

**PENERAPAN ALGORITMA C4.5 PADA SMK  
MUHAMMADIYAH 16 DESA PON UNTUK MENENTUKAN  
KELOMPOK BELAJAR OLIMPIADE**

**SKRIPSI**

**DISUSUN OLEH**

**DIAJENG VESHA RAMADHINI**

**NPM. 2209010046**



**UMSU**

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2026**

**PENERAPAN ALGORITMA C4.5 PADA SMK  
MUHAMMADIYAH 16 DESA PON UNTUK MENENTUKAN  
KELOMPOK BELAJAR OLIMPIADE**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer  
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer  
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**DIAJENG VESHA RAMADHINI**

**NPM. 2209010046**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2026**

## LEMBAR PENGESAHAN

Judul Skripsi : PENERAPAN ALGORITMA C4.5 PADA SMK  
MUHAMMADIYAH 16 DESA PON UNTUK  
MENENTUKAN KELOMPOK BELAJAR OLIMPIADE  
Nama Mahasiswa : DIAJENG VESHA RAMADHINI  
NPM : 2209010046  
Program Studi : SISTEM INFORMASI

Menyetujui  
Komisi Pembimbing



(Amrullah, S.Kom., M.Kom.)  
NIDN. 0125118604

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung,  
S.Kom, M.Kom)  
NIDN. 0117088902

Dekan



(Dr. Al-Kholiqul Rizki, S.Kom., M.Kom.)  
NIDN. 0127099201

## PERNYATAAN ORISINALITAS

### PENERAPAN ALGORITMA C4.5 PADA SMK MUHAMMADIYAH 16 DESA PON UNTUK MENENTUKAN KELOMPOK BELAJAR OLIMPIADE

#### SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, April 2026

Yang membuat pernyataan



Diajeng Vesha Ramadhini

NPM. 2209010046

**PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN  
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Diajeng Vesha Ramadhini  
NPM : 2209010046  
Program Studi : Sistem Informasi  
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**PENERAPAN ALGORITMA C4.5 PADA SMK MUHAMMADIYAH 16  
DESA PON UNTUK MENENTUKAN KELOMPOK BELAJAR  
OLIMPIADE**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, April 2026

Yang membuat pernyataan

Diajeng Vesha Ramadhini

NPM. 2209010046

## RIWAYAT HIDUP

### DATA PRIBADI

Nama Lengkap : Diajeng Vesha Ramadhini  
Tempat dan Tanggal Lahir : Rampah Estate, 08 November 2004  
Alamat Rumah : Desa Pon Dusun I, Sei Bamban, Kab.  
Serdang Bedagai, Sumatera Utara, 20995  
Telepon/Faks/HP : 0895339500057  
E-mail : diajengvesharamadhini@gmail.com  
Instansi Tempat Kerja : -  
Alamat Kantor : -

### DATA PENDIDIKAN

SD : SD Muhammadiyah Desa Pon TAMAT: 2016  
SMP : SMP Negeri 1 Sei Bamban TAMAT: 2019  
SMA : SMA Negeri 1 Sei Bamban TAMAT: 2022

## KATA PENGANTAR



Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Akrim, M.Pd. Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom. Wakil dekan I Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Bapak Mhd. Basri, S.Si, S.Kom., M.Kom. Wakil dekan III Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
5. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom. Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
6. Bapak Mulkan Azhari, S.Kom, M.Kom. Sekretaris Program Studi Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
7. Bapak Amrullah, S.Kom., M.Kom sebagai dosen pembimbing, penulis mengucapkan terima kasih yang sebesar-besarnya atas kesabaran, bimbingan, serta dukungan yang tiada henti selama proses penyusunan skripsi ini. Berkat arahan dan motivasi yang diberikan, penulis dapat menyelesaikan skripsi ini dengan baik.
8. Penulis mengucapkan terima kasih kepada staff serta guru SMK Muhammadiyah 16 Desa Pon selaku tempat penelitian penulis yang telah memberikan kesempatan, dukungan, serta fasilitas selama proses penyusunan skripsi ini.
9. Penulis menyampaikan rasa terima kasih yang tulus kepada kedua orang tua tercinta ibu dan ayah atas doa, dukungan, dan kasih sayang yang selalu mengiringi setiap langkah selama penyusunan skripsi ini. Segala

pengorbanan dan perhatian yang diberikan menjadi sumber kekuatan dan semangat bagi penulis untuk terus berusaha hingga skripsi ini dapat terselesaikan.

