,06344
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (BTA)} &= 0,58075 - \left(\left(\frac{301}{512} \right) 0 + \left(\frac{140}{512} \right) 0 + \left(\frac{71}{512} \right) 0 \right) \\
 &= 0,58075 - 0 \\
 &= 0,58075
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Agama)} &= 0,58075 - \left(\left(\frac{512}{512} \right) 0 \right) \\
 &= 0,58075 - 0 \\
 &= 0,58075
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Merokok)} &= 0,58075 - \left(\left(\frac{94}{512} \right) 0,80274 + \left(\frac{418}{512} \right) 0 \right) \\
 &= 0,58075 - 0,14738 \\
 &= 0,43337
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Kendaraan)} &= 0,58075 - \left(\left(\frac{147}{512} \right) 0,99917 + \left(\frac{365}{512} \right) 0 \right) \\
 &= 0,58075 - 0,14738 \\
 &= 0,43337
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Jenjang)} &= 0,58075 - \left(\left(\frac{160}{512} \right) 0,54356 + \left(\frac{217}{512} \right) 0,71823 + \left(\frac{135}{512} \right) 0,32450 \right) \\
 &= 0,58075 - 0,55983 \\
 &= 0,02092
 \end{aligned}$$

Untuk mengetahui dengan mudah masing-masing hasil dari perhitungan, maka hal tersebut ditampilkan dalam tabel 4.6.

Tabel 4.6 Hasil Perhitungan *Entropy* dan *Information Gain*

		Jumlah Kasus	No	Yes	Entropy	Information Gain
Total		512	71	441	0,58075	
Wilayah						0,00582
	1	139	19	120	0,57551	
	2	30	1	29	0,21084	
	3	343	51	292	0,60655	
Administrasi						0,06344
	C	501	60	441	0,52867	
	I	11	11	0	0,00000	
BTA						0,58075
	H	301	0	301	0,00000	
	M	140	0	140	0,00000	
	A	71	71	0	0,00000	
Agama						0,58075
	I	512	0	512	0,00000	
Merokok						0,43337
	T	94	71	23	0,80274	
	F	418	0	418	0,00000	
Kendaraan						0,29388
	H	147	71	76	0,99917	
	D	365	0	365	0,00000	
Jenjang						0,02092
	1	160	20	140	0,54356	
	2	217	43	174	0,71823	
	3	135	8	127	0,32450	

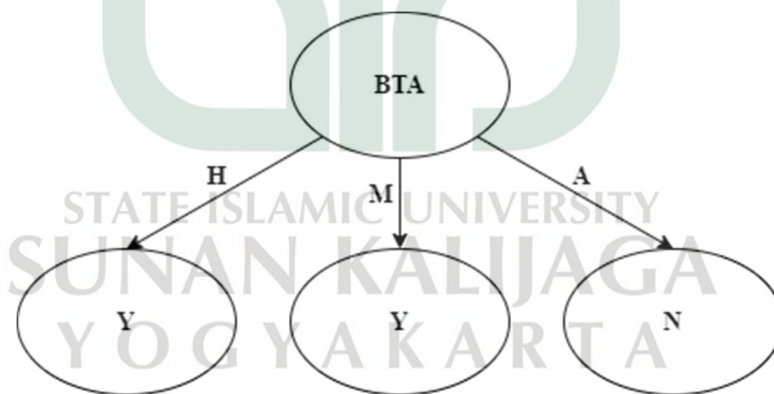
Dari tabel diatas terlihat bahwa *information gain* tertinggi yaitu BTA dan agama dengan hasil 0,580749976. Karena ada dua, maka dipilih salah satunya. Dalam penelitian ini dipilih BTA sebagai simpul akar dari pohon keputusan. Atribut BTA memiliki 3 nilai, yaitu H, M, dan A yang akan dijadikan cabang dari simpul akar BTA.

Tabel 4.7 Jumlah Data Atribut BTA

Atribut	Nilai	Jumlah Kasus	Diterima	Tidak
BTA				
	H	301	301	0
	M	140	140	0
	A	71	0	71

Dilihat dari jumlah data terhadap hasil klasifikasi ketiga nilai dari BTA pada tabel 4.8 dapat diputuskan bahwa pohon keputusan tidak berlanjut. Pohon keputusan berhenti, karena :

- Pada nilai H, 301 data (semua) terklasifikasi diterima
- Pada nilai M, 140 data (semua) terklasifikasi diterima
- Pada nilai A, 71 data (semua) terklasifikasi tidak diterima

Gambar 4.2 Pohon Keputusan *ID3*

4.3.2. Algoritma C4.5

Algoritma C4.5 adalah pengembangan dari algoritma *ID3*. Perbedaan algoritma ini dengan *ID3* adalah pada cara penentuan simpul. Jika pada *ID3*

menggunakan *information gain*, maka pada C4.5 menggunakan *gain ratio*. Garis besar dari algoritma ini adalah *entropy*, *information gain*, *split information*, dan *gain ratio*. *Gain ratio* merupakan suksesor karena *information gain* belum dapat secara optimal menentukan simpul.

