

**SELEKSI CALON SISWA BARU DENGAN METODE  
*DECISION TREE* MENGGUNAKAN ALGORITMA C5.0  
(STUDI KASUS : SMK N. 1 DOLOKSANGGUL)**

**SKRIPSI**

**SRI ADELINA SITUMORANG**

**198160056**



**FAKULTAS TEKNIK**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**UNIVERSITAS MEDAN AREA**

**2024**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 7/11/24

Access From (repository.uma.ac.id)7/11/24

**SELEKSI CALON SISWA BARU DENGAN METODE  
DECISION TREE MENGGUNAKAN ALGORITMA C5.0  
(STUDI KASUS : SMK N. 1 DOLOKSANGGUL)**

**SKRIPSI**

*Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Memperoleh  
Gelar Sarjana Universitas Medan Area*



**Oleh :  
SRI ADELINA SITUMORANG  
198160056**

**FAKULTAS TEKNIK**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**UNIVERSITAS MEDAN AREA**

**2024**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 7/11/24

Access From (repository.uma.ac.id)7/11/24

## LEMBAR PENGESAHAN

Judul Skripsi : Seleksi Calon Siswa Baru Dengan Metode *Decision Tree*  
Menggunakan Algoritma C5.0 (Studi Kasus : SMK N.1  
Doloksanggul)  
Nama : Sri Adelina Situmorang  
NPM : 198160056  
Fakultas : Teknik  
Prodi : Teknik Informatika

Disetujui Oleh Komisi Pembimbing

Dosen Pembimbing

Andre Hasudungan Lubis, S.Ti, M.Sc

Mengetahui

Dekan Fakultas Teknik

  
Dede Satriano, S.T, M.T  
NIDN : 0102027402

Ketua Program Studi

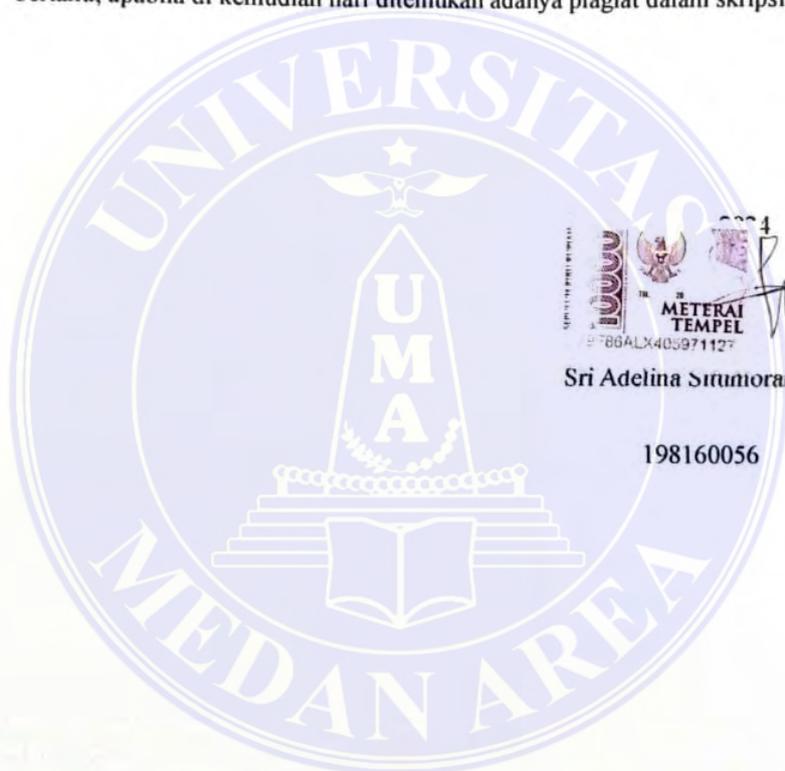
  
Rizki Muliono, S.Kom, M.Kom  
NIDN : 0109038902

Tanggal Lulus : 05 September 2024

iii

### HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi ini adalah hasil penelitian dan ide saya sendiri, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain yang telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah. Saya bersedia menerima sanksi akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.



Sri Adelina Situmorang

198160056

 Dipiindai dengan CamScanner

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS  
AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

---

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Sri Adelina Situmorang  
NPM : 198160056  
Program Studi : Teknik Informatika  
Fakultas : Fakultas Teknik  
Jenis karya : Tugas Akhir/Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Non-eksklusif (Non-Exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul : Seleksi calon siswa baru dengan metode *decision tree* menggunakan algoritma C5.0 (Studi kasus : SMK N 1 Doloksanggul).

Dengan Hak Bebas Royalti Non-ekskusif ini Universitas Medan Area berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir/skripsi/tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal :

Yang menyatakan

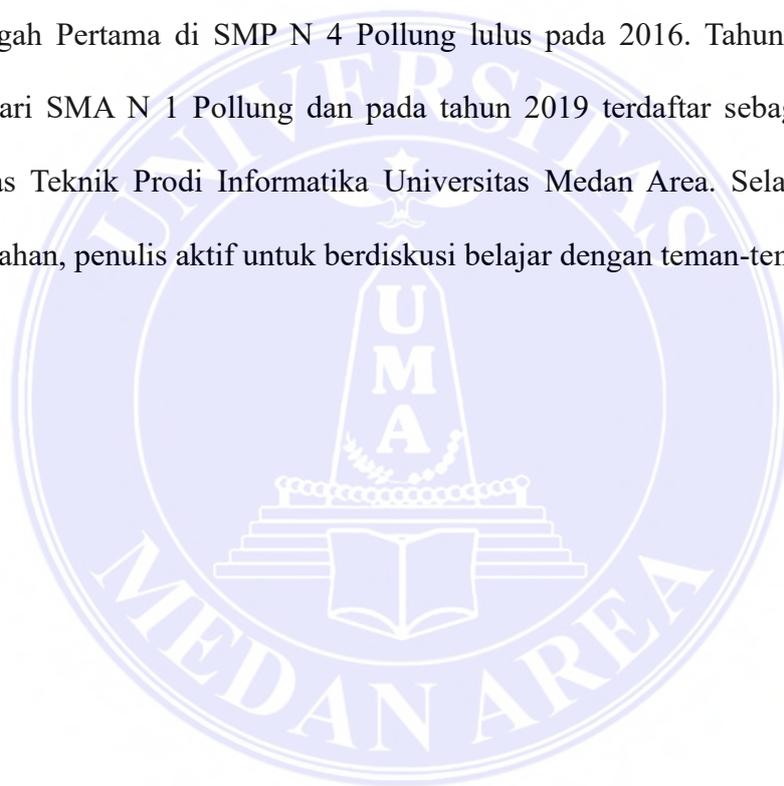


(Sri Adelina Situmorang)

 Dipindai dengan CamScanner

## RIWAYAT HIDUP

Sri Adelina Situmorang, lahir di Hutajulu pada tanggal 13 Mei 2001 dari pasangan orang tua ayah Ramli Situmorang dan ibu Permawan Lumban Gaol. Penulis merupakan anak ke 9 dari 9 bersaudara. Penulis pertama sekali sekolah di SD N 175782 Pollung pada tahun 2007-2013 dan kemudian melanjutkan Sekolah Menengah Pertama di SMP N 4 Pollung lulus pada 2016. Tahun 2019 Penulis lulus dari SMA N 1 Pollung dan pada tahun 2019 terdaftar sebagai mahasiswa Fakultas Teknik Prodi Informatika Universitas Medan Area. Selama mengikuti perkuliahan, penulis aktif untuk berdiskusi belajar dengan teman-teman.



## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan karunia-Nya. Sehingga penulis dapat menyelesaikan tugas akhir dengan judul “Seleksi Calon Siswa Baru Menggunakan Metode *Decision Tree* Dengan Algoritma C5.0 (Studi Kasus : SMK N. 1 Doloksanggul)” ini dengan sebaik-baiknya. Laporan tugas akhir/skripsi disusun berdasarkan hasil pengamatan dilapangan.

Penyusunan laporan tugas akhir/ skripsi ini merupakan syarat yang harus di tempuh untuk memenuhi kelulusan dalam menempuh Gelar Sarjana Jenjang Strata (S-1) sesuai dengan kurikulum Jurusan Teknik Informatika Universitas Medan Area yang berlaku pada saat ini.

Penyusunan laporan tugas akhir/skripsi ini tidak akan sesuai tanpa bimbingan, nasihat serta petunjuk dari berbagai pihak. Untuk itu, saya sebagai penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Kepada kedua orangtua penulis yang selalu setia mendukung dan mendoakan penulis dengan penuh cinta yang senantiasa berharap dalam doa dan dengan cinta mengajari, mendukung, menyemangati dan memenuhi kebutuhan penulis dalam perkuliahan selama menempuh pendidikan di Universitas Medan Area. Kepada Bapak Ramli Situmorang yang menjadi panutan dalam hidup penulis dan Ibu Permawan Lumban Gaol yang dengan penuh kesabaran dan kelembutan menjadi ibu dalam keluarga. Semoga penulis bisa menjadi

anak yang membanggakan untuk keluarga kita dan menjadi anak yang takut akan Tuhan.

2. Kepada abang, kakak, kakak ipar, abang ipar serta semua keponakan penulis, yang juga turut membantu penulis dalam mendukung dan mendoakan penulis dalam menyelesaikan tulisan ini. Penulis sangat berterimakasih kepadaNya karena menghadirkan mereka di dalam kehidupan penulis, Tuhan senantiasa memberkati keluarga kita.
3. Kepada ompung Tiorika Simanullang yang juga turut mendukung dan mendoakan penulis dengan penuh cinta dan dengan kesabaran memberi banyak nasihat kepada penulis.
4. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc selaku Rektor Universitas Medan Area.
5. Bapak Dr. Eng., Supriatno, S.T, M.T selaku Dekan Fakultas Teknik Universitas Medan Area.
6. Bapak Rizki Muliono, S.Kom, M.Kom selaku Kepala Program Studi Teknik Informatika.
7. Bapak Dosen Pembimbing yaitu Bapak Andre Hasudungan Lubis, S.Ti, M.Sc, yang dengan sabar memberikan waktu, pemikiran dan tenaganya untuk memberikan bimbingan, dukungan serta semangat kepada penulis dalam pengerjaan skripsi ini baik hingga tulisan ini dapat terselesaikan tepat pada waktunya.
8. Seluruh dosen dan staff Universitas Medan Area khususnya dosen prodi Teknik Informatika yang telah membagi ilmunya kepada penulis selama perkuliahan.

9. My roommate Tita Larose Togatorop yang senantiasa setia menemani dan memberi dukungan kepada penulis, terimakasih sudah mau berteman dengan penulis.
10. Diri saya sendiri yang telah berusaha semaksimal mungkin, terimakasih sudah bertahan, terimakasih karena tidak menyerah, dan terimakasih karena memilih untuk tetap hidup.
11. Tidak lupa penulis untuk berterimakasih pada Tuhan Yesus Kristus karena atas berkat kasih karuniaNya penulis bisa sampai di titik ini.
12. Orang-orang terdekat yang terkasih yang telah banyak mendukung dan menyemangati serta berbagai suka duka bersama penulis selama belajar di Universitas Medan Area.
13. Teman-teman seperjuangan dalam Bimbingan Skripsi Bapak Andre Hasudungan Lubis, S.Ti, M.Sc, yang telah berjuang bersama-sama dalam penyelesaian skripsi ini mulai dari proses awal hingga akhir.
14. Semua pihak yang tidak dapat disebut satu per satu yang juga mendukung, menyemangati dan mendoakan penulis agar menyelesaikan studi di Universitas Medan Area. Kiranya kasih setiaNya selalu menyertai kita dimana pun berada.

Dalam penulisan skripsi ini penulis juga menyadari keterbatasan penulis, oleh karena itu penulis memohon maaf atas segala kekurangan yang terdapat dalam tulisan ini. Penulis juga dengan terbuka mengharapkan kritik, saran dan sumbangan pemikiran. Agar tulisan ini menjadi lebih baik lagi, kiranya tulisan ini bermanfaat bagi pembaca. Tuhan Yesus memberkati.

Medan, September 2024

Hormat Saya,

Penulis



Sri Adelina Situmorang

198160056

## ABSTRAK

Penerimaan Peserta Didik Baru (PPDB) merupakan proses yang krusial dalam menentukan kualitas siswa di sebuah lembaga pendidikan. Salah satu permasalahan yang sering dihadapi dalam PPDB adalah kesulitan dalam menyeleksi calon siswa secara efektif dan efisien. Proses seleksi yang dilakukan secara manual membutuhkan waktu yang lama dan rentan terhadap kesalahan. Penelitian ini bertujuan untuk menganalisis dan mengembangkan model seleksi peserta didik baru menggunakan algoritma C5.0, yakni sebuah metode klasifikasi berbasis pohon keputusan (*decision tree*). Algoritma C5.0 dipilih karena memiliki keunggulan dalam hal kecepatan pemrosesan, kemampuan menangani data dalam jumlah besar, dan tingkat akurasi yang tinggi. Dataset yang digunakan dalam penelitian ini terdiri dari 1013 data calon siswa dari SMK N 1 Doloksanggul. Data tersebut mencakup atribut-atribut seperti poin zonasi, prestasi, afirmasi, perpindahan tugas orang tua, dan status diterima atau tidaknya siswa. Hasil penelitian menunjukkan bahwa algoritma C5.0 memiliki tingkat akurasi sebesar 99.01%, dengan presisi sebesar 89.47% dan recall 100%. Atribut Zonasi dan Prestasi memiliki peran signifikan dalam pengambilan keputusan dengan Information Gain tertinggi, masing-masing sebesar 0.352 dan 0.346. Sebaliknya, atribut Afirmasi dan Perpindahan Tugas Orang Tua memiliki kontribusi yang lebih kecil dalam proses klasifikasi. Dengan demikian, algoritma C5.0 terbukti sangat efektif dalam memprediksi status penerimaan calon siswa baru. Model ini diharapkan dapat membantu lembaga pendidikan dalam menyusun proses seleksi siswa yang lebih objektif dan efisien.