10. Penulis menyampaikan terima kasih kepada kakak dan adik tercinta atas segala dukungan, doa, dan semangat yang selalu diberikan selama proses penyusunan skripsi ini. Penulis juga sangat menghargai setiap perhatian dan motivasi yang diberikan, yang turut membantu penulis dalam melewati berbagai proses hingga sampai pada tahap ini.
11. Penulis mengucapkan terima kasih kepada teman-teman terdekat, khususnya Silvia Haryani, Suci Handayani dan Fadhila Arumtini, yang sudah selalu ada dan menemani setiap proses selama masa perkuliahan dan penyusunan skripsi ini. Banyak momen yang dilalui bersama, hari ke hari, bulan ke bulan serta tahun ke tahun, yang membuat perjalanan ini terasa lebih ringan dan penuh makna. Terima kasih untuk kebersamaan, dukungan, dan semangat yang tidak pernah berhenti diberikan semua itu menjadi kenangan berharga yang akan selalu diingat oleh penulis.
12. Penulis mengucapkan terima kasih kepada seseorang yang belum dapat penulis sebutkan secara gamblang, namun memiliki peran penting dalam proses penyusunan skripsi ini. Kehadiran, dukungan, dan perhatian yang selalu diberikan telah menjadi sumber semangat bagi penulis untuk terus melangkah hingga mencapai tahap ini. Terima kasih untuk segala hal yang mungkin tidak terlihat, tetapi sangat berarti dan selalu dirasakan oleh penulis dalam setiap proses yang dilalui.
13. Penulis juga mengucapkan terima kasih kepada teman-teman atas kebersamaan dan dukungan selama proses penyusunan skripsi ini. Terima kasih sudah saling menemani dan membantu hingga penulis dapat menyelesaikan skripsi ini.
14. Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan bantuan, dukungan, dan kontribusi dalam proses penyusunan skripsi ini, termasuk pihak-pihak yang tidak dapat penulis sebutkan satu per satu. Segala bentuk bantuan yang diberikan sangat berarti hingga skripsi ini dapat terselesaikan.

15. Terakhir penulis mengucapkan terima kasih kepada diri penulis sendiri yang telah bertahan sejauh ini, yang tetap melangkah meskipun sering merasa lelah, ragu, bahkan kehilangan arah di tengah proses penyusunan skripsi ini. Tidak semua hari berjalan mudah, ada banyak fase yang penuh tekanan, kebingungan, dan rasa ingin menyerah, namun penulis tetap memilih untuk tidak berhenti. Terima kasih karena terus mengusahakan yang terbaik, bahkan ketika hasilnya belum sesuai harapan. Terima kasih karena tetap kuat di saat tidak ada yang benar-benar mengerti apa yang sedang dirasakan. Setiap air mata, setiap usaha yang mungkin tidak terlihat, dan setiap langkah kecil yang diambil, semuanya memiliki arti dalam perjalanan ini. Penulis juga belajar untuk menerima proses, memahami bahwa tidak semua hal harus sempurna, dan bahwa bertahan adalah bagian dari keberhasilan itu sendiri. Dari semua yang telah dilalui, penulis menyadari bahwa kekuatan terbesar datang dari dalam diri, dari kemauan untuk terus mencoba, memperbaiki, dan tidak menyerah. Akhirnya, penulis bersyukur karena telah sampai di titik ini. Semua perjalanan, luka, dan pelajaran yang didapatkan menjadi bagian yang tidak terpisahkan dari proses ini. Terima kasih untuk diri sendiri, karena sudah berani memulai, berjuang, dan menyelesaikan semuanya hingga akhir.

**PENERAPAN ALGORITMA C4.5 PADA SMK MUHAMMADIYAH 16  
DESA PON UNTUK MENENTUKAN KELOMPOK BELAJAR  
OLIMPIADE**

**ABSTRAK**

Penentuan kelompok belajar olimpiade di SMK Muhammadiyah 16 Desa Pon masih dilakukan secara manual berdasarkan penilaian guru, observasi, dan nilai rapor, sehingga berpotensi menimbulkan subjektivitas dan kurang optimal dalam mengelompokkan siswa sesuai kemampuan dan minatnya. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam mengklasifikasikan siswa ke dalam kelompok belajar olimpiade secara lebih objektif dan berbasis data.

Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan teknik data mining menggunakan algoritma C4.5. Data yang digunakan meliputi nilai mata pelajaran (MU dan TKDJ), nilai rata-rata rapor, pemahaman, dan perilaku siswa. Tahapan penelitian meliputi pengumpulan data, transformasi data, perhitungan entropy dan gain, serta pembentukan pohon keputusan. Hasil penelitian menunjukkan bahwa atribut dengan nilai gain tertinggi adalah gain terbesar yaitu kategori pemahaman sebesar 0,527, lalu disusul nilai gain kategori rata-rata rapor sebesar 0,424, gain kategori matematika sebesar 0,320, gain kategori sains komputer sebesar 0,247, dan gain kategori perilaku sebesar 0,209. Model yang dihasilkan mampu mengklasifikasikan siswa ke dalam kategori kelompok olimpiade matematika, sains komputer, dan tidak mengikuti olimpiade secara lebih sistematis.

Kata Kunci: Data Mining; Algoritma C4.5; Klasifikasi; Pohon Keputusan; Kelompok Belajar Olimpiade.

**APPLICATION OF THE C4.5 ALGORITHM AT SMK  
MUHAMMADIYAH 16 DESA PON TO DETERMINE OLYMPIAD  
STUDY GROUPS**

**ABSTRACT**

*The determination of Olympiad study groups at SMK Muhammadiyah 16 Desa Pon is still conducted manually based on teacher assessments, observations, and report card grades, which may lead to subjectivity and less optimal grouping of students according to their abilities and interests. This study aims to apply the C4.5 algorithm to classify students into Olympiad study groups in a more objective and data-driven manner.*

*The research method used is a quantitative approach with data mining techniques using the C4.5 algorithm. The data includes subject scores (MU and TKDJ), average report grades, student understanding, and behavior. The research stages consist of data collection, data transformation, entropy and gain calculation, and decision tree construction. The results show that the attribute with the highest gain value is the understanding category with a gain of 0.527, followed by the average report score category with a gain of 0.424, the mathematics category with a gain of 0.320, the computer science category with a gain of 0.247, and the behavior category with a gain of 0.209. The resulting model is able to classify students into Mathematics Olympiad, Computer Science Olympiad, and non-participation groups in a more systematic manner.*

*Keywords: Data Mining; C4.5 Algorithm; Classification; Decision Tree; Olympiad Study Groups*

## DAFTAR ISI

<b>LEMBAR PENGESAHAN</b> .....	<b>i</b>
<b>PERNYATAAN ORISINALITAS</b> .....	<b>ii</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI</b> .....	<b>iii</b>
<b>KARYA ILMIAH UNTUK KEPENTINGAN</b> .....	<b>iii</b>
<b>RIWAYAT HIDUP</b> .....	<b>iv</b>
<b>KATA PENGANTAR</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR TABEL</b> .....	<b>xii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>BAB I</b> .....	<b>1</b>
<b>PENDAHULUAN</b> .....	<b>1</b>
1.1    Latar Belakang Masalah.....	1
1.2    Rumusan Masalah .....	3
1.3    Batasan Masalah.....	4
<b>BAB II</b> .....	<b>6</b>
<b>LANDASAN TEORI</b> .....	<b>6</b>
2.1    Pengertian Sistem.....	6
2.2    Sistem Informasi .....	6
2.2.1    Konsep Dasar Sistem Informasi .....	6
2.3    Data Mining .....	8
2.3.1    Tahapan <i>Data Mining</i> .....	8
2.4    Klasifikasi .....	9
2.5    Decision <i>Tree Method</i> .....	10
2.6    Algoritma C4.5.....	10
2.7    Pendidikan dan Kelompok Belajar.....	12
2.8    Data .....	13
2.9    Python .....	13
2.10   Rapid Miner .....	14
2.11 <i>Unified Modelling Language (UML)</i> .....	14
2.9.1 <i>Activity Diagram</i> .....	15
2.9.2 <i>Use Case Diagram</i> .....	17