4.3.2.1. Pembentukan Simpul Akar

Pada algoritma sebelumnya telah dilakukan perhitungan *entropy* dan *information gain* dari seluruh data. Maka selanjutnya adalah melakukan perhitungan *split information* dan *gain ratio*.

$$\begin{aligned} \text{SplitInfo (Wilayah)} &= - \left(\left(\frac{139}{512} \right) \log_2 \frac{139}{512} + \left(\frac{30}{512} \right) \log_2 \frac{30}{512} + \left(\frac{343}{512} \right) \log_2 \frac{343}{512} \right) \\ &= -(-0,51068 + (-0,23983) + (-0,38717)) \\ &= 1,13768 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Administrasi)} &= - \left(\left(\frac{501}{512} \right) \log_2 \frac{501}{512} + \left(\frac{11}{512} \right) \log_2 \frac{11}{512} \right) \\ &= -(-0,03066 + (-0,11904)) \\ &= 0,14969 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (BTA)} &= - \left(\left(\frac{301}{512} \right) \log_2 \frac{301}{512} + \left(\frac{140}{512} \right) \log_2 \frac{140}{512} + \left(\frac{71}{512} \right) \log_2 \frac{71}{512} \right) \\ &= -(-0,45055 + (-0,51152) + (-0,39525)) \\ &= 1,35732 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Agama)} &= - \left(\left(\frac{512}{512} \right) \log_2 \frac{512}{512} \right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Merokok)} &= - \left(\left(\frac{94}{512} \right) \log_2 \frac{94}{512} + \left(\frac{418}{512} \right) \log_2 \frac{418}{512} \right) \\ &= -(-0,44896 + (-0,23891)) \end{aligned}$$

$$\begin{aligned}
 &= 0,68788 \\
 \text{SplitInfo (Kendaraan)} &= -\left(\left(\frac{147}{512}\right)\log_2 \frac{147}{512} + \left(\frac{365}{512}\right)\log_2 \frac{365}{512}\right) \\
 &= -(-0,51689 + (-0,34807)) \\
 &= 0,86496 \\
 \text{SplitInfo (Jenjang)} &= -\left(\left(\frac{160}{512}\right)\log_2 \frac{160}{512} + \left(\frac{217}{512}\right)\log_2 \frac{217}{512} + \left(\frac{135}{512}\right)\log_2 \frac{135}{512}\right) \\
 &= -(-0,5244 + (-0,52489) + (-0,50709)) \\
 &= 1,55638
 \end{aligned}$$

Setelah dilakukan perhitungan *split information*, maka selanjutnya menghitung *gain ratio* sebagai perhitungan akhir dari algoritma ini dan acuan dalam penentuan simpul akar.

$$\text{GainRatio (Wilayah)} = \frac{0,00582}{1,13768}$$

$$= 0,00511$$

$$\text{GainRatio (Administrasi)} = \frac{0,06344}{0,14969}$$

$$= 0,42377$$

$$\text{GainRatio (BTA)} = \frac{0,58075}{1,35732}$$

$$= 0,42786$$

$$\text{GainRatio (Agama)} = \frac{0,58075}{0,00000}$$

$$= \sim$$

$$\text{GainRatio (Merokok)} = \frac{0,43337}{0,68788}$$

$$= 0,63002$$

$$\text{GainRatio (Kendaraan)} = \frac{0,29388}{0,86496}$$

$$= 0,33976$$

$$\begin{aligned} \text{GainRatio (Jenjang)} &= \frac{0,02092}{1,55638} \\ &= 0,01344 \end{aligned}$$

Untuk dapat mengetahui nilai *gain ratio* terbesar lebih mudah, maka hasil-hasil dari perhitungan diatas ditampilkan pada tabel 4.8 berikut.

Tabel 4.8 Hasil Perhitungan *Split Information* dan *Gain Ratio*

		Jumlah Kasus	No	Yes	Entropy	Information Gain
Total		512	0,58075			
Wilayah				0,00582	1,13768	0,00511
	1	139	0,57551			
	2	30	0,21084			
	3	343	0,60655			
Administrasi				0,06344	0,14970	0,42377
	C	501	0,52867			
	I	11	0,00000			
BTA				0,58075	1,35732	0,42786
	H	301	0,00000			
	M	140	0,00000			
	A	71	0,00000			
Agama				0,58075	0,00000	~
	I	512	0,00000			
Merokok				0,43337	0,68788	0,63002
	T	94	0,80274			
	F	418	0,00000			
Kendaraan				0,29388	0,86496	0,33976
	H	147	0,99917			
	D	365	0,00000			
Jenjang				0,02092	1,55638	0,01344
	1	160	0,54356			
	2	217	0,71823			
	3	135	0,32450			

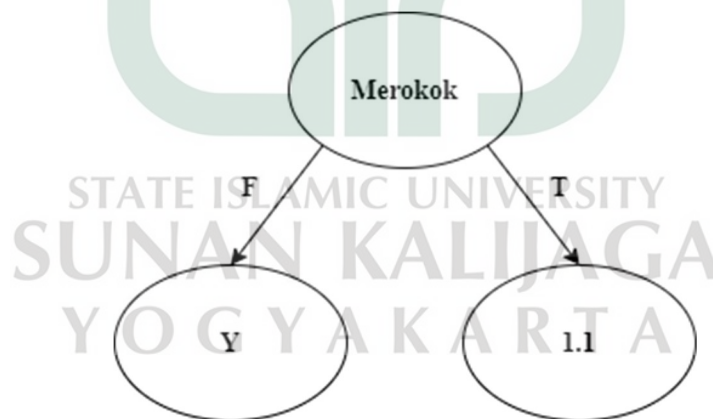
Dari hasil perhitungan diatas, terlihat bahwa atribut merokok memiliki *gain ratio* tertinggi yang berarti atribut tersebut menjadi simpul akar dari pohon

keputusan. Terdapat dua nilai dalam atribut merokok, yaitu T dan F yang dijadikan cabang. Untuk mengetahui pembagian jumlah data dari kedua nilai dan mengetahui apakah cabang berlanjut atau berhenti, maka dibawah ini ditampilkan jumlah data dari dua nilai atribut merokok terhadap klasifikasi.

Tabel 4.9 Jumlah Data Atribut Merokok Terhadap Klasifikasi

Atribut	Nilai	Jumlah Kasus	Diterima	Tidak
Merokok				
	T	94	23	71
	F	418	48	0

Dari tabel 4.10 dapat disimpulkan bahwa untuk nilai F berhenti karena semua mutlak masuk dalam klasifikasi diterima. Berbeda dengan nilai F, nilai T dapat dilakukan percabangan lagi karena adanya pembagia data dalam klasifikasi.