Kata Kunci: PPDB, Algoritma C5.0, *Decision Tree*, Zonasi, Prestasi, Klasifikasi.

## ABSTRACT

*The New Student Admission (PPDB) process is crucial in determining the quality of students in an educational institution. One common issue faced during PPDB is the difficulty of effectively and efficiently selecting prospective students. The manual selection process is time-consuming and prone to errors. This study aimed to analyze and develop a new student selection model using the C5.0 algorithm, a decision tree-based classification method. The C5.0 algorithm was chosen for its advantages in processing speed, ability to handle large datasets, and high accuracy levels. The dataset used in this research consisted of 1.013 prospective students from SMK N 1 Doloksanggul. The data included attributes such as zoning points, achievements, affirmation, parental job transfer, and acceptance status. The results indicated that the C5.0 algorithm achieved an accuracy rate of 99.01%, with a precision of 89.47% and recall of 100%. The attributes of Zoning and Achievement played significant roles in decision-making, with the highest Information Gain values of 0.352 and 0.346, respectively. Conversely, the attributes of Affirmation and Parental Job Transfer contributed less to the classification process. Thus, the C5.0 algorithm proved highly effective in predicting the acceptance status of new students. This model is expected to assist educational institutions in establishing a more objective and efficient student selection process.*

**Keywords:** PPDB, C5.0 Algorithm, Decision Tree, Zoning, Achievement, Classification.

 Diketahui dengan CamScanner

## DAFTAR ISI

<b>LEMBAR PENGESAHAN .....</b>	<b>Error! Bookmark not defined.</b>
<b>HALAMAN PERNYATAAN.....</b>	<b>iii</b>
<b>RIWAYAT HIDUP.....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>ABSTRAK .....</b>	<b>xi</b>
<b>ABSTRACT .....</b>	<b>Error! Bookmark not defined.</b>
<b>DAFTAR ISI.....</b>	<b>xiii</b>
<b>DAFTAR TABEL .....</b>	<b>xv</b>
<b>DAFTAR GAMBAR .....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah .....	7
1.4 Tujuan Penelitian .....	7
1.5 Manfaat Penelitian .....	7
1.6 Sistematika Penulisan .....	<b>Error! Bookmark not defined.</b>
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>9</b>
2.1 Seleksi Calon Siswa Baru.....	9
2.1.1 Jalur zonasi .....	10
2.2.2 Jalur Prestasi .....	10
2.1.3 Jalur Afirmasi.....	10
2.1.4 Jalur Perpindahan Tugas Orang Tua .....	11
2.2 Metode <i>Decision Tree</i> .....	11
2.3 Algoritma C5.0 .....	13
2.3.1 Perhitungan Algoritma C5.0 .....	13
2.5 <i>Python</i> .....	14
2.6 <i>Confusion Matrix</i> .....	15
2.7 Penelitian Terdahulu .....	16

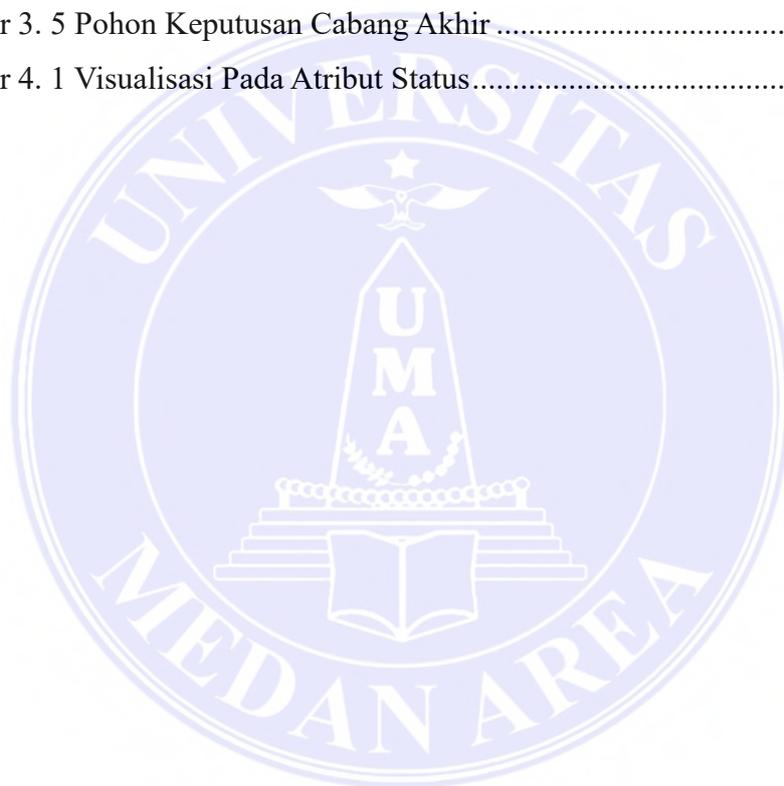
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>19</b>
3.1 Tahapan Penelitian .....	19
3.1.1 <i>Study Literatur</i> .....	20
3.1.2 Pengumpulan Data.....	20
3.1.3 Pengolahan Data .....	20
3.2 Perhitungan Manual .....	21
3.2.1 Pilih Akar Atribut.....	22
3.2.2 Cabang Untuk Setiap Nilai .....	27
3.2.3 Pembagian Kasus Setiap Cabang.....	28
3.2.4 Pembagian Kasus Pada <i>Node</i> 1.2.....	32
3.2.5 <i>Decision Tree</i> .....	36
3.2.6 <i>General Rule</i> .....	37
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>40</b>
4.1 Hasil .....	40
4.1.1 Proporsi Data .....	40
4.1.2 Evaluasi Algoritma C5.0 dan C4.5 .....	42
4.2 Pembahasan .....	45
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>45</b>
5.1 Kesimpulan .....	47
5.2 Saran .....	47
<b>DAFTAR PUSTAKA.....</b>	<b>48</b>
<b>LAMPIRAN.....</b>	<b>53</b>

## DAFTAR TABEL

	Halaman
Tabel 2. 1 Confusion Matrix (Normawati & Prayogi, 2021) .....	15
Tabel 2. 2 Penelitian Terdahulu .....	16
Tabel 3. 1 Data Siswa .....	21
Tabel 3. 2 Hasil Perhitungan Entropy, Gain dan Gain Ratio untuk Node 1 .....	26
Tabel 3. 3 Tabel Perhitungan Node 1.1 .....	31
Tabel 3. 4 Perhitungan Node 1.2 .....	35
Tabel 4.1 Rangkuman Data Siswa Baru .....	40
Tabel 4.2 Evaluasi Algoritma C5.0 Berdasarkan Metrik .....	42
Tabel 4.3 Confusion Matrix Menggunakan C5.0 .....	43
Tabel 4.4 Gain dan Entropy C5.0 dan C4.5 .....	43
Tabel 4. 5 Evaluasi Algoritma C4.5 Berdasarkan Metrik .....	44
Tabel 4.6 Confusion Matrix Menggunakan C4.5 .....	45

## DAFTAR GAMBAR

	halaman
Gambar 3. 1 Tahapan Penelitian .....	19
Gambar 3. 2 Pohon Keputusan Hasil Perhitngan Node 1 .....	27
Gambar 3. 3 Pohon Keputusan Node 1.1 .....	32
Gambar 3. 4 Pohon Keputusan Node 1.2 .....	36
Gambar 3. 5 Pohon Keputusan Cabang Akhir .....	37
Gambar 4. 1 Visualisasi Pada Atribut Status .....	41



## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Menurut Undang-Undang RI Pasal 5 No. 20 Tahun 2003, pendidikan sangat penting bagi setiap warga negara, karena berkaitan dengan hak dan kewajiban warga negara. Pasal 1 dan 5 berbunyi sebagai berikut: “Setiap warga negara mempunyai hak yang sama atas pendidikan yang bermutu dan setiap warga negara mempunyai kesempatan untuk meningkatkan pendidikan sepanjang hayat.” Menurut Ki Hajar Dewantara (Bapak Pendidikan Nasional Indonesia), pendidikan merupakan prasyarat bagi perkembangan kehidupan anak, sedangkan tujuannya adalah mengarahkan seluruh sumber daya alam anak, sehingga mereka sebagai manusia dan anggota masyarakat dapat melakukannya untuk mencapai tingkat keamanan dan kebahagiaan yang setinggi-tingginya. Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana dan proses pembelajaran agar peserta didik mewujudkan potensi dirinya dalam hal kekuatan *spiritual* keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang perlu dikembangkan dalam diri, masyarakat, bangsa dan negara (Berek dkk., 2023).

Perkembangan besar sedang terjadi di bidang pendidikan, artinya pendidikan juga akan mengalami perubahan di masa yang akan datang. Pemerintah mewajibkan lembaga pendidikan untuk menggunakan teknologi. Salah satunya adalah dalam Penerimaan Peserta Didik Baru (PPDB) (Maharani Aulia, 2022).

PPDB merupakan tahap pertama yang dilalui oleh siswa sebelum proses pembelajaran (Novitasari, 2023). PPDB dilaksanakan melalui seleksi calon siswa dengan syarat dan mekanisme tertentu. Pemilihan calon siswa baru sangat penting bagi keberhasilan akademik siswa. Metode seleksi yang baik dapat memastikan bahwa siswa yang diterima memiliki kesempatan untuk berhasil di sekolah. Oleh karena itu, metode yang efektif dalam menyeleksi siswa baru sangatlah penting.

Faktor seleksi siswa baru dilakukan dengan beberapa cara, seperti: jalur prestasi, zonasi, perpindahan tugas orang tua dan afirmasi. Prestasi adalah hasil dari tindakan yang dilakukan dan diciptakan baik secara individu maupun kelompok (Masrifah, 2020). Jalur prestasi (nilai) adalah sesuatu yang abstrak dan ideal. Nilai bukanlah objek yang berwujud, bukan fakta, bukan sekedar baik dan buruk yang perlu bukti empiris, melainkan apresiasi yang bersifat sukarela dan tidak sukarela. Dalam sistem pendidikan akan diberlakukan, sistem zonasi yaitu sistem penerimaan siswa baru dengan memfokuskan jarak antara sekolah dan tempat tinggal. Kegiatan siswa di luar sekolah, baik di sekolah maupun di tempat lain, dengan tujuan untuk memperluas pengetahuan, memahami keterkaitan antar mata pelajaran, memadukan keterampilan dan minat serta meningkatkan keimanan dan ketaqwaan siswa kepada Tuhan Yang Maha Esa serta kesadaran berbangsa dan bernegara memperkuat moral dan lainnya adalah definisi prestasi non akademik (Kenopan & Bangun, 2023). Jalur perpindahan tugas orang tua merupakan jalur yang disediakan bagi calon peserta didik yang mengikuti perpindahan tugas orang tua/wali yang dibuktikan dengan surat penugasan orang tua/wali. Jalur afirmasi merupakan program pemerintah yang memberikan pendidikan kepada anak-anak dari seluruh kalangan di Indonesia. Hal itu

disampaikan dalam Peraturan Mendikbud 51 Tahun 2018 dan Peraturan Menteri Pendidikan dan Kebudayaan 20 Tahun 2019.

Instansi atau lembaga pendidikan ini sering mengalami kendala dalam menyeleksi calon siswa baru (Puspita dkk., 2021). Permasalahan yang dihadapi dalam menyeleksi siswa baru seringkali tidak terorganisir dengan baik yang membuat data siswa tersebut sulit ditemukan. Oleh karena itu, pada penelitian ini dikembangkan model PPDB untuk mengatasi permasalahan tersebut. PPDB juga memiliki beberapa kendala, seperti masih banyak kekurangan dalam penanganan, pengolahan informasi yang masih dilakukan secara manual sehingga membutuhkan waktu yang cukup lama dan efisiensi informasi yang belum tinggi. Karena keterbatasan tersebut, diperlukan suatu algoritma yang mendukung PPDB. Algoritma yang digunakan untuk menyelesaikan kasus ini adalah algoritma C5.0.

Algoritma C5.0 merupakan metode klasifikasi *decision tree* karena menggunakan pohon atau struktur hirarki dalam pengambilan keputusan dan merupakan kelanjutan dari metode *decision tree* sebelumnya yaitu algoritma C4.5 (Karlitasari dkk., 2023). Algoritma C5.0 adalah algoritma yang dapat memproses masalah lebih cepat karena *multi-threaded*, tetapi sebaliknya menghasilkan pengklasifikasi yang persis sama. Selain itu juga dapat menghasilkan kumpulan aturan, dan peningkatannya cukup besar (Karlitasari dkk., 2023).