2.9.3	<i>Flowchart</i> .....	18
2.10	Literatur <i>Review</i> .....	20
<b>BAB III</b>	.....	<b>26</b>
<b>ANALISA DAN PERANCANGAN SISTEM</b>	.....	<b>26</b>
3.1	Analisis Permasalahan .....	26
3.2	Algoritma Sistem .....	27
3.2.1	Pengumpulan Data.....	27
3.2.2	Pengolahan Data .....	28
3.2.3	Transformasi Data.....	31
3.2.4	Penerapan Algoritma C4.5.....	33
3.3	Pemodelan dan Perancangan Sistem.....	38
3.3.1	Pemodelan sistem .....	39
3.4	Perancangan Antarmuka .....	43
<b>BAB IV</b>	.....	<b>49</b>
<b>IMPLEMENTASI DAN PENGUJIAN SISTEM</b>	.....	<b>49</b>
4.1	Data Perhitungan Algoritma C4.5 .....	49
4.1.1	Perhitungan Algoritma C4.5 .....	51
4.1.2	Pohon Keputusan .....	56
4.3	Implementasi Algoritma C4.5 Berbasis Web dalam Penentuan Siswa Olimpiade .....	59
4.3.1	Halaman <i>Login</i> .....	59
4.3.2	Tampilan <i>Dashboard</i> .....	60
4.3.3	Halaman Input Data Siswa .....	61
4.3.4	Halaman Data Tersimpan .....	63
4.3.5	Halaman Proses Algoritma .....	65
4.3.6	Halaman Hasil Kelompok.....	66
4.2	Hasil Pengujian Sistem .....	68
4.2.1	Hasil Pengujian Blackbox Testing.....	68
<b>BAB V</b>	.....	<b>72</b>
<b>PENUTUP</b>	.....	<b>72</b>
5.1	Kesimpulan .....	72
5.2	Saran.....	73
<b>DAFTAR PUSTAKA</b>	.....	<b>73</b>

## DAFTAR TABEL

Tabel 2.1 Komponen Activity Diagram.....	16
Tabel 2.2 Komponen Use Case Diagram.....	18
Tabel 2.3 Simbol Flowchart.....	19
Tabel 2.4 Literatur Review .....	21
Tabel 3.1 Data Penelitian .....	28
Tabel 3.2 Contoh data sample nilai rekap siswa .....	29
Tabel 3.3 Skema Diskretisasi.....	31
Tabel 4.4 Rekap Nilai yang telah ditransformasi.....	31
Tabel 4.5 Hasil Perhitungan Gain dan Entropy .....	37
Tabel 4. 1 Data Siswa .....	49
Tabel 4. 2 Tabel Blackbox Testing .....	68

## DAFTAR GAMBAR

Gambar 2.1 Tahapan data mining .....	9
Gambar 3.1 Flowchart Sistem Klasifikasi Kelompok Belajar .....	39
Gambar 3.2 Use Case Sistem Klasifikasi Kelompok Belajar .....	40
Gambar 3.3 Activity Diagram Sistem Klasifikasi Kelompok Belajar .....	42
Gambar 3.4 Halaman Login .....	43
Gambar 3.5 Halaman Dashboard .....	44
Gambar 3.6 Halaman Input Data Siswa .....	45
Gambar 3.7 Halaman Data Tersimpan .....	45
Gambar 3.8 Halaman Proses Algoritma C4.5 .....	46
Gambar 3.9 Halaman Hasil Kelompok .....	47
Gambar 4. 1 Pohon Keputusan Sistem .....	57
Gambar 4. 2 Tampilan Halaman Login .....	59
Gambar 4. 3 Tampilan Halaman Dashboard .....	60
Gambar 4. 4 Halaman Input .....	61
Gambar 4. 5 Halaman Input Manual .....	61
Gambar 4. 6 Halaman Data Tersimpan .....	63
Gambar 4. 7 Code Konversi Nilai .....	64
Gambar 4. 8 Halaman Proses Algoritma .....	65
Gambar 4. 9 Halaman Hasil Kelompok .....	66
Gambar 4. 10 Code Rule Pengelompokkan .....	67