Gambar 4.3 Pohon Keputusan C4.5

4.3.2.2. Pembentukan Simpul Keputusan 1.1

Sama halnya dengan pembentukan akar, dalam pembentukan simpul ini juga dipilih yang memiliki *gain ratio* tertinggi. Hanya menjadi lebih spesifik yaitu

mengacu pada atribut merokok dengan nilai T. Sebelum dilakukan pembentukan simpul keputusan 1.1, perlu dilakukan perhitungan jumlah objek data seluruh atribut kecuali yang telah ditetapkan sebagai simpul akar terhadap klasifikasi. Jumlah objek data yang dihitung spesifik pada data dengan atribut merokok bernilai T atau berarti benar merokok.

Tabel 4.10 Jumlah Objek Data dengan Atribut Merokok Bernilai T

Atribut	Jumlah Kasus		Hasil Klasifikasi	
			Y	N
Total	94		23	71
Wilayah	1	25	6	19
	2	3	2	1
	3	66	15	51
Adm	C	83	23	60
	I	11	0	11
BTA	H	12	12	0
	M	11	11	0
	A	71	0	71
Agama	I	94	23	71
Kendaraan	H	71	0	71
	D	23	23	0
Jenjang	1	31	11	20
	2	49	6	43
	3	14	6	8

Setelah diketahui masing-masing jumlah data, maka selanjutnya dilakukan perhitungan *entropy* total dan masing-masing atribut dengan nilai atribut merokok bernilai T.

$$\begin{aligned}
 \text{Entropy (Total)} &= \left(\left(\frac{-23}{94} \right) \log_2 \frac{23}{94} \right) + \left(\left(\frac{-71}{94} \right) \log_2 \frac{71}{94} \right) \\
 &= (-0,24468 (-2,03103)) + (-0,75532 (-0,40484))
 \end{aligned}$$

$$= (2,27571) + (1,16016)$$

$$= 0,802738$$

Menghitung *entropy* atribut wilayah terhadap atribut merokok dengan nilai T.

$$\begin{aligned} \text{Entropy (Wilayah, 1)} &= \left(\left(\frac{-6}{25} \right) \log_2 \frac{6}{25} \right) + \left(\left(\frac{-19}{25} \right) \log_2 \frac{19}{25} \right) \\ &= (-0,24 (-2,05889)) + (-0,76 (-0,39593)) \\ &= (0,49413) + (0,30091) \\ &= 0,79504 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Wilayah, 2)} &= \left(\left(\frac{-2}{3} \right) \log_2 \frac{2}{3} \right) + \left(\left(\frac{-1}{25} \right) \log_2 \frac{1}{25} \right) \\ &= (-0,66667 (-0,58496)) + (-0,33333 (-1,58496)) \\ &= (0,38998) + (0,52832) \\ &= 0,91830 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Wilayah, 3)} &= \left(\left(\frac{-15}{66} \right) \log_2 \frac{15}{66} \right) + \left(\left(\frac{-51}{66} \right) \log_2 \frac{51}{66} \right) \\ &= (-0,22727 (-2,13750)) + (-0,77273 (-0,37197)) \\ &= (0,38998) + (0,52832) \\ &= 0,77323 \end{aligned}$$

Menghitung *entropy* atribut administrasi terhadap atribut merokok dengan nilai T.

$$\begin{aligned} \text{Entropy (Adm, C)} &= \left(\left(\frac{-23}{83} \right) \log_2 \frac{23}{83} \right) + \left(\left(\frac{-60}{83} \right) \log_2 \frac{60}{83} \right) \\ &= (-0,27711 (-1,85148)) + (-0,72289 (-0,46815)) \\ &= (0,51306) + (0,33842) \\ &= 0,85148 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Adm, I)} &= \left(\left(\frac{-0}{11} \right) \log_2 \frac{0}{11} \right) + \left(\left(\frac{-11}{11} \right) \log_2 \frac{11}{11} \right) \\ &= 0 \end{aligned}$$

Menghitung *entropy* atribut BTA terhadap atribut merokok dengan nilai T.

$$\text{Entropy (BTA, H)} = \left(\left(\frac{-12}{12} \right) \log_2 \frac{12}{12} \right) + \left(\left(\frac{-0}{12} \right) \log_2 \frac{0}{12} \right)$$

$$= 0$$

$$\text{Entropy (BTA, M)} = \left(\left(\frac{-11}{11} \right) \log_2 \frac{11}{11} \right) + \left(\left(\frac{-0}{11} \right) \log_2 \frac{0}{11} \right)$$

$$= 0$$

$$\text{Entropy (BTA, A)} = \left(\left(\frac{-0}{71} \right) \log_2 \frac{0}{71} \right) + \left(\left(\frac{-71}{71} \right) \log_2 \frac{71}{71} \right)$$

$$= 0$$

Menghitung *entropy* atribut agama terhadap atribut merokok dengan nilai T.

$$\text{Entropy (Agama, I)} = \left(\left(\frac{-23}{94} \right) \log_2 \frac{23}{94} \right) + \left(\left(\frac{-71}{94} \right) \log_2 \frac{71}{94} \right)$$

$$= (-0,24468 (-2,03103)) + (-0,75532 (-0,40484))$$

$$= (0,49695) + (0,30578)$$

$$= 0,80274$$

Menghitung *entropy* atribut kendaraan terhadap atribut merokok dengan nilai T.

$$\text{Entropy (Kendaraan, H)} = \left(\left(\frac{-0}{71} \right) \log_2 \frac{0}{71} \right) + \left(\left(\frac{-71}{71} \right) \log_2 \frac{71}{71} \right)$$

$$= 0$$

$$\text{Entropy (Kendaraan, D)} = \left(\left(\frac{-23}{23} \right) \log_2 \frac{23}{23} \right) + \left(\left(\frac{-0}{23} \right) \log_2 \frac{0}{23} \right)$$

$$= 0$$

Menghitung *entropy* atribut jenjang terhadap atribut merokok dengan nilai T.