*Decision tree* merupakan salah satu metode yang paling umum digunakan, khususnya dalam klasifikasi data (Tangkelayuk, 2022). *Decision tree* adalah sistem klasifikasi penambangan data yang menggunakan pohon keputusan untuk memprediksi kelas data baru. Dalam kasus ini, metode *decision tree* sangat baik digunakan karena merupakan algoritma yang mudah diinterpretasikan, membantu

mengidentifikasi faktor-faktor yang paling penting dalam pengambilan keputusan. *Decision tree* membagi data menjadi subset dan membuat keputusan berdasarkan *subset* tersebut. *Decision tree* ini mengelompokkan data sampai setiap *subset* data homogen (Alamsyah dkk., 2023). Dalam lima tahun terakhir telah terjadi peningkatan yang signifikan dalam kualitas algoritma dan aplikasinya di berbagai industri. Beberapa penelitian berfokus pada modifikasi metode konstruksi pohon untuk meningkatkan akurasi. *Decision tree* juga digunakan dalam penelitian di bidang keamanan, pemasaran, dan kesehatan (Alamsyah dkk., 2023).

Penelitian sebelumnya (Rachmadi dkk., 2021) yang berjudul “Rekomendasi Jurusan Dengan Menggunakan *Decision Tree* Pada Sistem Penerimaan Peserta Didik Baru SMK Widya Dharma Turen”. Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma *decision tree* yang dibantu dengan metode pengembangan *waterfall* menghasilkan sistem proses Penerimaan Peserta Didik Baru (PPDB) berbasis web yang mengurangi efisiensi sekolah dalam menerima calon siswa yang memenuhi kriteria permintaan.

Pada penelitian (Pamungkas dkk., 2022) yang berjudul “Penerapan Metode *Fuzzy Tsukamoto* Untuk Penentuan Jurusan Sesuai Minat Bakat Penerimaan Siswa Baru Pada SMK”. Tujuan dari penelitian ini adalah untuk membantu siswa memilih jurusan yang sesuai dengan minat dan kemampuannya. Hasil dari penelitian ini adalah sebuah aplikasi untuk menentukan jurusan siswa berdasarkan minat, keterampilan dan nilai siswa dari sekolah sebelumnya.

Pada penelitian (Aesyid dkk., 2021) yang berjudul “Deteksi Dini Mahasiswa *Drop Out* Menggunakan C5.0”. Berdasarkan hasil pengukuran, diketahui bahwa

kombinasi karakteristik bahasa dan sosial, latar belakang pendidikan agama dan orang tua yang menjadi pendidik/pengajar, serta kemampuan membaca Alquran berkorelasi positif dengan keberhasilan pendidikan karakter. Nilai akurasi sebesar 60,91% menunjukkan bahwa algoritma *decision tree* C4.5 dapat digunakan untuk memprediksi tingkat keberhasilan pelatihan karakter. Oleh karena itu, diperlukan kajian yang jauh lebih dalam, terutama untuk mengetahui ciri-ciri yang sangat mempengaruhi keberhasilan pendidikan karakter, dengan harapan akurasi yang lebih baik. Serta efisiensi dalam pengelompokan atribut.

Pada penelitian (Hartianto & Erikson, 2022) yang berjudul “Optimasi Prediksi Kelulusan Mahasiswa dengan Menggunakan C5.0 dan *Regresi Linear*”. Hasil penelitian menunjukkan bahwa penggunaan algoritma C5.0 menghasilkan kinerja yang lebih baik dalam memprediksi kelulusan dengan akurasi sebesar 86%, sedangkan *regresi linier* menghasilkan akurasi sebesar 75%. Dalam hal ini, algoritma C5.0 juga memiliki kepekaan yang lebih tinggi dalam mengidentifikasi mahasiswa yang berpeluang gagal, yang sangat penting untuk mencegah mahasiswa putus kuliah. Hasil penelitian ini menunjukkan bahwa dengan menggunakan algoritma C5.0 dapat mengoptimalkan prediksi kelulusan dibandingkan dengan *regresi linier*. Menggunakan algoritma C5.0 dapat membantu universitas melakukan perencanaan akademik dan mengidentifikasi mahasiswa yang berisiko gagal untuk secara proaktif mencegah putus kuliah.

Terdapat algoritma *decision tree* dengan hasil yang baik yaitu algoritma C5.0. Algoritma C5.0 adalah proses *decision tree* yang mirip dengan Algoritma C4.5, dengan persamaan dibuat dalam perhitungan *entropy* dan *gain*. Ketika perhitungan algoritma C4.5 berhenti pada perhitungan *entropy* dan *gain*, perhitungan

dilanjutkan pada C5.0 untuk menghitung *rasio gain* menggunakan hasil perhitungan *entropy* dan *gain* yang telah dihitung sebelumnya. Keunggulan algoritma C5.0 terletak pada kemampuannya menangani masalah seperti *missing value* dan data dalam jumlah besar. Algoritma ini juga dapat mengolah data pelatihan dalam waktu singkat dan menggunakannya untuk menguji data. Algoritma C5.0 ini juga menyediakan metode akselerasi yang dapat meningkatkan akurasi.

Pada penelitian sebelumnya tentang algoritma C5.0 (Benediktus & Oetama, 2020b) yang berjudul “Algoritma Klasifikasi *Decision Tree C5.0* untuk Memprediksi Performa Akademik Siswa”. Berdasarkan penelitian ini, penggunaan algoritma *Decision Tree C5.0* sangat efektif dalam memprediksi prestasi akademik mahasiswa dengan tingkat akurasi hingga 71,667%.

Berdasarkan beberapa penelitian yang telah dijelaskan sebelumnya, penelitian ini menggunakan algoritma C5.0 dengan metode *decision tree* sebagai metode analisisnya, karena metode ini merupakan metode yang baik sebagai metode analisis.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan diatas, maka masalah yang akan dibahas yaitu Bagaimana melakukan seleksi peserta didik baru dengan menggunakan algoritma C5.0?

### 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini ialah:

1. Ruang lingkup penelitian dibatasi pada data calon siswa SMK N 1 Doloksanggul.
2. Mempertimbangkan faktor-faktor yang dapat diukur dan dikumpulkan seperti prestasi, zonasi, faktor sosial sebagai atribut, dan perpindahan tugas orang tua.
3. Penelitian ini akan membangun sebuah model untuk mengklasifikasi penerimaan peserta didik baru dengan menggunakan *algoritma C5.0*.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menganalisis seleksi calon siswa baru dengan metode *Decision Tree* menggunakan Algoritma C5.0.

### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Bagi Penulis
  - a. Untuk meningkatkan pemahaman atau kemampuan mahasiswa dalam menerapkan ilmu yang diperoleh selama belajar di perguruan tinggi.
  - b. Memenuhi salah satu syarat kelulusan strata satu (S1), Program studi Teknik Informatika Universitas Medan Area (UMA).

## 2. Bagi SMK N. 1 Doloksanggul

Hasil penelitian ini diharapkan dapat memberikan dasar pengambilan keputusan dalam pemilihan calon siswa baru sehingga dapat dilakukan secara objektif.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Seleksi Calon Siswa Baru

Secara umum, PPDB (Penerimaan Peserta Didik Baru) dapat diartikan sebagai proses administrasi tahunan untuk menyeleksi siswa berdasarkan hasil akademik untuk melanjutkan studi ke jenjang yang lebih tinggi dan merupakan proses yang krusial dalam menentukan kualitas siswa di sebuah lembaga pendidikan. Penerimaan Peserta Didik Baru (PPDB) merupakan proses penting dalam sistem pendidikan Indonesia (Pd, 2023). Tujuan seleksi siswa baru adalah untuk menyeleksi siswa baru yang memiliki kesiapan akademik yang cukup untuk mengikuti dan menyelesaikan pendidikan, sehingga dapat berperan sebagai pedoman seleksi siswa baru yang efektif untuk mencapai tujuan peningkatan mutu prestasi akademik siswa (Nasser dkk., 2021). Manfaat dan keuntungan penerapan metode *decision tree* dengan algoritma C5.0 yaitu *decision tree* memungkinkan untuk mengidentifikasi faktor yang paling berpengaruh dalam pemilihan siswa baru. Dengan bantuan analisis struktur pohon, dimungkinkan untuk menemukan cabang yang memiliki pengaruh signifikan terhadap keputusan pemilihan. Ini dapat membantu untuk memahami kualitas yang penting ketika memilih siswa baru. Berikut adalah variabel yang digunakan dalam menyeleksi calon siswa baru dalam penelitian ini.

### **2.1.1 Jalur zonasi**

Jalur zonasi adalah jarak antara sekolah dengan tempat tinggal calon siswa.(Maharani Aulia, 2022). Jalur penerimaan ini diatur oleh Permendikbud Nomor 44 Tahun 2019 oleh Kementerian Pendidikan dan Kebudayaan (Kemendikbud), dengan maksimal jarak zonasi antara rumah dan SMA/SMK pilihan adalah 9 sampai dengan 10 kilometer. Sementara itu, kuota penerimaan untuk jalur zonasi SMA/SMK ini adalah 50% dari daya tampung sekolah pilihan.

### **2.2.2 Jalur Prestasi**

Jalur prestasi merupakan jalur pendaftaran bagi calon siswa baru yang berprestasi dalam kompetisi akademik dan non akademik serta berdomisili di luar komunitas sekolah (Ansar, 2019). Daya tampung untuk jalur prestasi nilai akademik SMK sebanyak 25% dari daya tampung sekolah terdiri dari jalur prestasi nilai rapor 20%, prestasi hasil lomba akademik 2% dan prestasi hasil lomba non-akademik 3%. Jalur prestasi nilai rapor diperuntukkan bagi calon peserta didik baru SMK yang sistem penilaiannya merupakan total nilai rata-rata rapor SMP semester 1 sampai semester 5 dengan bobot 70% ditambah dengan nilai akreditasi sekolah dari SMP dengan bobot 30%.

### **2.1.3 Jalur Afirmasi**

Jalur afirmasi merupakan jalur pendaftaran bagi siswa baru yang masuk dari keluarga kurang mampu dan penyandang disabilitas (Siswadi dkk., 2023). Daya tampung jalur afirmasi sebesar 20% dari daya tampung sekolah yang terbagi

atas keluarga tidak mampu paling sedikit 17% dan penyandang disabilitas paling banyak 3%.

#### 2.1.4 Jalur Perpindahan Tugas Orang Tua

Jalur perpindahan tugas orang tua merupakan jalur yang disediakan bagi calon peserta didik yang mengikuti perpindahan tugas orang tua/wali yang dibuktikan dengan surat penugasan orang tua/wali. Jalur perpindahan tugas orang tua juga tidak optimal dilakukan karena tidak banyak perpindahan tugas orang tua.

### 2.2 Metode *Decision Tree*

*Database* adalah hal biasa di dunia digital saat ini. Ada banyak teknik yang dapat digunakan untuk *data mining*. Salah satu metode yang sering digunakan adalah metode klasifikasi data. Diantara berbagai teknik klasifikasi tersebut, *decision tree* ternyata menjadi populer karena metode ini cukup sederhana, ekspresif, efektif dan efisien (Arum dkk., 2022). Istilah *data mining* mengacu pada proses menemukan sekumpulan pola atau karakteristik yang menggambarkan dan membedakan kelas data untuk memprediksi data yang belum memiliki kelas data (Teknorama dkk., 2023).

*Decision tree* memiliki struktur pohon yang terdiri dari simpul akar (titik awal) yang merupakan simpul teratas dari simpul yang tidak memiliki input dan tidak dapat memiliki keluaran atau memiliki lebih dari satu keluaran, simpul internal (titik cabang) adalah simpul percabangan yang *input output* nya terdiri dari dua simpul dan satu daun (kepala).

Setiap simpul akar dari *decision tree* merepresentasikan keputusan pada

atribut, sedangkan simpul internal merepresentasikan keluaran atau hasil keputusan, dan simpul daun hanya memiliki masukan tanpa *output* (Studi dkk., 2022). Setelah *decision tree* dibangun, seperangkat aturan dapat dibuat untuk mengevaluasi keadaan agar tegangan sistem tetap stabil. Aturan ini dapat digunakan sebagai petunjuk pengoperasian bagi operator jaringan untuk memantau dan mengevaluasi stabilitas tegangan secara *real time* (Arum dkk., 2022).

Tujuan utama dari *decision tree* itu sendiri adalah untuk menemukan sekumpulan aturan keputusan yang dapat digunakan untuk memprediksi sekumpulan hasil (Arum dkk., 2022). Metode ini memiliki efisiensi yang relatif tinggi dan kompleksitas waktu yang relatif rendah, serta tidak menggunakan akurasi klasifikasi sebagai fungsi evaluasi. *Decision tree* banyak digunakan untuk memecahkan masalah di bidang representasi informasi, klasifikasi, optimasi kombinatorial, geometri komputasi, dll. Keuntungan utama menggunakan *decision tree* yang sangat populer adalah dapat menyederhanakan proses pengambilan keputusan yang kompleks dan memudahkan pembuat keputusan untuk menginterpretasikan solusi masalah. Selain itu, *decision tree* juga berguna dalam eksplorasi data karena mengungkap hubungan tersembunyi antara beberapa kandidat variabel *input* dan variabel target (Teknorama dkk., 2023). Seperti algoritma *data mining* pada umumnya, *decision tree* memiliki beberapa kelemahan. Jika beberapa kategori dan kriteria digunakan, mungkin ada tumpang tindih. Dalam hal ini, *decision tree* dapat lebih mudah dipengaruhi oleh sampel acak atau bantuan parsial yang diulang berkali-kali selama proses pengambilan keputusan (Arum dkk., 2022).