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Perkembangan dalam teknologi informasi saat ini telah menjangkau semua bidang kehidupan, dan kemajuan ini juga terlihat di bidang pendidikan. Pendidikan adalah proses yang disengaja yang dilakukan dengan tujuan untuk mengembangkan lingkungan dan proses pembelajaran yang memungkinkan siswa untuk sepenuhnya menyadari potensi mereka sendiri (Dwi Nur Hidayat & Ahmad Bagus., 2022). Kemajuan teknologi juga membawa digitalisasi. Digitalisasi dapat diartikan sebagai sebuah proses menyimpan seluruh sifat dan informasi dari teks, suara, gambar, atau multimedia dalam sebuah string elektronik dari nol dan satu (Amrullah & Syahril, 2018). Pada tingkat Sekolah Menengah Kejuruan (SMK), potensi tersebut menjadi semakin penting karena setiap siswa memiliki minat dan kemampuan akademik yang berbeda-beda. Dengan adanya pengembangan potensi tersebut maka akan banyak siswa-siswa berprestasi. Prestasi pada siswa merupakan suatu kebanggaan dari suatu sekolah (Afandi et al., 2023). SMK Muhammadiyah 16 Desa Pon merupakan salah satu institusi pendidikan yang secara rutin mengikuti berbagai kegiatan kompetisi akademik atau olimpiade demi menunjang potensi para siswa, baik di tingkat kabupaten maupun provinsi. Namun, proses penentuan kelompok belajar olimpiade selama ini masih dilakukan secara manual berdasarkan penilaian guru, observasi umum, serta nilai rapor. Metode manual tersebut sering

menimbulkan masalah seperti ketidaktepatan dalam penempatan siswa, subjektivitas, serta pengolahan data siswa dalam jumlah yang cukup besar.

Dalam konteks penentuan kelompok belajar olimpiade dengan potensi serta minat siswa yang berbeda-beda, penentuan klasifikasi siswa yang akurat menjadi sangat penting. Kategori kelayakan siswa untuk kelompok belajar olimpiade seperti kelompok olimpiade matematika, kelompok olimpiade sains computer dan tidak ikut kelompok olimpiade. Dengan adanya data ringkasan yang lengkap dan terorganisir mengenai siswa, yang mencakup informasi seperti nama, nilai dalam mata pelajaran Olimpiade rata-rata nilai setiap semester, pengalaman di Olimpiade, serta keterlibatan di kelas, teknik penambangan data dapat diterapkan untuk mengolah data ini dan menjadikannya informasi yang bermanfaat dalam mendukung keputusan yang diambil.

Data mining juga merupakan proses analisis untuk menemukan pola dan memprediksi sekumpulan data untuk memperoleh pengetahuan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (C, 2024). Hasil data mining nantinya akan akurat tergantung dengan metode algoritma yang dipakai. Penggunaan algoritma juga harus disesuaikan dengan jenis data yang perlu diolah (Rongcai et al., n.d.). Dalam konteks ini salah satu algoritma yang dianggap akan akurat nantinya untuk klasifikasi dalam pengambilan keputusan penentuan kelompok olimpiade adalah Algoritma C4.5. Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara

algoritma-algoritma yang lain. (Putra, 2024). Karena berbagai keunggulan yang dimilikinya, algoritma C4.5 diharapkan mampu mengurangi kesalahan yang dilakukan manusia dan secara tepat mengklasifikasi potensi kemampuan siswa dalam olimpiade secara akurat dan tepat yang nantinya dapat meningkatkan kualitas pendidikan. Dengan adanya latar belakang diatas maka, peneliti tertarik untuk melakukan penelitian dengan judul **“Penerapan Algoritma C4.5 Pada Smk Muhammadiyah 16 Desa Pon Untuk Menentukan Kelompok Belajar Olimpiade”**.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang diuraikan sebelumnya, maka dirumuskan bahwa rumusan masalah dalam penelitian ini adalah :

1. Bagaimana penerapan algoritma C4.5 dalam membentuk model klasifikasi untuk menentukan kategori kelompok belajar olimpiade seperti kelompok olimpiade Matematika, kelompok Olimpiade Sains Komputer atau tidak ikut kelompok olimpiade?
2. Bagaimana implementasi sistem yang mampu mengidentifikasi potensi siswa untuk mengikuti olimpiade berdasarkan atribut-atribut penting dalam data siswa selama satu semester menggunakan pendekatan pohon keputusan C4.5?
3. Bagaimana hasil klasifikasi siswa yang diperoleh dari algoritma C4.5 dalam pengelompokan siswa yang mengikuti dan tidak mengikuti kelompok belajar olimpiade?