$$\text{Entropy (Jenjang, I)} = \left(\left(\frac{-11}{31} \right) \log_2 \frac{11}{31} \right) + \left(\left(\frac{-20}{31} \right) \log_2 \frac{20}{31} \right)$$

$$= (-0,35484 (-1,49476)) + (-0,64516 (-0,63227))$$

$$= (0,53041) + (0,40791)$$

$$= 0,93832$$

$$\begin{aligned}
 \text{Entropy (Jenjang, 2)} &= \left(\left(\frac{-6}{49} \right) \log_2 \frac{6}{49} \right) + \left(\left(\frac{-43}{49} \right) \log_2 \frac{43}{49} \right) \\
 &= (-0,12245 (-3,02975)) + (-0,87755 (-0,18845)) \\
 &= (0,37099) + (0,16537) \\
 &= 0,53636
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy (Jenjang, 3)} &= \left(\left(\frac{-6}{14} \right) \log_2 \frac{6}{14} \right) + \left(\left(\frac{-8}{14} \right) \log_2 \frac{8}{14} \right) \\
 &= (-0,42857 (-1,22239)) + (-0,57143 (-0,80735)) \\
 &= (0,52388) + (0,46135) \\
 &= 0,98523
 \end{aligned}$$

Setelah didapatkan hasil *entropy*, maka selanjutnya adalah melakukan perhitungan *information gain*.

$$\begin{aligned}
 \text{Gain (Wilayah)} &= 0,802738 - \left(\left(\frac{25}{94} \right) 0,79504 + \left(\frac{3}{94} \right) 0,91830 + \left(\frac{66}{94} \right) 0,77323 \right) \\
 &= 0,80274 - 0,78366 \\
 &= 0,01908
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Administrasi)} &= 0,802738 - \left(\left(\frac{83}{94} \right) 0,85148 + \left(\frac{11}{94} \right) 0 \right) \\
 &= 0,80274 - 0,75184 \\
 &= 0,05090
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (BTA)} &= 0,802738 - \left(\left(\frac{12}{94} \right) 0 + \left(\frac{11}{94} \right) 0 + \left(\frac{71}{94} \right) 0 \right) \\
 &= 0,80274 - 0 \\
 &= 0,80274
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (Agama)} &= 0,802738 - \left(\left(\frac{94}{94} \right) 0,80274 \right) \\
 &= 0
 \end{aligned}$$

$$\text{Gain (Kendaraan)} = 0,802738 - \left(\left(\frac{71}{94} \right) 0 + \left(\frac{23}{94} \right) 0 \right)$$

$$= 0,80274$$

$$\begin{aligned} \text{Gain (Jenjang)} &= 0,802738 - \left(\left(\frac{31}{94} \right) 0,93832 + \left(\frac{49}{94} \right) 0,53636 + \left(\frac{14}{94} \right) 0,98523 \right) \\ &= 0,80274 - 0,73577 \\ &= 0,06697 \end{aligned}$$

Seluruh *information gain* dari seluruh atribut telah didapat, selanjutnya melakukan perhitungan *split information* untuk mendapatkan *gain ratio*.

$$\begin{aligned} \text{SplitInfo (Wilayah)} &= - \left(\left(\frac{25}{512} \right) \log_2 \frac{25}{512} + \left(\frac{3}{512} \right) \log_2 \frac{3}{512} + \left(\frac{66}{512} \right) \log_2 \frac{66}{512} \right) \\ &= -(-0,50817 + (-0,15861) + (-0,35822)) \\ &= 1,02500 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Administrasi)} &= - \left(\left(\frac{83}{94} \right) \log_2 \frac{83}{94} + \left(\frac{11}{94} \right) \log_2 \frac{11}{94} \right) \\ &= -(-0,15854 + (-0,3622)) \\ &= 0,52074 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (BTA)} &= - \left(\left(\frac{12}{94} \right) \log_2 \frac{12}{94} + \left(\frac{11}{94} \right) \log_2 \frac{11}{94} + \left(\frac{71}{94} \right) \log_2 \frac{71}{94} \right) \\ &= -(-0,3791 + (-0,3622) + (0,30578)) \\ &= -(-0,3791 + (-0,3622) + (0,30578)) \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Agama)} &= - \left(\left(\frac{94}{94} \right) \log_2 \frac{94}{94} \right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Kendaraan)} &= - \left(\left(\frac{71}{94} \right) \log_2 \frac{71}{94} + \left(\frac{23}{94} \right) \log_2 \frac{23}{94} \right) \\ &= -(-0,30578 + (-0,49695)) \\ &= 0,80274 \end{aligned}$$

$$\begin{aligned} \text{SplitInfo (Jenjang)} &= - \left(\left(\frac{31}{94} \right) \log_2 \frac{31}{94} + \left(\frac{49}{94} \right) \log_2 \frac{49}{94} + \left(\frac{14}{94} \right) \log_2 \frac{14}{94} \right) \\ &= -(-0,52779 + (-0,48994) + (-0,40916)) \end{aligned}$$

$$= 1,42689$$

Entropy, *information gain*, dan *split information* telah didapatkan. Maka tahap terakhir adalah mencari *gain ratio* untuk menentukan simpul 1.1.

$$\begin{aligned} \text{GainRatio (Wilayah)} &= \frac{0,01908}{1,02500} \\ &= 0,01861 \end{aligned}$$

$$\begin{aligned} \text{GainRatio (Administrasi)} &= \frac{0,05090}{0,52074} \\ &= 0,09774 \end{aligned}$$

$$\begin{aligned} \text{GainRatio (BTA)} &= \frac{0,80274}{1,04709} \\ &= 0,76664 \end{aligned}$$

$$\begin{aligned} \text{GainRatio (Agama)} &= \frac{0}{0} \\ &= \sim \end{aligned}$$

$$\begin{aligned} \text{GainRatio (Kendaraan)} &= \frac{0,80274}{0,80274} \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{GainRatio (Jenjang)} &= \frac{0,06697}{1,42689} \\ &= 0,04693 \end{aligned}$$

Hasil dari perhitungan *entropy*, *information gain*, *split information*, dan *gain ratio* diatas ditampilkan dalam tabel 4.11 dan 4.12 dibawah ini.