## 2.3 Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma yang merupakan penyempurnaan dari algoritma C4.5 yang menggunakan representasi berbentuk pohon yang setiap *node* mempresentasikan atribut kemudian cabang mempresentasikan nilai dari atribut dan memiliki yang namanya daun dimana fungsinya adalah kelas (Sungkar & Qurohman, 2021).

### 2.3.1 Perhitungan Algoritma C5.0

Tahapan dari algoritma c5.0 adalah sebagai berikut :

- a. Memilih atribut sebagai *node* akar,
- b. Buat cabang untuk masing-masing *node*,
- c. Membagi kasus dari tiap cabang,
- d. Ulangi langkah yang sama untuk setiap cabang hingga semua kasus memiliki kelas yang sama. Untuk menghitung nilai *entropy* dapat dihitung dengan persamaan :

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (2.1)$$

dengan :

$S$  : Nilai kasus

$n$  : Jumlah kelas pada variabel  $A_i$

$p_i$  : Proporsi dari  $S$  dan  $S_i$

Untuk menghitung nilai *gain* digunakan persamaan berikut :

$$Gain(A_i) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} + Entropy(A_{ij}) \quad (2.2)$$

dengan :

$S$  : Nilai kasus

$A_{ij}$  : Variabel ke  $i$  pada kategori ke  $j$

$A_i$  : Variabel ke  $i$

$m$  : Jumlah kategori pada variabel  $A$

$|S_i|$  : Jumlah kasus pada kategori ke  $i$

$|S|$  = jumlah kasus dalam  $S$

Setelah didapat nilai *entropy* dan *gain*, selanjutnya adalah menghitung nilai *gain ratio*. Rumus dasar dari perhitungan *gain ratio* adalah sebagai berikut :

$$\text{Gain Ratio} = \frac{\text{Gain}(A)}{\sum_{i=1}^n \text{Entropy}(S_i)} \quad (2.3)$$

dengan :

$\text{Gain}(A)$  : Nilai *gain* dari suatu variabel ke  $i$

$\sum_{i=1}^n \text{Entropy}(A_{ij})$  : Jumlah nilai *entropy* dalam suatu variabel ke  $i$  kategori ke  $j$ .

## 2.5 Python

*Python* adalah bahasa pemrograman berorientasi objek interaktif yang menyediakan struktur data tingkat tinggi (Heryanto & Harjanti, 2023). *Python* adalah bahasa pemrograman luas yang dirancang untuk meningkatkan keterbacaan kode sehingga sintaks yang digunakan lebih mudah dipahami. *Python* dianggap sebagai bahasa yang menggabungkan fitur dan kemampuan dengan sintaks kode yang jelas. Bahasa pemrograman *Python* dirancang khusus untuk memudahkan programmer membuat program yang baik dari segi efisiensi waktu,

kemudahan pengembangan dan kompatibilitas dengan sistem. *Python* dapat digunakan untuk membuat aplikasi mandiri atau (skrip) (Dina Aqmila, 2022).

### 2.6 Confusion Matrix

*Confusion Matrix* adalah tabel yang melaporkan klasifikasi bahan uji yang benar dan klasifikasi bahan uji yang salah (Normawati & Prayogi, 2021). Contoh *matriks confusion* klasifikasi biner ditunjukkan pada Tabel 2.1.

Tabel 2. 1 *Confusion Matrix* (Normawati & Prayogi, 2021)

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan:

TP ( <i>True Positive</i> )	=	jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
TN ( <i>True Negative</i> )	=	jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
FP ( <i>False Positive</i> )	=	jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1
FN ( <i>False Negative</i> )	=	jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, dan *recall* seperti berikut:

$$accuracy = \frac{TP+TN}{Total} \tag{2.3}$$

$$precision = \frac{TP}{TP+FP} \tag{2.4}$$

$$accuracy = \frac{TP}{TP+FN} \tag{2.5}$$

## 2.7 Penelitian Terdahulu

Hasil penelitian sebelumnya menjadi salah satu acuan dasar penelitian. Karena tugas penelitian sebelumnya adalah memperluas dan memperdalam teori yang digunakan dalam penelitian ini. Biasanya, penelitian sebelumnya adalah sumber penelitian yang kemudian peneliti coba bandingkan dengan penelitian saat ini. Hasil penelitian sebelumnya juga menjadi sumber inspirasi yang sangat membantu selama penelitian. Selain itu, peneliti dapat mengidentifikasi kekuatan dan kelemahan yang dapat dikembangkan. Adapun penelitian terdahulu dalam penelitian ini dapat dilihat pada Tabel 2. 2 sebagai berikut:

Tabel 2. 2 Penelitian Terdahulu

No	Nama peneliti dan tahun	Judul	Metode	Hasil Penelitian
1	Sulaiman Sinaga, Rahmat W. Sembiring, S. Sumarno. (2022)	Penerapan <i>Algoritma Naïve Bayes</i> untuk Klasifikasi Prediksi Penerimaan Siswa Baru.	Klasifikasi	Berdasarkan pemeriksaan terhadap 30 data uji yang diolah dengan <i>Rapid Miner</i> diperoleh akurasi sebesar 76,67% yaitu 26 siswa diterima dan 4 tidak diterima. Singkatnya, proses prediksi berupa klasifikasi dengan menggunakan teknik <i>Naive Bayes</i> dapat lebih cepat, akurat dan mencapai akurasi yang tinggi.
2	Suprpto, Fahrul Fahrezi J. T, Edora. (2023)	Prediksi Pengangkatan Karyawan Dengan Metode Klasifikasi <i>Algoritma C5.0</i>	Klasifikasi	Metode klasifikasi algoritma C5.0 dapat diprediksi dengan mengimplementasikan <i>decision tree</i> , yang dapat mempercepat akurasi pengambilan

				keputusan pelamar kerja dan karyawan. Skor klasifikasi menggunakan data pelatihan dan data pengujian untuk memprediksi pengangkatan karyawan menggunakan confusion matriks, dimana akurasi data pengujian adalah 91,00 - akurasi data pelatihan adalah 96,75%. Diukur terhadap kurva RC dari data pelatihan, dan memiliki akurasi klasifikasi yang sangat baik yaitu 0,989.
3	Ismail & M. Ilham.(2022)	Sistem Pendukung Keputusan Penerimaan Siswa Baru SMAN 7 WATANSOPPENG Menggunakan Metode <i>Simple Additive Weighting</i>	<i>Simple Additive Weighting</i>	Ketika menerapkan metode <i>simple additive weighting (SAW)</i> , opsi terbaik dipilih dari beberapa opsi yang ada, karena bobot setiap fitur ditentukan dalam proses klasifikasi untuk menghasilkan hasil yang lebih efisien. Hasil penelitian ini menunjukkan bahwa terdapat sistem pendukung keputusan penerimaan siswa baru di SMAN 7 Watansoppeng yang menggunakan metode <i>Simple Additive Weighting (SAW)</i> berbasis Web yang dapat mempermudah dalam pemilihan siswa baru.
4	Benediktus & Oetama. (2020)	<i>Algoritma Klasifikasi Decision Tree C5.0</i>	Klasifikasi	Penggunaan algoritma <i>Decision Tree C5.0</i> sangat efektif dalam

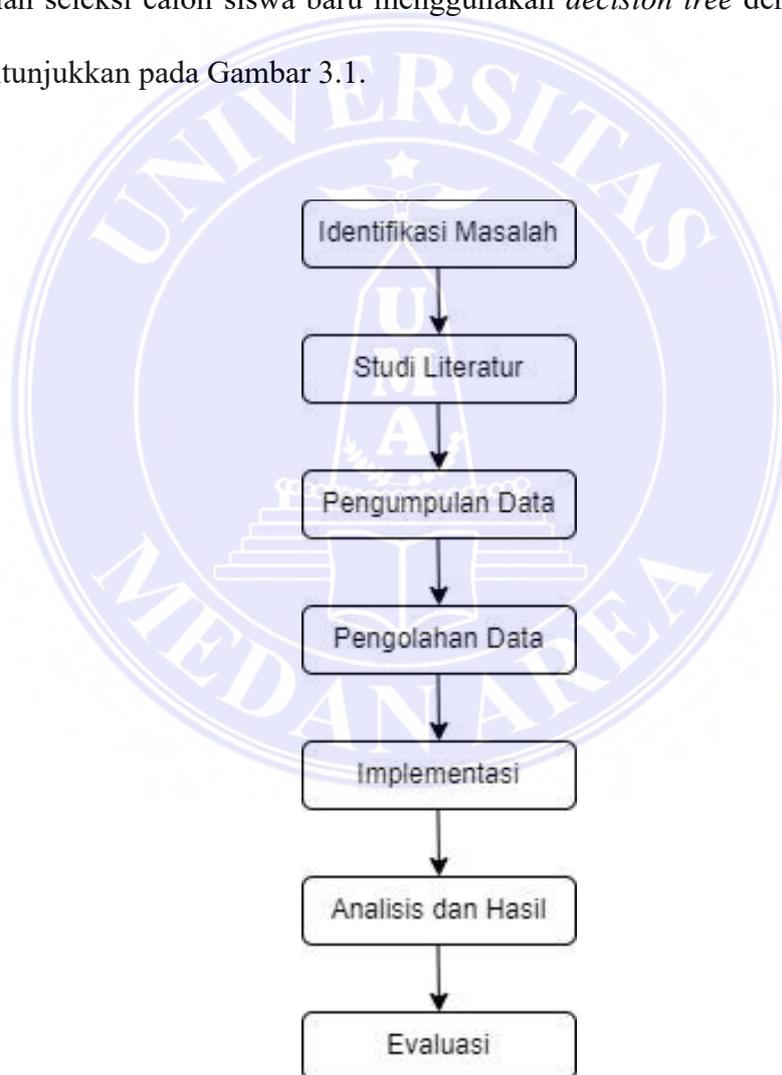
		untuk Memprediksi Performa Akademik Siswa		memprediksi prestasi akademik mahasiswa, mencapai akurasi sebesar 71,667%. Sehingga, <i>decision tree</i> yang disajikan menyatakan bahwa untuk mendapatkan nilai yang memuaskan, siswa harus benar-benar aktif dan selalu datang ke sekolah agar tidak terlambat masuk kelas. Namun, pasti ada siswa yang tidak aktif tetapi mendapat nilai memuaskan, atau sebaliknya, yang sangat aktif tetapi mendapat nilai kurang memuaskan. Namun penelitian ini menunjukkan bahwa mayoritas siswa aktif justru dapat mencapai nilai yang memuaskan.
5	Falda Junisman Zebua, Ribka Permatasari Br Manalu, Marlince Novita Karoseri Nababan. (2021)	Prediksi Kelulusan Mahasiswa Menggunakan Perbandingan Algoritma C5.0 dengan <i>Regression Linear</i>	<i>Regression Linear</i>	Berdasarkan masa studi, 651 (91%) siswa lulus ujian dengan benar, yang terdiri dari 427 laki-laki dan 224 perempuan (P), sedangkan 64 (9%) tidak lulus ujian (9%). 11 mahasiswi untuk musim 2017-2020. Hasil perbandingan nilai R2 algoritma C5.0 mencapai 96,85% (pelatihan) dan 93,72% (pengujian), dan nilai R2 <i>regresi linier</i> adalah 33,31% (pelatihan) dan 40,30% (pengujian).

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Tahapan Penelitian

Pada tahap penelitian, agar penelitian berjalan dengan baik. Prosedur penelitian seleksi calon siswa baru menggunakan *decision tree* dengan algoritma C5.0 ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Tahapan Penelitian

### 3.1.1 *Study Literatur*

Dengan mengumpulkan referensi dari penelitian terdahulu yang menggunakan metode yang sama maupun berbeda. *Study literatur* dari jurnal-jurnal terpercaya yang telah terindeksi *IEEE*, *ScienceDirect*, *SINTA*, maupun yang lainnya. Kemudian dilakukan *riview* pada *literatur* tersebut untuk menemukan topik penelitian.

### 3.1.2 *Pengumpulan Data*

Adapun yang dibahas pada sub bab ini adalah mengenai sumber dan jenis data serta metode pengumpulan data yang digunakan. Pada penelitian ini jenis data yang digunakan adalah data primer yang diperoleh langsung dari SMK N. 1 Doloksanggul. Data tersebut kemudian dibagi menjadi dua kategori pengumpulan data, yaitu:

- a. Data pelatihan yang digunakan untuk mengembangkan metode dan mengidentifikasi metode yang cocok dengan parameter yang tepat.
- b. Data testing digunakan untuk menguji hasil seleksi calon siswa baru dengan menggunakan metode yang diterapkan pada data *training*.

### 3.1.3 *Pengolahan Data*

Pengolahan data bertujuan untuk mengubah data mentah menjadi data yang berkualitas sehingga data layak untuk diolah pada tahapan selanjutnya. Tahapan ini dilakukan pada data mentah untuk menghilangkan data yang bermasalah atau *inkonsisten*. Data yang bermasalah seperti data yang niose atau yang mengandung *error*.