### 1.3 Batasan Masalah

Berdasarkan latar belakang dan permasalahan yang terkait yang telah diuraikan diatas, adapun Batasan masalah dalam penelitian ini ialah

1. Penelitian ini hanya berfokus pada penerapan algoritma C4.5 sebagai metode klasifikasi untuk menentukan kelompok belajar olimpiade.
2. Data yang digunakan terbatas pada data siswa SMK Muhammadiyah 16 Desa Pon, meliputi nilai mata pelajaran terkait olimpiade, minat siswa, nilai rata-rata rapor, pengalaman olimpiade, serta keaktifan di kelas.
3. Klasifikasi kelompok belajar olimpiade dibatasi pada tiga kategori, yaitu:
  - a. Kelompok Olimpiade Matematika
  - b. Kelompok Olimpiade Sains Komputer
  - c. Tidak Mengikuti Kelompok Olimpiade
4. Pengolahan data hanya dilakukan pada tahun 2024-2025 selama satu semester.
5. Penelitian ini hanya diperuntukan untuk SMK Muhammadiyah 16 Desa Pon dengan jurusan Teknik Komputer Jaringan.

### 1.4 Tujuan Penelitian

Adapun tujuan penelitian ini ialah :

1. Menganalisis data siswa yang meliputi nilai mata pelajaran terkait olimpiade, nilai rata-rata rapor, serta keaktifan di kelas sebagai dasar dalam proses klasifikasi penentuan kelompok belajar olimpiade di SMK Muhammadiyah 16 Desa Pon.
2. Menerapkan algoritma C4.5 untuk membangun model klasifikasi yang mampu menentukan kategori kelompok belajar olimpiade, yaitu kelompok olimpiade

Matematika, olimpiade Sains Komputer atau tidak mengikuti kelompok olimpiade.

### **1.5 Manfaat Penelitian**

Berdasarkan latar belakang, rumusan masalah, batasan masalah dan tujuan penelitian yang telah disebutkan sebelumnya, adapun manfaat dari penelitian ini sebagai berikut :

1. Bagi Penulis, penelitian ini memberikan pengalaman serta penerapan langsung dalam penggunaan algoritma C4.5 untuk menyelesaikan permasalahan nyata di lingkungan pendidikan. Selain itu, penelitian ini juga menambah wawasan penulis mengenai implementasi teknik data mining sebagai pendukung dalam proses pengambilan keputusan.
2. Bagi Sekolah (SMK Muhammadiyah 16 Desa Pon), hasil penelitian ini diharapkan dapat membantu pihak sekolah dalam menentukan kelompok belajar olimpiade secara lebih objektif, akurat, dan berbasis data. Dengan adanya sistem yang diusulkan, tingkat subjektivitas guru dalam proses seleksi siswa dapat diminimalkan, terutama ketika jumlah data siswa cukup besar, serta memberikan rekomendasi yang dapat dimanfaatkan untuk meningkatkan kualitas pembinaan olimpiade.
3. Bagi Universitas Muhammadiyah Sumatera Utara, penelitian ini diharapkan dapat menjadi sumber referensi ilmiah yang menambah khazanah penelitian di bidang Sistem Informasi, khususnya yang berkaitan dengan penerapan algoritma data mining dalam dunia pendidikan.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Pengertian Sistem**

Menurut Fat pengertian sistem adalah suatu himpunan suatu “benda” nyata atau abstrak (*a set of thing*) yang terdiri dari bagian atau komponen yang saling berkaitan, berhubungan, berketergantungan, saling mendukung, yang secara keseluruhan Bersatu dalam kesatuan (*Unity*) untuk mencapai tujuan tertentu secara efisien dan efektif.

#### **2.2 Sistem Informasi**

Menurut (Rojabi, M. A. (2025) pengertian sistem Informasi sebagai kumpulan komponen yang saling terhubung untuk mengumpulkan, memproses, menyimpan, dan menyebarkan data dan informasi. Tujuan utamanya adalah untuk mendukung operasi, manajemen, dan pengambilan keputusan di dalam suatu organisasi.

##### **2.2.1 Konsep Dasar Sistem Informasi**

Sistem informasi merupakan pendekatan berpikir yang melihat suatu organisasi, proses, atau permasalahan sebagai sebuah sistem yang terdiri dari terdiri dari bagian-bagian yang saling berhubungan dan bergantung satu sama lain. Dalam hal ini, suatu sistem tidak dapat diapresiasi hanya dengan menganalisis satu unsur secara terpisah; sebaliknya, sistem tersebut harus dilihat sebagai suatu kesatuan yang terintegrasi atau menyeluruh. Hal ini sejalan dengan pendapat Stair dan Reynolds (yang menyatakan bahwa sistem informasi adalah kombinasi terstruktur dari manusia, perangkat keras, perangkat lunak, data, dan prosedur yang saling

terhubung untuk menghasilkan informasi yang bernilai bagi pengambilan keputusan. Oleh karena itu, sistem informasi menekankan pentingnya memahami hubungan antarkomponen tersebut serta dampak perubahan suatu komponen terhadap keseluruhan sistem.