Tabel 4.11 Hasil Perhitungan *Entropy*, *Information Gain*, *Split Information*, dan *Gain Ratio*

		Jumlah Kasus	Entropy	Information Gain	Split Info	Gain Ratio
Total		94	0,802738			
Wilayah				0,01908	1,02500	0,01861
	1	25	0,79504			
	2	3	0,91830			

Tabel 4.12 Hasil Perhitungan *Entropy*, *Information Gain*, *Split Information*, dan *Gain Ratio* (Lanjutan)

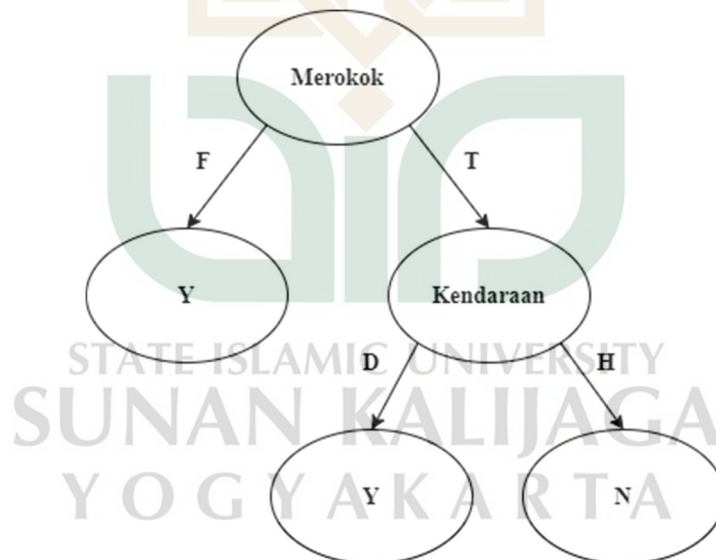
		Jumlah Kasus	Entropy	Information Gain	Split Info	Gain Ratio
	3	66	0,77323			
Adm				0,05090	0,52074	0,09774
	C	83	0,85148			
	I	11	0,00000			
BTA				0,80274	1,04709	0,76664
	H	12	0,00000			
	M	11	0,00000			
	A	71	0,00000			
Agama				0,00000	0,00000	~
	I	94	0,80274			
Kendaraan				0,80274	0,80274	1,00000
	H	25	0,79504			
	D	23	0,00000			
Jenjang				0,06697	1,42689	0,04693
	1	31	0,93832			
	2	49	0,53636			
	3	14	0,98523			

Berdasarkan perhitungan dalam tabel diatas dapat dilihat bahwa *gain ratio* tertinggi adalah atribut kendaraan, yaitu sebesar 1,00000. Oleh karena itu, atribut kendaraan dijadikan simpul keputusan 1.1. Atribut ini memiliki dua nilai yaitu H dan D yang menjadi cabang dari simpul keputusan kendaraan. Selanjutnya untuk mengetahui keberlanjutan dari simpul, maka perlu diperhatikan jumlah data masing-masing nilai pada atribut kendaraan terhadap klasifikasi.

Tabel 4.13 Jumlah Data Atribut Kendaraan Terhadap Klasifikasi

Atribut	Nilai	Jumlah Kasus	Diterima	Tidak
Kendaraan				
	H	71	71	0
	D	23	0	23

Dapat disimpulkan bahwa pohon keputusan berhenti pada simpul keputusan 1.1, karena masing-masing nilai secara keseluruhan memihak pada salah satu klasifikasi. Pada nilai H semua data terklasifikasi tidak diterima, sedangkan semua data pada nilai D terklasifikasi diterima. Dengan seperti itu, maka pohon keputusan dari algoritma C4.5 selesai dan digambarkan pada gambar 4.4 Dibawah ini.

**Gambar 4.4** Pohon Keputusan C4.5

4.4. Hasil dengan Perbandingan Data

Hasil dari penerapan algoritma pada data latih yang berupa *rule* atau aturan dipresentasikan dengan pohon keputusan. Untuk melihat *performance* dan keefektifan dari kedua algoritma ini diperlukan pengujian dengan data latih dengan

data uji menggunakan *confusion matrix* dan waktu eksekusi. Dalam penelitian ini dilakukan penambahan 250 data *dummy*. Contoh data *dummy* ditampilkan dalam tabel 4.15 dan 4.16 dibawah ini.

Tabel 4.14 Contoh Data Dummy

Hasil	Wilayah	Adm	BTA	Agama	Merokok	Kendaraan	Jenjang
Y	2	C	H	I	F	D	1
N	3	C	A	I	T	H	1
Y	1	C	H	I	F	D	2
Y	2	C	H	I	F	D	3
Y	3	C	H	I	F	D	2
Y	1	C	H	I	F	D	1
Y	2	C	H	I	F	H	2
Y	3	C	H	I	F	D	3
Y	1	C	H	I	F	D	3
Y	2	C	H	I	F	D	1
Y	3	C	H	I	F	D	2