### 3.2 Perhitungan Manual

Berikut contoh kasus perhitungan manual yang menggunakan algoritma *decision tree* C5.0 untuk membangun *decision tree* data siswa yang akan dipilih sebagai dasar pemilihan calon siswa baru di SMK N. 1 Doloksanggul.

Tabel 3. 1 Data Siswa

No	Nama	Zonasi (m)	Prestasi	Afirmasi	Perindahan Orang Tua	Status
1	Bu	72	90	Ya	Tidak	Diterima
2	De	172	89	Tidak	Tidak	Diterima
3	Jo	198	76	Tidak	Tidak	Tidak Diterima
4	Fra	270	77	Tidak	Tidak	Tidak Diterima
5	Gre	313	85	Ya	Tidak	Diterima
6	Pa	355	80	Tidak	Tidak	Diterima
7	Se	1058	76	Ya	Tidak	Diterima
8	Ad	1156	81	Tidak	Tidak	Diterima
9	Kri	1780	79	Tidak	Tidak	Tidak Diterima
10	Ke	1877	75	Ya	Tidak	Tidak Diterima
11	Re	1896	75	Ya	Tidak	Tidak Diterima
12	Ni	2314	79	Tidak	Tidak	Tidak Diterima
13	Ke	2429	84	Tidak	Tidak	Diterima
14	Er	2543	77	Ya	Tidak	Diterima
15	Reg	2917	82	Tidak	Tidak	Diterima
16	Gi	3070	78	Ya	Ya	Diterima
17	En	3312	90	Tidak	Tidak	Diterima
18	So	4182	89	Ya	Tidak	Diterima
19	Yu	4290	83	Tidak	Ya	Diterima
20	Feb	5430	84	Tidak	Ya	Diterima

### 3.2.1 Pilih Akar Atribut

Ketika sebuah atribut ditentukan sebagai *root*, itu didasarkan pada nilai validasi tertinggi dan atribut yang ada. Hitung nilai *gain* dan *entropy* menggunakan rumus yang ditunjukkan pada Bab II. Dengan menggunakan Rumus 2.1 Bab II dapat dihasilkan nilai *entropy* dan dengan menggunakan Rumus 2.2 dapat dihasilkan nilai *gain* yang nantinya akan digunakan sebagai *root* dalam membangun *decision tree*.

#### 1. Entropy Total

$$\begin{aligned}
 Entropy\ Total(14,6) &= - \sum_{i=1}^2 P_i \log_2(P_i) \\
 &= \left( -\frac{14}{20} \log_2 \left( \frac{14}{20} \right) \right) + \left( -\frac{6}{20} \log_2 \left( \frac{6}{20} \right) \right) \\
 &= \left( -\frac{14}{20} \times \frac{\ln \left( \frac{14}{20} \right)}{\ln 2} \right) + \left( -\frac{6}{20} \times \frac{\ln \left( \frac{6}{20} \right)}{\ln 2} \right) \\
 &= 0.879
 \end{aligned}$$

*Entropy* total digunakan untuk menghitung nilai hasil yang diterima (14) dan tidak diterima (6), di mana nilai 20 adalah jumlah dari semua kasus.

#### 2. Entropy Atribut Zonasi

Untuk atribut zonasi terdapat 2 nilai, yaitu: “Zonasi  $\leq 2543$ ” dan “Zonasi  $> 2543$ ” dengan nilai *entropy* sebagai berikut :

*Entropy* Zonasi  $\leq 2543$

(Total kasus = 14, hasil diterima = 8, hasil tidak diterima = 6).

$$Entropy\ Zonasi\ \leq\ 2543 = \left( -\frac{8}{14} \log_2 \left( \frac{8}{14} \right) \right) + \left( -\frac{6}{14} \log_2 \left( \frac{6}{14} \right) \right)$$

$$= 0.983$$

*Entropy* Zonasi > 2543

(Total kasus = 6, hasil diterima = 6, dan hasil tidak diterima = 0).

$$\begin{aligned} \text{Entropy Zonasi} > 2543 &= \left(-\frac{6}{6} \log_2 \left(\frac{6}{6}\right)\right) + \left(-\frac{0}{6} \log_2 \left(\frac{0}{6}\right)\right) \\ &= 0 \end{aligned}$$

### 3. *Entropy* Atribut Prestasi

Untuk atribut prestasi terdiri dari 2 nilai, yaitu: “Prestasi ≤ 81”,

“Prestasi > 81” dengan nilai *entropy* sebagai berikut :

*Entropy* Prestasi ≤ 81

(Jumlah kasus = 11, hasil diterima = 5, hasil tidak diterima = 6).

$$\begin{aligned} \text{Entropy Prestasi} \leq 81 &= \left(-\frac{5}{11} \log_2 \left(\frac{5}{11}\right)\right) + \left(-\frac{6}{11} \log_2 \left(\frac{6}{11}\right)\right) \\ &= 1.095 \end{aligned}$$

*Entropy* Prestasi > 81

(Jumlah kasus = 9, hasil diterima = 9, hasil tidak diterima = 0).

$$\begin{aligned} \text{Entropy Prestasi} > 81 &= \left(-\frac{9}{9} \log_2 \left(\frac{9}{9}\right)\right) + \left(-\frac{0}{9} \log_2 \left(\frac{0}{9}\right)\right) \\ &= 0 \end{aligned}$$

### 4. *Entropy* Atribut Afirmasi

Untuk atribut afirmasi terdapat 2 nilai, yaitu: “Ya” dan “Tidak”

dengan nilai *entropy* keduanya ialah sebagai berikut :

*Entropy* Afirmasi Ya

(Jumlah kasus = 8, hasil diterima = 6, hasil tidak diterima = 2).

$$\begin{aligned} \text{Entropy Afirmasi Ya} &= \left(-\frac{6}{8} \log_2 \left(\frac{6}{8}\right)\right) + \left(-\frac{2}{8} \log_2 \left(\frac{2}{8}\right)\right) \\ &= 0.810 \end{aligned}$$

*Entropy Afirmasi Tidak*

(Jumlah kasus = 12, hasil diterima = 8, hasil tidak diterima = 4).

$$\begin{aligned} \text{Entropy Afirmasi Tidak} &= \left(-\frac{8}{12} \log_2 \left(\frac{8}{12}\right)\right) + \left(-\frac{4}{12} \log_2 \left(\frac{4}{12}\right)\right) \\ &= 0.915 \end{aligned}$$

5. *Entropy Atribut Perpindahan Tugas Orang Tua*

Untuk atribut perpindahan tugas orang tua terdapat 2 nilai, yaitu: “Ya” dan “Tidak” dengan nilai *entropy* keduanya ialah sebagai berikut :

*Entropy Perpindahan Tugas Orang Tua Ya*

(Jumlah kasus = 3, hasil diterima = 3, hasil tidak diterima = 0).

$$\begin{aligned} \text{Entropy Perpindahan Tugas Orang Tua Ya} \\ &= \left(-\frac{3}{3} \log_2 \left(\frac{3}{3}\right)\right) + \left(-\frac{0}{3} \log_2 \left(\frac{0}{3}\right)\right) \\ &= 0 \end{aligned}$$

*Entropy Afirmasi Tidak*

(Jumlah kasus = 17, hasil diterima = 11, hasil tidak diterima = 6).

*Entropy Perpindahan Tugas Orang Tua Tidak*

$$\begin{aligned} &= \left(-\frac{11}{17} \log_2 \left(\frac{11}{17}\right)\right) + \left(-\frac{6}{17} \log_2 \left(\frac{6}{17}\right)\right) \\ &= 0.658 \end{aligned}$$

Selanjutnya adalah melakukan perhitungan nilai *gain* di setiap atribut.

6. *Gain* (Total, Zonasi)

*Gain* untuk zonasi adalah :

$$= 0.879 - \left( \left( \frac{14}{20} \times 0.983 \right) + \left( \frac{6}{20} \times 0 \right) \right)$$

$$= 0,191$$

Dengan total = 20, jumlah kasus (zonasi  $\leq 2543 = 14$ , zonasi dan zonasi  $>2543 = 6$ ).

7. *Gain* (Total, Prestasi)

*Gain* untuk prestasi adalah :

$$= 0.879 - \left( \left( \frac{11}{20} \times 1.095 \right) + \left( \frac{9}{20} \times 0 \right) \right)$$

$$= 0.277$$

Dengan total = 20, jumlah kasus (prestasi  $\leq 81 = 11$  dan prestasi  $> 81 = 9$ ).

8. *Gain* (Total, Afirmasi)

*Gain* untuk afirmasi adalah :

$$= 0.879 - \left( \left( \frac{8}{20} \times 0.810 \right) + \left( \frac{12}{20} \times 0.915 \right) \right)$$

$$= 0,006$$

Dengan total = 20, jumlah kasus (afirmasi ya = 8 dan afirmasi tidak = 12).

9. *Gain* (Total, Perpindahan Tugas Orang Tua)

*Gain* untuk Perpindahan Tugas Orang Tua adalah :

$$= 0.879 - \left( \left( \frac{3}{20} \times 0 \right) + \left( \frac{17}{20} \times 0.658 \right) \right)$$

$$= 0,32$$

Dengan total = 20, jumlah kasus (perpindahan tugas orang tua ya = 3 dan perpindahan tugas orang tua tidak = 17).

Selanjutnya adalah melakukan perhitungan nilai *gain ratio* dari setiap atribut.

10. *Gain Ratio* (Zonasi)

*Gain Ratio* Zonasi adalah :

$$= \frac{0.191}{0.983 + 0} = 0.194$$

11. *Gain Ratio* (Prestasi)

*Gain Ratio* Prestasi adalah :

$$= \frac{0.277}{1.095 + 0} = 0.252$$

12. *Gain Ratio* (Afirmasi)

*Gain Ratio* Afirmasi adalah :

$$= \frac{0.006}{0.810+0.915} = 0.003$$

13. *Gain Ratio* (Perpindahan Tugas Orang Tua)

*Gain Ratio* Perpindahan Tugas Orang Tua adalah :

$$= \frac{0.32}{0+ 0.658} = 0.486$$

Adapun hasil perhitungan nilai *entropy*, *gain*, dan *gain ratio* secara lengkap disajikan pada Tabel 3 2 berikut:

Tabel 3. 2 Hasil Perhitungan Entropy, Gain dan Gain Ratio untuk Node 1

		Jumlah Kasus	Diterima	Tidak Diterima	Entropy
Total		20	14	6	0.879
Zonasi	<=2543	14	8	6	0.983
	> 2543	6	6	0	0
Gain		0.191			

<i>Gain Ratio</i>		0.194			
Prestasi	$\leq 81$	11	5	6	1.095
	$> 81$	9	9	0	0
<i>Gain</i>		0.277			
<i>Gain Ratio</i>		0.252			
Afirmasi	Ya	8	6	2	0.81
	Tidak	12	8	4	0.915
<i>Gain</i>		0.006			
<i>Gain Ratio</i>		0.003			
Perpindahan Tugas Orang Tua	Ya	3	3	0	0
	Tidak	17	11	6	0.658
<i>Gain</i>		0.32			
<i>Gain Ratio</i>		0.486			

Dengan hasil yang ada di Tabel 3. 2 dapat dilihat *gain ratio* tertinggi ialah perpindahan tugas orang tua, maka atribut prestasi dijadikan sebagai node akar.

### 3.2.2 Cabang Untuk Setiap Nilai

Pada atribut perpindahan tugas orang tua memiliki 2 nilai “ya” dan “tidak”. Dari nilai atribut ini, maka nilai “perpindahan tugas orang tua = ya” memiliki hasil = “diterima” sehingga tidak dibutuhkan perhitungan kembali.

Sedangkan “perpindahan tugas orang tua = tidak” memiliki hasil = “diterima” dan “tidak diterima” perlu dilakukan perhitungan kembali. Gambar *decision tree* sementara dapat dilihat pada Gambar 3 2.



Gambar 3. 2 Pohon Keputusan Hasil Perhitngan Node 1

Dari Gambar 3. 2 bisa dilihat yang dijadikan sebagai *node* akar ialah perpindahan tugas orang tua dimana di bagian ini dibagi 2 nilai ialah “perpindahan tugas orang tua = ya”, dan “perpindahan tugas orang tua = tidak”. Untuk nilai dari kelas “perpindahan tugas orang tua = ya” tidak dibutuhkan perhitungan lagi disebabkan hasilnya ialah “diterima”. Sedangkan untuk nilai “perpindahan tugas orang tua = tidak” diperlukan perhitungan kembali karena bernilai “diterima” dan “tidak diterima”.

### 3.2.3 Pembagian Kasus Setiap Cabang

Berikut untuk *node* 1 sebagai akar, sesuai dengan cara ada diatas maka menghitung kembali nilai *entropy* dari atribut yang masih ada yaitu status zonasi, prestasi, dan afirmasi. Setelah menghitung entropy, selanjutnya dihitung kembali *gain* dan *gain ratio* untuk setiap atribut.