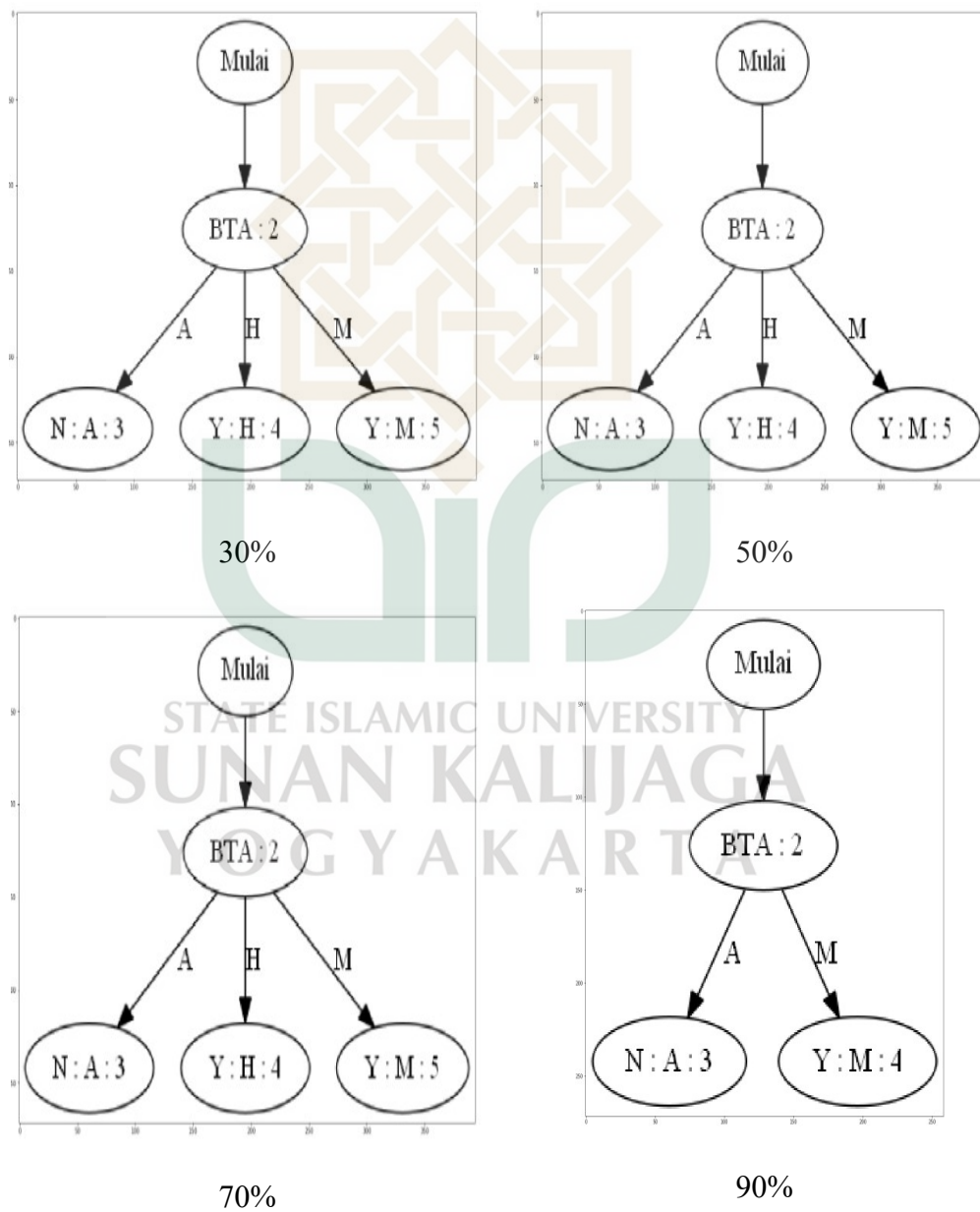
Pengujian dilakukan dalam beberapa bagian data. Dalam penelitian ini 512 dianggap sebagai data latih dan 250 data uji (data dummy). Maka data total adalah 762 data. Dari data tersebut dilakukan 4 pembagian untuk melihat tingkat *performance* dan efektif. Pembagian data dapat dilihat pada tabel 4.15 dibawah ini.

Tabel 4.15 Pembagian Data

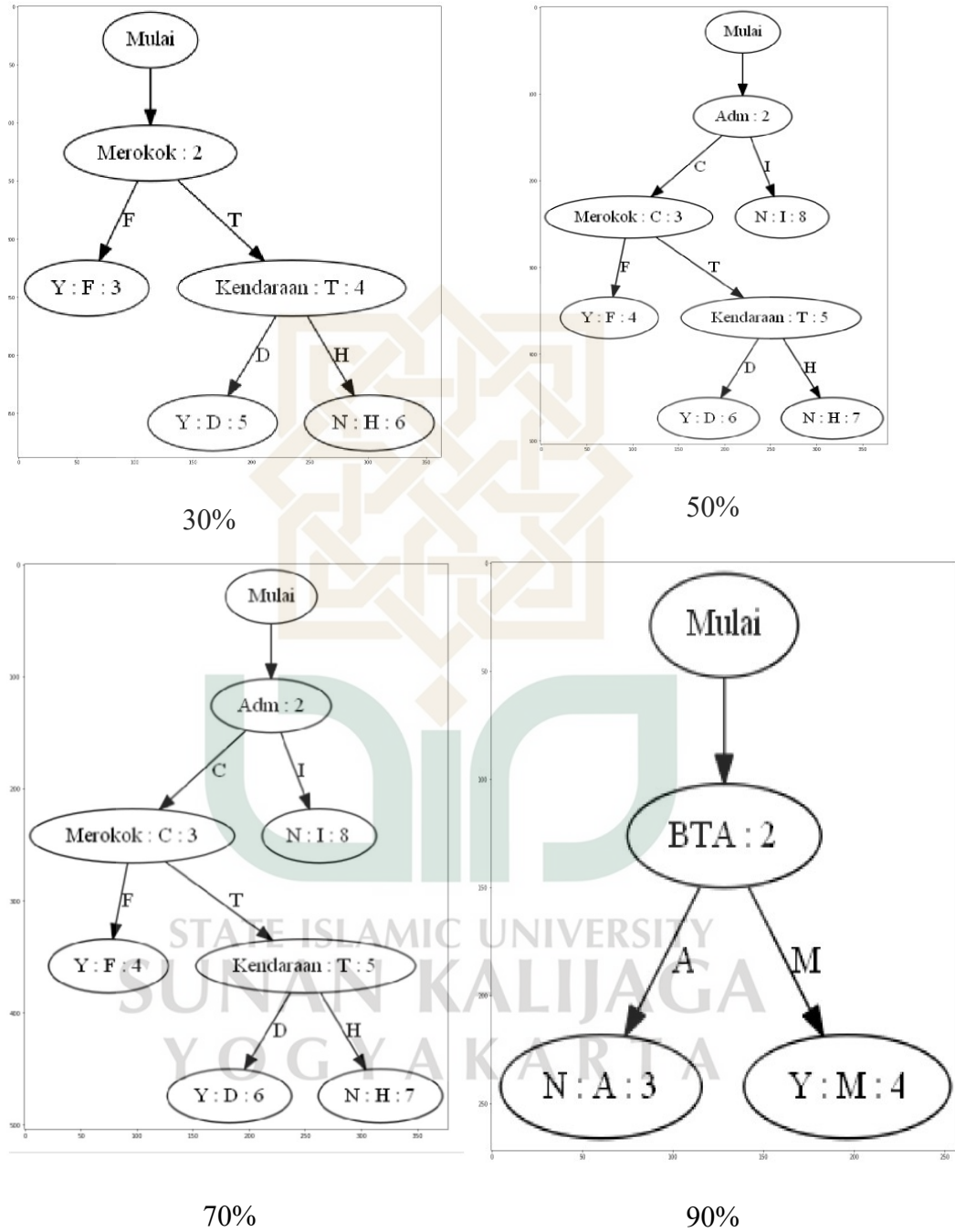
	Data Latih	Data Uji
30%	512	250
50%	381	381
70%	209	553
90%	77	685

4.4.1. Uji *Confusion Matrix*

Berdasarkan pengujian yang dilakukan 4 kali dengan perbandingan data uji dari algoritma *ID3* dan *C4.5* menghasilkan pohon keputusan dan *confusion matrix* yang berbeda. Maka dibawah ini ditampilkan pohon keputusan masing-masing algoritma dengan jumlah data uji yang berbeda.



Gambar 4.5 Pohon Keputusan *ID3* (4 data)



Gambar 4.6 Pohon Keputusan C4.5 (4 data)

Tabel 4.16 Confusion Matrix dengan Perbandingan Data Uji

Metode		Pengujian Data Uji (%)			
		30	50	70	90
ID3	Recall	100	100	100	100
	Precision	97,83	98,4	98,9	98,48
	Specificity	79,16	93,15	94,12	96,55
	Accuracy	98	98,69	99,06	98,93
C4.5	Recall	99,56	99,67	99,78	100
	Precision	100	100	100	98,48
	Specificity	100	100	100	96,55
	Accuracy	99,6	99,74	99,81	98,93

Pengujian dilakukan berdasarkan jumlah data uji. Pengujian data dilakukan mulai dari 30% sampai 90% dari 762 data (termasuk data *dummy*). Pada 4.16 dapat dilihat bahwa hasil pengujian tertinggi pada algoritma *ID3* untuk *recall* didapatkan pada seluruh pembagian data mencapai 100%, *precision* pada data 70% mencapai 98,9%, *specificity* pada data 90% mencapai 96,55%, dan *accuracy* pada data 70% mencapai 99,06%. Dan hasil pengujian tertinggi pada algoritma *C4.5* untuk *recall* didapatkan pada data 90% mencapai 100%, *precision* pada data 30%, 50%, dan 70% mencapai 100%, *specificity* pada data 30%, 50%, dan 70% mencapai 100%, dan *accuracy* pada data 70% mencapai 99,81%.

Dari pengujian algoritma *ID3* dan *C4.5* yang telah dilakukan membuktikan bahwa algoritma *C4.5* adalah algoritma yang memiliki tingkat *recall*, *precision*, dan *accuracy* yang tinggi.

4.4.2. Uji Waktu

Sama halnya dengan pengujian *confusion matrix*. Dalam pengujian waktu ini, akan menguji empat bagian data. Pembagian data sama dengan uji sebelumnya,

250 (30%), 381 (50%), 553 (70%), 685 (90%). Algoritma dengan waktu terpendek, maka dianggap sebagai algoritma paling efektif.

Tabel 4.17 Uji Waktu dengan Perbandingan Data Uji

Metode	Pengujian Data Uji (%)			
	30	50	70	90
ID3	01.835329	01.764932	0.362961	0.246897
C4.5	0.438028	0.553007	0.598286	0.346025

Dilihat dari tabel diatas, dari kedua algoritma di dua titik pembagian data pada 30% dan 50% algoritma C4.5 memiliki waktu lebih pendek yaitu 0.438028 dan 0.553007 dibandingkan algoritma *ID3* yang waktunya hingga 01.835329 dan 01.764932. Namun pada pembagian 70% dan 90% algoritma *ID3* memiliki waktu eksekusi yang lebih pendek yaitu 0.362961 dan 0.246897. Padahal pada C4.5 mencapai 0.598286 dan 0.346025. Disimpulkan bahwa waktu eksekusi yang terpendek dari kedua algoritma berada pada 90% data. Pada *ID3* memiliki waktu lebih pendek dari C4.5 dengan waktu eksekusi 0.246897 dan 0.346025. Maka disimpulkan bahwa algoritma yang paling efektif adalah algoritma *ID3*.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa dari pengujian *confusion matrix* algoritma C4.5 lebih baik dalam mengklasifikasikan penerimaan beasiswa dibandingkan *ID3* pada kasus penerimaan penerimaan Beasiswa Solopeduli tahun 2017/2018. Hal ini tampak pada nilai *recall*, *precision*, *specificity*, dan *accuracy* yang semua dicapai pada algoritma C4.5 yaitu *recall* dengan nilai 100% pada data 90%, *precision* 100% pada tiga titik pembagian data, *specificity* 100% pada tiga pembagian data, dan *accuracy* 99,81% pada 70% data. Sedangkan dari pengujian waktu eksekusi, algoritma *ID3* lebih efektif dibandingkan C4.5 pada kasus penerimaan penerimaan Beasiswa Solopeduli tahun 2017/2018 dengan hasil 0.246897 pada *ID3* dan 0.346025 pada C4.5 di data 90%.