#### 1. Entropy Atribut Zonasi

Untuk atribut zonasi terdapat 2 nilai, yaitu : zonasi  $\leq 2543$  dan zonasi  $> 2543$  dengan nilai entropynya sebagai berikut :

Zonasi  $\leq 2543$

(Total kasus = 14, hasil diterima = 8, hasil tidak diterima = 6).

$$\begin{aligned} \text{Entropy Zonasi } \leq 2543 &= \left( -\frac{8}{14} \log_2 \left( \frac{8}{14} \right) \right) + \left( -\frac{6}{14} \log_2 \left( \frac{6}{14} \right) \right) \\ &= 0.983 \end{aligned}$$

Zonasi  $> 2543$

(Total kasus = 3, hasil diterima = 3, hasil tidak diterima = 0).

$$\text{Entropy Zonasi } > 2543 = \left( -\frac{3}{3} \log_2 \left( \frac{3}{3} \right) \right) + \left( -\frac{0}{3} \log_2 \left( \frac{0}{3} \right) \right)$$

$$= 0$$

## 2. Entropy Atribut Prestasi

Untuk atribut prestasi terdapat 2 nilai, yaitu : prestasi  $\leq 81$  dan prestasi  $> 81$  dengan nilai *entropynya* sebagai berikut :

Prestasi  $\leq 81$

(Total kasus = 10, hasil diterima = 4, hasil tidak diterima = 6).

$$\begin{aligned} \text{Entropy Prestasi } \leq 81 &= \left( -\frac{4}{10} \log_2 \left( \frac{4}{10} \right) \right) + \left( -\frac{6}{10} \log_2 \left( \frac{6}{10} \right) \right) \\ &= 0.969 \end{aligned}$$

Prestasi  $> 81$

(Total kasus = 7, hasil diterima = 7, hasil tidak diterima = 0).

$$\begin{aligned} \text{Entropy Prestasi } > 81 &= \left( -\frac{7}{7} \log_2 \left( \frac{7}{7} \right) \right) + \left( -\frac{0}{7} \log_2 \left( \frac{0}{7} \right) \right) \\ &= 0 \end{aligned}$$

## 3. Entropy Atribut Afirmasi

Untuk atribut afirmasi terdapat 2 nilai, yaitu : ya dan tidak dengan nilai *entropynya* sebagai berikut :

Afirmasi Ya

(Total kasus = 7, hasil diterima = 5, hasil tidak diterima = 2).

$$\begin{aligned} \text{Entropy Afirmasi Ya} &= \left( -\frac{5}{7} \log_2 \left( \frac{5}{7} \right) \right) + \left( -\frac{2}{7} \log_2 \left( \frac{2}{7} \right) \right) \\ &= 0.859 \end{aligned}$$

Afirmasi Tidak

(Total kasus = 10, hasil diterima = 6, hasil tidak diterima = 4).

$$\begin{aligned} \text{Entropy Afirmasi Tidak} &= \left(-\frac{6}{10} \log_2 \left(\frac{6}{10}\right)\right) + \left(-\frac{4}{10} \log_2 \left(\frac{4}{10}\right)\right) \\ &= 0.969 \end{aligned}$$

Selanjutnya ialah melakukan perhitungan *gain* setiap atribut

4. *Gain* (Total, Zonasi)

*Gain* untuk zonasi adalah :

$$\begin{aligned} &= 0.658 - \left(\left(\frac{14}{17} \times 0.983\right) + \left(\frac{3}{17} \times 0\right)\right) \\ &= -0.151 \end{aligned}$$

Dengan total = 17, total kasus (zonasi  $\leq 2543 = 14$  dan zonasi  $> 2543 = 3$ ).

5. *Gain* (Total, Prestasi)

*Gain* untuk prestasi adalah :

$$\begin{aligned} &= 0.658 - \left(\left(\frac{10}{17} \times 0.969\right) + \left(\frac{7}{17} \times 0\right)\right) \\ &= 0.089 \end{aligned}$$

Dengan total = 17, total kasus (afirmasi 1 = 10 dan afirmasi 2 = 7).

6. *Gain* (Total, Afirmasi)

*Gain* untuk afirmasi adalah :

$$\begin{aligned} &= 0.658 - \left(\left(\frac{7}{17} \times 0.859\right) + \left(\frac{10}{17} \times 0.969\right)\right) \\ &= -0.264 \end{aligned}$$

Dengan total = 17, total kasus (afirmasi ya = 7 dan afirmasi tidak = 10).

Selanjutnya adalah melakukan perhitungan nilai *gain ratio* dari setiap atribut.

7. *Gain Ratio* (Zonasi)

*Gain Ratio* Zonasi adalah :

$$= \frac{-0.151}{0.983 + 0} = -0.153$$

8. *Gain Ratio* (Prestasi)

*Gain Ratio* Prestasi adalah :

$$= \frac{0.089}{0.969 + 0} = 0.091$$

9. *Gain Ratio* (Afirmasi)

*Gain Ratio* Afirmasi adalah :

$$= \frac{-0.264}{0.859 + 0.969} = -0.144$$

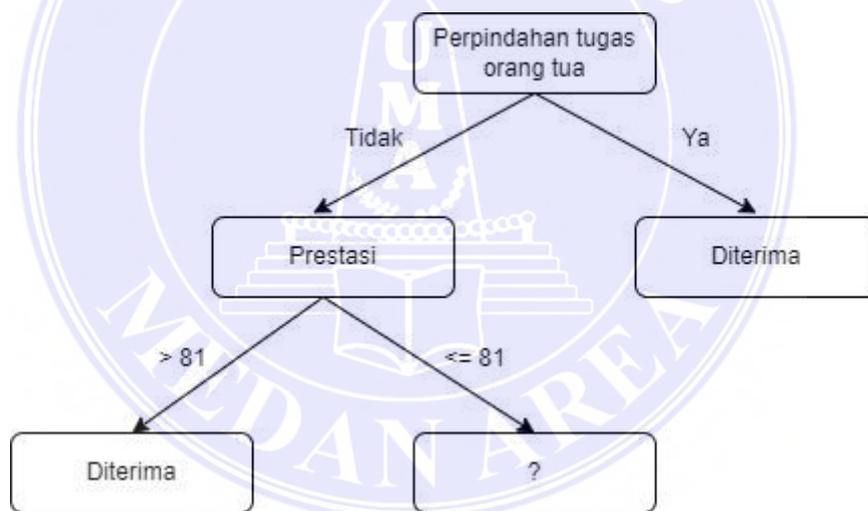
Setelah keseluruhan nilai *entropy*, *gain* dan *gain ratio* yang didapatkan, maka hasil dari perhitungan dibuat ke dalam Tabel 3.3.

Tabel 3.3 Tabel Perhitungan Node 1.1

		Jumlah Kasus	Diterima	Tidak Diterima	<i>Entropy</i>
<b>Total</b>		<b>17</b>	<b>11</b>	<b>6</b>	<b>0.658</b>
Zonasi	<= 2543	14	8	6	0.983
	> 2543	3	3	3	0
<i>Gain</i>		-0.151			
<i>Gain Ratio</i>		-0.153			
Prestasi	<= 81	10	4	6	0.969
	> 81	7	7	0	0
<i>Gain</i>		0.089			
<i>Gain Ratio</i>		0.091			
Afirmasi	Ya	7	5	2	0.859
	Tidak	10	6	4	0.969
<i>Gain</i>		-0.264			
<i>Gain Ratio</i>		-0.144			

Dari Tabel 3.3 maka dilihat bahwa atribut prestasi dengan nilai *gain ratio* tertinggi, yang berarti zonasi merupakan simpul terdekat. Ada 2 atribut dari prestasi yaitu “prestasi  $\leq 81$ ” dan “prestasi  $> 81$ ”. Nilai “prestasi  $> 81$ ” mempunyai hasil = “diterima” sehingga tidak dibutuhkan perhitungan kembali.

Sedangkan nilai “prestasi  $\leq 81$ ” memiliki hasil = “diterima” dan hasil “tidak diterima” sehingga nilai atribut “menikah” perlu dibutuhkannya perhitungan lagi. Setelah mendapatkan cabang kedua maka akan disambung dengan pohon keputusan sebelumnya yang masih belum terlihat keseluruhannya. Untuk melihat cabang pohon keputusan berikutnya dapat dilihat pada gambar pohon keputusan sementara dapat dilihat pada Gambar 3.3.



Gambar 3.3 Pohon Keputusan Node 1.1

### 3.2.4 Pembagian Kasus Pada Node 1.2

Berikutnya untuk *node* 1.2, sesuai dengan cara yang ada diatas maka menghitung kembali nilai *entropy* dari atribut yang masih ada yaitu zonasi dan afirmasi. Setelah menghitung *entropy*, selanjutnya dihitung kembali *gain* dan *gain ratio* untuk di setiap atribut.

1. *Entropy* Atribut Zonasi

Untuk atribut zonasi terdapat 2 nilai, yaitu : (zonasi  $\leq$  2543 dan zonasi  $>$  2543) dengan nilai *entropy*nya ialah sebagai berikut :

Zonasi  $\leq$ 2543

(Total kasus = 10, hasil diterima = 4, dan hasil tidak diterima = 6).

$$\begin{aligned} \text{Entropy Zonasi } \leq 2543 &= \left( -\frac{4}{10} \log_2 \left( \frac{4}{10} \right) \right) + \left( -\frac{6}{10} \log_2 \left( \frac{6}{10} \right) \right) \\ &= 0.969 \end{aligned}$$

Zonasi  $>$  2543

(Total kasus = 0, hasil diterima = 0, dan hasil tidak diterima = 0).

$$\begin{aligned} \text{Entropy Zonasi } > 2543 &= \left( -\frac{0}{0} \log_2 \left( \frac{0}{0} \right) \right) + \left( -\frac{0}{0} \log_2 \left( \frac{0}{0} \right) \right) \\ &= 0 \end{aligned}$$

2. *Entropy* Atribut Afirmasi

Untuk atribut afirmasi terdapat 2 nilai, yaitu: ya dan tidak dengan nilai *entropy*nya ialah sebagai berikut :

Afirmasi Ya

(Total kasus = 4, hasil diterima = 2, dan hasil tidak diterima = 2).

$$\begin{aligned} \text{Entropy Afirmasi Ya} &= \left( -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right) + \left( -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right) \\ &= 1 \end{aligned}$$

Afirmasi Tidak

(Total kasus = 6, hasil diterima = 2, dan hasil tidak diterima = 4).

$$\begin{aligned} \text{Entropy Afirmasi Tidak} &= \left( -\frac{2}{6} \log_2 \left( \frac{2}{6} \right) \right) + \left( -\frac{4}{6} \log_2 \left( \frac{4}{6} \right) \right) \\ &= 0.915 \end{aligned}$$

Selanjutnya adalah melakukan perhitungan nilai *gain* setiap atribut

3. *Gain* (Total, Zonasi)

*Gain* untuk zonasi adalah

$$= 0.969 - \left( \left( \frac{10}{10} \times 0.969 \right) + 0 \right)$$

$$= 0$$

Dengan total = 10, total kasus (zonasi  $\leq 2543 = 10$  dan zonasi  $> 2543 = 0$ ).

4. *Gain* (Total, Afiriasi)

*Gain* untuk afiriasi adalah

$$= 0.969 - \left( \left( \frac{4}{10} \times 1 \right) + \left( \frac{6}{10} \times 0.915 \right) \right)$$

$$= 0.02$$

Dengan total = 10, total kasus (afiriasi ya = 4 dan afiriasi tidak = 6).

Selanjutnya adalah melakukan perhitungan nilai *gain ratio* dari setiap atribut.

5. *Gain Ratio* (Zonasi)

*Gain Ratio* Zonasi adalah :

$$= \frac{0}{0.969} = 0$$

6. *Gain Ratio* (Afiriasi)

*Gain Ratio* Afiriasi adalah :

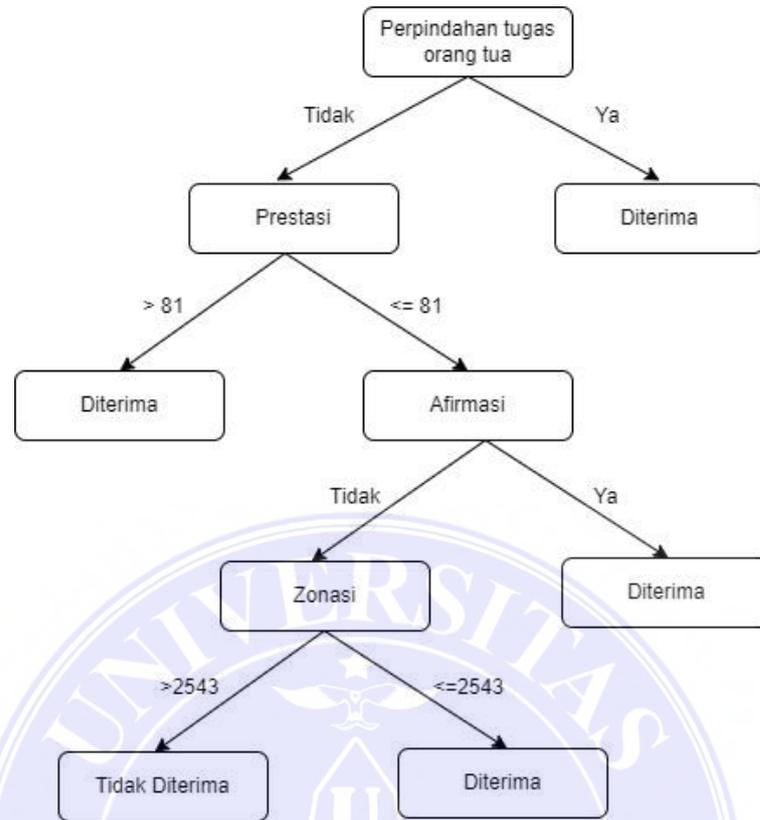
$$= \frac{0.02}{1 + 0.915} = 0.010$$

Setelah keseluruhan nilai *entropy*, *gain* dan *gain ratio* didapatkan, maka hasil dari perhitungan dibuat kedalam Tabel 3 4.

Tabel 3. 4 Perhitungan Node 1.2

		<b>Jumlah Kasus</b>	<b>Diterima</b>	<b>Tidak Diterima</b>	<b>Entropy</b>
<b>Total</b>		<b>10</b>	<b>4</b>	<b>6</b>	<b>0.969</b>
Zonasi	$\leq 2543$	10	4	6	0.969
	$> 2543$	0	0	0	0
<i>Gain</i>		0			
<i>Gain Ratio</i>		0			
Afirmasi	Ya	4	2	2	1
	Tidak	6	2	4	0.915
<i>Gain</i>		0.02			
<i>Gain Ratio</i>		0.010			

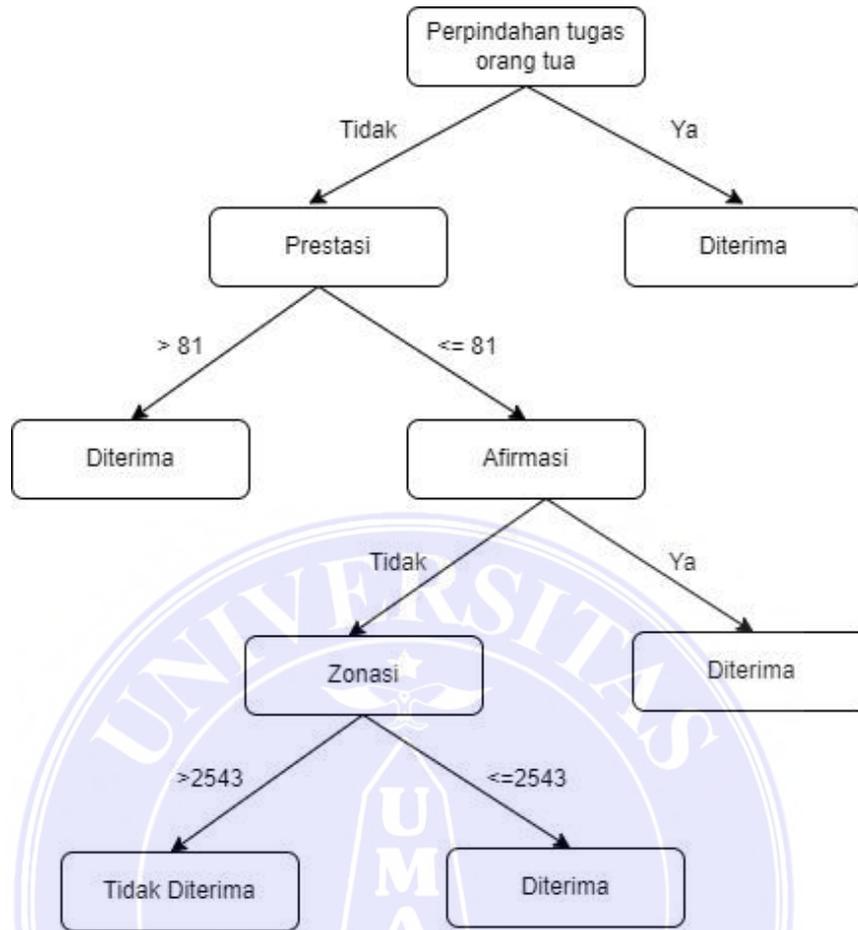
Dari Tabel 3.4 maka dilihat bahwa atribut afirmasi dengan nilai *gain ratio* tertinggi, yang berarti zonasi ialah *node* terakhir. Untuk afirmasi yang bernilai “ya” memiliki hasil “diterima” dan afirmasi tidak “tidak diterima” sehingga tidak dibutuhkan perhitungan kembali. Untuk zonasi yang bernilai  $>2543$  tidak memiliki kasus maka tidak diperlukan perhitungan sedangkan untuk zonasi  $\leq 2543$  memiliki hasil “tidak diterima” sehingga tidak dibutuhkan kembali perhitungan kembali. Untuk melihat cabang pohon keputusan dapat dilihat pada Gambar 3. 4.



Gambar 3. 4 Pohon Keputusan Node 1.2

### 3.2.5 Decision Tree

Pada perhitungan *entropy*, *gain* dan *gain ratio* di atas, maka diperoleh sebuah *decision tree* akhir pada Gambar 3. 5.



Gambar 3. 5 Pohon Keputusan Cabang Akhir

### 3.2.6 General Rule

Berdasarkan hasil *decision tree* yang terbentuk pada Gambar 3 5, dapat disimpulkan bahwa aturannya adalah sebagai berikut :

1. *If* perpindahan tugas orang tua = ya *then* diterima.
2. *If* perpindahan tugas orang tua = tidak *and* prestasi > 81 *then* diterima.
3. *If* perpindahan tugas orang tua = tidak *and* prestasi <= 81 *and* afirmasi = ya *then* diterima.
4. *If* perpindahan tugas orang tua = tidak *and* prestasi <= 81 *and* afirmasi = tidak *and* zonasi > 2543 *then* tidak diterima.

5. *If* perpindahan tugas orang tua = tidak *and* prestasi  $\leq 81$  *and* afirmasi = tidak *and* zonasi  $\leq 2543$  *then* diterima.

Keterangan :

1. Jika siswa memiliki sebuah nilai dengan perpindahan tugas orang tua = ya, maka akan dihasilkan nilai diterima.
2. Jika siswa memiliki nilai perpindahan tugas orang tua = tidak, maka dapat dilihat di nilai atribut berikutnya yakni prestasi (beralaskan perhitungan *gain ratio* tertinggi setelah perpindahan tugas orang tua). Apabila dengan kasus ini siswa mempunyai nilai prestasi  $> 81$ , maka hasilnya diterima.
3. Jika siswa memiliki nilai perpindahan tugas orang tua = tidak, maka dapat dilihat di nilai atribut berikutnya yakni prestasi (beralaskan perhitungan *gain ratio* tertinggi setelah perpindahan tugas orang tua). Apabila dengan kasus ini siswa mempunyai nilai prestasi  $\leq 81$ . Kemudian akan dilihat kembali nilai atribut berikutnya, yaitu afirmasi = ya, maka hasil diterima.
4. Jika siswa memiliki nilai perpindahan tugas orang tua = tidak, maka dapat dilihat di nilai atribut berikutnya yakni prestasi (beralaskan perhitungan *gain ratio* tertinggi setelah perpindahan tugas orang tua). Apabila dengan kasus ini siswa mempunyai nilai prestasi  $\leq 81$ . Kemudian akan dilihat kembali nilai atribut berikutnya, yaitu afirmasi = tidak dan atribut terakhir yaitu zonasi  $\leq 2543$ , maka hasilnya diterima.

5. Jika siswa memiliki nilai perpindahan tugas orang tua = tidak, maka dapat dilihat di nilai atribut berikutnya yakni prestasi (beralaskan perhitungan *gain ratio* tertinggi setelah perpindahan tugas orang tua). Apabila dengan kasus ini siswa mempunyai nilai prestasi  $\leq 81$ . Kemudian akan dilihat kembali nilai atribut berikutnya, yaitu afirmasi  $\neq$  tidak dan atribut terakhir yaitu zonasi  $> 2543$ , maka hasilnya tidak diterima.
6. Untuk setiap atribut, *entropy*, *gain* dan *gain ratio* harus dihitung secara terpisah untuk memeriksa nilai *gain ratio* maksimum untuk membangkitkan simpul akar (lihat Gambar 3 2). *Decision tree* tersebut sesuai dengan informasi yang diperoleh, sehingga *decision tree* tersebut dapat dijadikan dasar untuk mengolah siswa dalam seleksi calon siswa baru di SMK N. 1 Doloksanggul.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil pembahasan yang telah dilakukan, dapat disimpulkan bahwa Algoritma C5.0 menunjukkan performa yang sangat baik dalam proses seleksi peserta didik baru (PPDB). Dengan akurasi 99.01%, presisi 89.47%, dan *recall* 100%, C5.0 sangat efektif dalam memprediksi kelas dengan tingkat kesalahan yang sangat rendah. Dalam model C5.0, atribut Zonasi dan Prestasi memiliki peran yang sangat signifikan dalam pengambilan keputusan. Zonasi memiliki *Information Gain* tertinggi (0.352), diikuti oleh Prestasi (0.346), yang menunjukkan bahwa kedua atribut ini sangat penting dalam mengurangi ketidakpastian prediksi. Sebaliknya, atribut Afirmasi dan Perpindahan Tugas Orang Tua memiliki kontribusi yang sangat kecil. Algoritma C5.0 menggunakan teknik *pruning* untuk menghindari *overfitting*, yang membantu dalam menjaga keakuratan model dan memastikan hasil yang stabil.

#### 5.2 Saran

Adapun saran untuk menyempurnakan penelitian ini adalah penelitian selanjutnya dapat menambah data yang lebih beragam untuk meningkatkan generalisasi model. Diperlukan pengembangan lebih lanjut pada atribut-atribut lain yang dapat memberikan informasi tambahan dalam seleksi peserta didik baru.

## DAFTAR PUSTAKA

- Adriansa, M., Yulianti, L., & Elfianty, L. (2022). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 2(2), 115–121. <https://doi.org/10.54367/jtiust.v7i1.1983>
- Aesy, U. S., Lahitani, A. R., Diwangkara, T. W., & Kurniawan, R. T. (2021). Deteksi Dini Mahasiswa Drop Out Menggunakan C5.0. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(2), 113–119. <https://doi.org/10.14421/jiska.2021.6.2.113-119>
- Alamsyah, M. F., Satriawan, T. P., & Ramadanis, F. N. (2023). Analisa Komparasi Algoritma Naïve Bayes , Decision Tree Dan KKN Untuk Klasifikasi Kebakaran Hutan Pada Wilayah Aljazair. 1(2).
- Ansar, N. R. (2019). Implementasi Manajemen Penerimaan Peserta Didik Baru Berbasis Online di SMK Negeri 6 Makassar. *Jurnal Dinamika Manajemen Pendidikan*, 4(1), 65. <https://doi.org/10.26740/jdmp.v4n1.p65-72>
- Arum, L., Putri, B., Matematika, P. S., Sains, F., Teknologi, D. A. N., Islam, U., Maulana, N., & Ibrahim, M. (2022). *KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS SEKOLAH MENGGUNAKAN METODE DECISION TREE ID3 KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS SEKOLAH MENGGUNAKAN METODE DECISION TREE ID3*.
- Benediktus, N., & Oetama, R. S. (2020a). Algoritma Klasifikasi Decision Tree C5.0 untuk Memprediksi Performa Akademik Siswa. *Ultimatics*, XII(1), 14–19.
- Benediktus, N., & Oetama, R. S. (2020b). The Decision Tree C5.0 Classification

- Algorithm for Predicting Student Academic Performance. *Ultimatics : Jurnal Teknik Informatika*, 12(1), 14–19. <https://doi.org/10.31937/ti.v12i1.1506>
- Berek, P. A. L., Cristiana, Y., Sanan, U., Fouk, M. F. W. A., Dana, E., Rohi, F. R., & Orte, J. S. (2023). *Jurnal Ilmiah Pendidikan Citra Bakti BELAJAR DENGAN PRESTASI AKADEMIK MAHASISWA*. 10, 106–118.
- Budiarto, E., Rino, R., Hariyanto, S., & Susilawati, D. (2022). Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4. 5. *Algor*, 2. <https://jurnal.buddhidharma.ac.id/index.php/algor/article/view/1019%0Ahttps://jurnal.buddhidharma.ac.id/index.php/algor/article/download/1019/638>
- Dina Aqmila. (2022). Perancangan Media Pembelajaran Bahasa Pemograman Python Menggunakan Aplikasi SCRATCH Untuk Siswa Sekolah Menengah Pertama (SMP). *Skripsi*.
- Fitrianah, D., Gunawan, W., & Sari, A. P. (2022). Studi Komparasi Algoritma Klasifikasi C5.0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir. *Techno.Com*, 21(1), 1–11. <https://doi.org/10.33633/tc.v21i1.5348>
- Hartianto, D., & Erikson, H. (2022). *Optimasi Prediksi Kelulusan Mahasiswa dengan Menggunakan C5 . 0 dan*. 2, 22–32.
- Heryanto, Y., & Harjanti, T. W. (2023). *Analisis Perbandingan Ruang dan Waktu pada Algoritma Sorting Menggunakan Bahasa Pemrograman Python*. 4(2), 342–347.
- Ismail, & M. Ilham. (2022). Sistem Pendukung Keputusan Penerimaan Siswa Baru Sman 7 Watansoppeng Menggunakan Metode Simple Additive Weighting. *Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika (JISTI)*,

5(1), 29–36. <https://doi.org/10.57093/jisti.v5i1.106>

Karlitasari, L., Sriyasa, I. W., Wahyudi, I., & Santosi, H. B. (2023). Prediksi Morfologi Jamur Menggunakan Algoritma C5.0. *Jurnal Teknoinfo*, 17(1), 271. <https://doi.org/10.33365/jti.v17i1.2372>

Maharani Aulia, P. (2022). *Implementasi Kebijakan Penerimaan Peserta Didik Baru (PPDB) Melalui Sistem Zonasi Pada Jenjang SMA Negeri Di Kota Padang Sumatera Barat Tahun Ajaran 2021 / 2022*.