5.2. Saran

Pada penelitian ini, peneliti tentu tidak jauh dari kekurangan. Oleh karena itu, terdapat beberapa saran untuk dijadikan perbaikan dari penelitian ini. Saran-saran tersebut meliputi :

1. Penelitian selanjutnya dapat dilakukan dengan melibatkan lebih banyak atribut agar didapat pohon keputusan yang lebih banyak menghasilkan simpul keputusan dengan artian akan semakin spesifik atau akurat dalam menentukan hasil klasifikasi.

2. Melakukan penelitian lanjutan dengan berbagai metode klasifikasi selain ID3 dan *C4.5* sehingga dapat meningkatkan hasil pengklasifikasian pada data beasiswa Yayasan Solopeduli.



DAFTAR PUSTAKA

- Agustina, D. M. (2016). Analisis Perbandingan Algoritma ID3 Dan C4. 5 Untuk Klasifikasi Penerima Hibah Pemasangan Air Minum Pada PDAM Kabupaten Kendal. *Journal of Applied Intelligent System*, 1(3), 234–244.
- Amin, R. K., Indwiarti, I., & Sibaroni, Y. (2015). Implementasi Klasifikasi Decision Tree Dengan Algoritma C4. 5 Dalam Pengambilan Keputusan Permohonan Kredit Oleh Debitur (Studi Kasus: Bank Pasar Daerah Istimewa Yogyakarta). *EProceedings of Engineering*, 2(1).
- Anggraeni, D. K., & Prabowo, Y. D. (2017). Pengembangan Aplikasi Decision Tree Iterative Dichotomiser 3 untuk Klasifikasi Positif atau Negatif Obesitas pada Balita. *Kalbiscientia*, 4.
- Efendi, M. S., & Wibawa, H. A. (2018). Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik. *JUITA: Jurnal Informatika*, 6(1), 29–35.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2, 36. <https://doi.org/10.15575/join.v2i1.71>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining. *Knowledge Discovery and Data Mining : Towards a Unifying Framework*, 87(1), 54–61. <https://doi.org/10.1511/1999.16.807>
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining concepts and techniques, third*

edition. Morgan Kaufmann Publishers. http://www.amazon.de/Data-Mining-Concepts-Techniques-Management/dp/0123814790/ref=tmm_hrd_title_0?ie=UTF8&qid=1366039033&sr=1-1

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*.

Elsevier.

Hermawati, F. A. (2013). Data Mining. 2013. *Andi: Yogyakarta*.

Hssina, B., Merbouha, A., Ezzikouri, H., & Erritali, M. (2014). A comparative study of decision tree ID3 and C4. 5. *International Journal of Advanced Computer Science and Applications*, 4(2), 0.

Iskandar, I., Hiryanto, L., & Hendryli, J. (2018). PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE C4. 5 DENGAN TEKNIK PRUNING. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 6(1), 64.

Kristanto, O. (2014). Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang. *Universitas Dian Nuswantoro, Semarang*.

Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises*. Dave Kuhlman Lutz.

Kusrini, E. T. L., & Taufiq, E. (2009). Algoritma data mining. *Yogyakarta: Andi Offset*.

Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. In *Discovering Knowledge in Data: An Introduction*

to Data Mining: Second Edition (Vol. 9780470908).

<https://doi.org/10.1002/9781118874059>

Maulana, A. (2016). Komparasi Algoritma Data Mining Untuk Akurasi Penentuan Beasiswa Kurang Mampu IAIN Syekh Nurjati Cirebon. *ITEj (Information Technology Engineering Journals)*, 1(2).

Romansyah, F., Sitanggang, I. S., & Nurdiati, S. (2009). Fuzzy Decision Tree dengan Algoritma ID3 pada Data Diabetes. *Internetworking Indonesia Journal*, 1(2), 1.

Sihombing, V. (2018). Klasifikasi Algoritma Iterative Dichotomizer (ID3) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, 1(2), 27–34.

Sihotang, H. (2005). Evaluasi Control Environment dalam Sistem Pengendalian Intern PT Bank Tabungan Negara (Persero). *Makalah Disampaikan Dalam Pendidikan SESPIBANK XL. LPPI. Jakarta*, 14–27.

Suharto, E., Widodo, A. P., & Suryono, S. (2019). Analyzing the Accuracy of Answer Sheet Data in Paper-based Test Using Decision Tree. *IJID (International Journal on Informatics for Development)*, 8(1), 1–7.

Susila, M. D. (2018). Penerapan Optimasi Algoritma C45 dengan Naïve Bayes pada Pemilihan Internet Service Provider. *Jurnal Eksplora Informatika*, 7(2), 16–26.

Suyanto, A. I. (2014). Searching, Reasoning, Planning, dan Learning (Revisi Kedua). *Bandung: Informatika Bandung*.

Tanjung, Y. P., Sentinuwo, S. R., & Jacobus, A. (2016). Penentuan Daya Listrik

Rumah Tangga Menggunakan Metode Decision Tree. *Jurnal Teknik Informatika*, 9(1).

Tyasti, A. E., Ispriyanti, D., & Hoyyi, A. (2015). Algoritma Iterative Dichotomiser 3 (Id3) Untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta). *Jurnal Gaussian*, 4(2), 237–246.

Utama, T. D. (2015). *Implementasi Algoritma Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha UNS*.

Zaki, M. J., & Wagner Meira, J. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.