Masrifah, R. (2020). Peran Profesionalitas Guru PAI dalam Meningkatkan Prestasi Belajar. *Jurnal Al-Hikmah*, 22(2), 66–79.

Nasser, A. A., Arifudin, O., Barlian, U. C., & Sauri, S. (2021). Sistem Penerimaan Siswa Baru Berbasis Web Dalam Meningkatkan Mutu Siswa Di Era Pandemi. *Biormatika: Jurnal Ilmiah Fakultas Keguruan Dan Ilmu Pendidikan*, 7(1), 100–109. <https://doi.org/10.35569/biormatika.v7i1.965>

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711. <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>

Novitasari, U. D. (2023). *Strategi promosi madrasah dalam proses penerimaan peserta didik baru pada masa pandemi covid-19 di man 2 kota madiun*.

Nugraha D Z. (2023). Implementasi Data Mining Untuk Menentukan Pola Penjualan Di Armada Computer Menggunakan Algoritma Apriori. *Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(1), 25–31.

Pamungkas, P. D., Remawati, D., & Widada, B. (2022). Penerapan Metode Fuzzy Tsukamoto Untuk Penentuan Jurusan Sesuai Minat Bakat Penerimaan Siswa

- Baru Pada SMK. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, 10(1). <https://doi.org/10.30646/tikomsin.v10i1.599>
- Puspita, K., Alkhalifi, Y., & Basri, H. (2021). Rancang Bangun Sistem Informasi Penerimaan Peserta Didik Baru Berbasis Website Dengan Metode Spiral. *Paradigma - Jurnal Komputer Dan Informatika*, 23(1), 35–42. <https://doi.org/10.31294/p.v23i1.10434>
- Rachmadi, L. N., Wibawa, A. P., & Pujiyanto, U. (2021). Rekomendasi Jurusan Dengan Menggunakan Decision Tree Pada Sistem Penerimaan Peserta Didik Baru SMK Widya Dharma Turen. *Belantika Pendidikan*, 4(1), 29–36. <https://doi.org/10.47213/bp.v4i1.95>
- Sinaga, S., Sembiring, R. W., & Sumarno, S. (2022). Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru. *Journal of Machine ...*, 1(1), 55–64. <https://journal.fkpt.org/index.php/malda/article/view/162%0Ahttps://journal.fkpt.org/index.php/malda/article/download/162/115>
- Siswadi, G. A., Filsafat, F., & Gadjah, U. (2023). *MERDEKA BELAJAR DALAM PEMBELAJARAN AGAMA HINDU*. 3(1), 23–32.
- Studi, P., Informasi, S., Teknik, F., & Batam, U. P. (2022). *PREDICTION TINGKAT POTENSI PEMBELI*.
- Sungkar, M. S., & Qurohman, M. T. (2021). Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer. *Jurnal Media Informatika Budidarma*, 5(3), 1166. <https://doi.org/10.30865/mib.v5i3.3116>
- Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode

KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119.  
<https://doi.org/10.35957/jatisi.v9i2.2048>

Teknorama, J., El, T., & Rahma, S. El. (2023). *PENERAPAN DATA MINING UNTUK MENGANALISIS KEPUASAN PESERTA PROGRAM INDONESIA BISA BACA QURAN MENGGUNAKAN ALGORITMA DECISION TREE ( C4 . 5 ) BERBASIS*. 1(1), 14–20.

Yudhana, A., Sunardi, S., & Hartanta, A. J. S. (2020). Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggajian Kayu Sengon. *Transmisi*, 22(4), 123–129. <https://doi.org/10.14710/transmisi.22.4.123-129>

Zebua, F. J., Br Manalu, R. P., & Nababan, M. N. K. (2021). Prediksi Kelulusan Mahasiswa Menggunakan Perbandingan Algoritma C5.0 Dengan Regression Linear. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 4(2), 230.  
<https://doi.org/10.37600/tekinkom.v4i2.400>

## LAMPIRAN

### Lampiran Kode Program

```
import pandas as pd

# Membaca data
file_path = '/content/dok.csv'
data = pd.read_csv(file_path)
data

# Visualisasi dari jumlah kelas

import matplotlib.pyplot as plt

data = data['Status'].value_counts()
data.plot(kind='pie', autopct='%0.2f%%')
plt.show()

from sklearn.preprocessing import LabelEncoder

# Encode categorical variables
label_encoders = {}
for column in ['Afirmasi', 'Perpindahan Tugas Orang Tua', 'Status']:
    le = LabelEncoder()
    data[column] = le.fit_transform(data[column])
    label_encoders[column] = le

data

# Split the data into features and target variable
X = data.drop(columns=['Nama', 'Status'])
y = data['Status']

from sklearn.model_selection import train_test_split

# Membagi data menjadi set pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

## Algoritma C5

```

from sklearn.metrics import accuracy_score, precision_score, recall_score, confusion_matrix
from tree_c5 import C5

# Inisialisasi dan latih model C5
c5_model = C5(depth=3, min_samples_split=10, max_leaf_nodes=5)
c5_model.fit(X_train, y_train)

# Prediksi menggunakan model yang telah dilatih
y_pred_c5 = c5_model.predict(X_test)

# Evaluasi hasil
accuracy_c5 = accuracy_score(y_test, y_pred_c5)
precision_c5 = precision_score(y_test, y_pred_c5)
recall_c5 = recall_score(y_test, y_pred_c5)
conf_matrix_c5 = confusion_matrix(y_test, y_pred_c5)

# Menampilkan hasil
print(f"Akurasi: {accuracy_c5}")
print(f"Presisi: {precision_c5}")
print(f"Recall: {recall_c5}")
print(f"Confusion Matrix:\n{conf_matrix_c5}")

from graphviz import Digraph

def visualize_tree(tree, filename='decision_tree'):
    """Visualisasikan pohon keputusan menggunakan graphviz."""
    dot = Digraph()

    def add_edges(tree, parent_name):
        if not isinstance(tree, dict):
            dot.node(str(id(tree)), label=str(tree))
            dot.edge(parent_name, str(id(tree)))
            return

        feature_name = list(tree.keys())[0]
        for value, subtree in tree[feature_name].items():
            node_name = f'{feature_name} = {value}'
            dot.node(node_name, label=node_name)
            dot.edge(parent_name, node_name)
            add_edges(subtree, node_name)

    root_name = 'root'
    dot.node(root_name, label=root_name)
    add_edges(tree, root_name)

    dot.render(filename, format='png', cleanup=True)
    print(f"Pohon keputusan telah disimpan sebagai '{filename}.png'")

visualize_tree(c5_model.tree, filename='decision_tree_C50')

```

## Algoritma C4.5

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, confusion_matrix
from tree_c45 import C45

# Inisialisasi dan latih model C45
c45_model = C45(depth=3, min_samples_split=10, max_leaf_nodes=5)
c45_model.fit(X_train, y_train)

# Prediksi menggunakan model yang telah dilatih
y_pred_c45 = c45_model.predict(X_test)

# Evaluasi hasil
accuracy_c45 = accuracy_score(y_test, y_pred_c45)
precision_c45 = precision_score(y_test, y_pred_c45)
recall_c45 = recall_score(y_test, y_pred_c45, zero_division=1)
conf_matrix_c45 = confusion_matrix(y_test, y_pred_c45)

# Menampilkan hasil
print(f"Akurasi: {accuracy_c45}")
print(f"Presisi: {precision_c45}")
print(f"Recall: {recall_c45}")
print(f"Confusion Matrix:\n{conf_matrix_c45}")

visualize_tree(c45_model.tree, filename='decision_tree_C45')
```



# UNIVERSITAS MEDAN AREA

## FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 (061) 7366878, 7360168, 7364348, 7366781, Fax (061) 7366998 Medan 20223  
Kampus II : Jalan Seliabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, (061) 8225602, Fax. (061) 8226331 Medan 20122  
Website: [www.teknik.uma.ac.id](http://www.teknik.uma.ac.id) E-mail: [univ\\_medanarea@uma.ac.id](mailto:univ_medanarea@uma.ac.id)

Nomor : 112/FT.6/01.10/II/2024  
Lamp : -  
Hal : Perpanjang SK Pembimbing Tugas Akhir

12 Februari 2024

Yth. Pembimbing Tugas Akhir  
**Andre Hasudungan Lubis, S. Ti, MSc**  
di  
Tempat

Dengan hormat,  
Sehubungan telah berakhirnya waktu masa berlaku SK pembimbing nomor 678/FT.6/01.10/IX/2023 tertanggal 20 September 2023 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut :

N a m a : Sri Adelina Situmorang  
N P M : 198160056  
Jurusan : Teknik Informatika

Oleh karena itu kami mengharapkan kesediaan saudara :

**Andre Hasudungan Lubis, S. Ti, MSc** (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

**"Seleksi Calon Siswa Baru dengan Metode *Decision Tree* menggunakan Algoritma C.5 (Studi kasus: SMKN1 Doloksangul)"**

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,  
  
Dr. Pang Supriatno, ST, MT



# UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎(061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122  
Website: www.teknik.uma.ac.id E-mail: univ\_medanarea@uma.ac.id

Nomor : 671 /FT.6/01.10/IX/2023

Lamp : -

2 September 2023

H a l : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Kepala Sekolah SMK N 1 Doloksanggul  
Jln. Bonan Dolok Km. 2,5 Purba Dolok, Doloksanggul  
Di  
Sumatera Utara

Dengan hormat,  
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

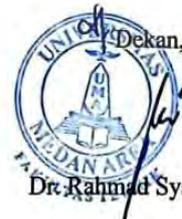
NO	N A M A	N P M	PRODI
1	Sri Adelina Situmorang	198160056	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

**Analisis Seleksi Calon Siswa Baru dengan Metode *Decision Tree* menggunakan Algoritma C5.0 (Studi kasus: SMKN1 Doloksanggul)**

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan,  
  
Dr. Rahmad Syah, S. Kom, M. Kom

Tembusan :  
1. Ka. BAMAI  
2. Mahasiswa  
3. File

**PEMERINTAH PROVINSI SUMATERA UTARA**  
**DINAS PENDIDIKAN**  
**SMK NEGERI 1 DOLOKSANGGUL**  
Jalan Bonandolok Km. 2,5 - Doloksanggul – Kab. Humbang Hasundutan  
Telp / Fax. 0633-31525 NPSN : 10208704  
Email : smkn1dolsa\_14@yahoo.co.id Web :  
[www.smkn1doloksanggul.sch.id](http://www.smkn1doloksanggul.sch.id)

---

Komp. Keahlian : Akuntansi dan Keuangan Lembaga, Otomatisasi dan Tata Kelola Perkantoran, Bisnis Daring dan Pemasaran, Tata Busana, Kecantikan Kulit dan Rambut, Usaha Perjalanan Wisata, Teknik Komputer dan Jaringan

---

**SURAT KETERANGAN SELESAI PENELITIAN**  
Nomor : 421.5/074/SMKN.1/DS/III/2024

Yang bertanda tangan dibawah ini :

Nama : Togar Halomoan Nainggolan, S. S.  
NIP : 19741111 200502 1 002  
Pangkat, Golongan Ruang : Pembina , IV/a  
Jabatan : Kepala Sekolah  
Unit Kerja : SMK Negeri 1 Doloksanggul  
Kecamatan Doloksanggul  
Kabupaten Humbang Hasundutan  
Provinsi Sumatera Utara  
NPSN : 10208704

Dengan ini menerangkan bahwa :

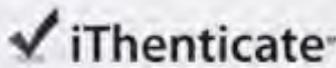
Nama : Sri Adelina Situmorang  
NIM : 198160056  
Jurusan : Teknik Informatika  
Judul Skripsi : *“Analisis Seleksi Calon Siswa Baru dengan Metode Decision Tree menggunakan Algoritma C5.0 (Studi Kasus : SMK Negeri 1 Doloksanggul)”*

Mahasiswa tersebut diatas benar telah melakukan penelitian di SMK Negeri 1 Doloksanggul pada tanggal 12 September s.d 12 Oktober 2023.

Demikian Surat Keterangan ini diperbuat sebenarnya untuk dapat dipergunakan seperlunya.

  
Doloksanggul, 07 Maret 2024  
Kepala SMK Negeri 1 Doloksanggul  
  
Togar Halomoan Nainggolan, S. S.  
Pembina  
NIP. 19741111 200502 1 002

 Ujireflecting Gambar



**iThenticate™** Similarity Report ID: oid:29477-67400250

---

PAPER NAME	AUTHOR
SELEKSI CALON SISWA BARU DENGAN METODE DECISION TREE MENGGUNAKAN ALGORITMA C5.0(STUDI KASUS _ SMKN_	SRI ADELINA SITUMORANG

---

WORD COUNT	CHARACTER COUNT
<b>7368 Words</b>	<b>41897 Characters</b>
PAGE COUNT	FILE SIZE
<b>48 Pages</b>	<b>615.8KB</b>
SUBMISSION DATE	REPORT DATE
<b>Sep 23, 2024 3:10 PM GMT+7</b>	<b>Sep 23, 2024 3:11 PM GMT+7</b>

---

**22% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 20% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

**Excluded from Similarity Report**

- Bibliographic material
- Cited material
- Abstract
- Small Matches (Less than 15 words)