Selain itu, sistem informasi berorientasi pada proses, sehingga setiap alur kerja mulai dari input, proses pengolahan, hingga output harus dianalisis secara menyeluruh untuk memastikan efektivitas dan efisiensi. O'Brien dan Marakas menegaskan bahwa informasi berkualitas hanya dapat dihasilkan melalui proses yang terdefinisi dengan baik dan dijalankan secara konsisten. Dengan demikian, pendekatan ini membantu dalam mengidentifikasi kelemahan proses, mengoptimalkan alur data, serta meningkatkan ketepatan informasi yang dihasilkan. (Saputri et al., 2025).

Pendekatan ini juga menuntut pemikiran berbasis tujuan, di mana sistem informasi harus dirancang untuk memenuhi kebutuhan spesifik organisasi. Setiap perancangan sistem, analisis kebutuhan, hingga implementasi teknologi harus selaras dengan tujuan organisasi dalam menghasilkan informasi yang relevan, akurat, dan tepat waktu (Laudon & Laudon, n.d.) Selain itu, cara berpikir sistem informasi menekankan pentingnya pengambilan keputusan berbasis data. Dalam era digital, data menjadi aset utama yang dapat diolah menjadi pengetahuan melalui analisis, termasuk melalui metode data mining dan algoritma klasifikasi seperti C4.5.

Terakhir, sistem informasi bersifat adaptif, artinya harus mampu mengikuti perkembangan teknologi yang cepat. Sistem perlu dievaluasi secara berkala untuk memastikan keamanan, kinerja, dan kecocokan dengan kebutuhan masa kini analisis

analisis sistem atau peneliti dapat melihat permasalahan secara komprehensif, menemukan akar masalah, serta menentukan solusi yang paling efektif dengan dukungan teknologi informasi yang tepat.

### **2.3 Data Mining**

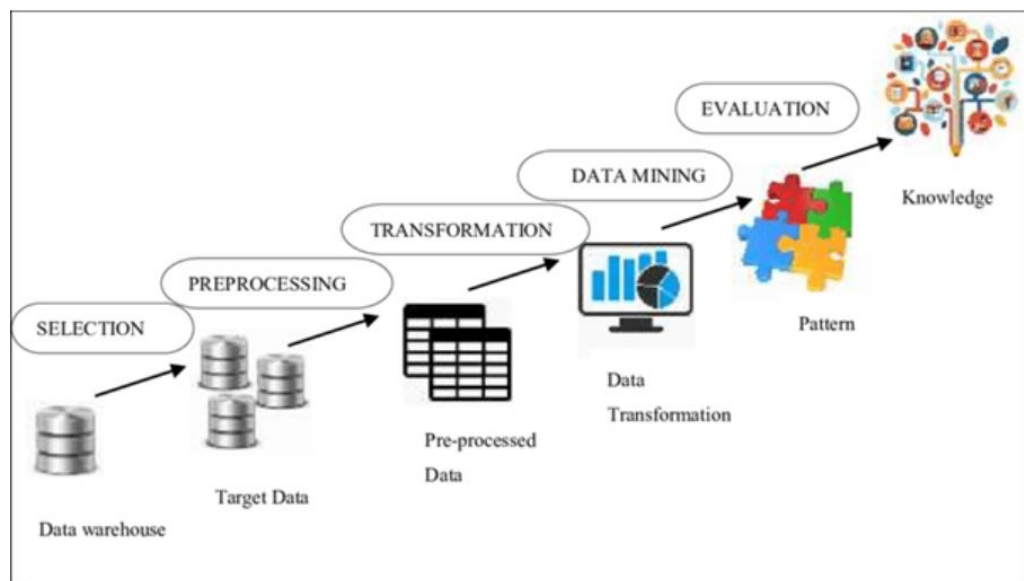
Menurut (Syahdan & Sindar, 2018) data mining merupakan proses iteratif dan interaktif untuk menemukan pola-pola atau model baru yang sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang.

#### **2.3.1 Tahapan *Data Mining***

*Data mining* merupakan kegiatan yang bertujuan untuk mengenali bermacam jenis informasi. Dari banyaknya data yang disimpan dalam basis data, elemen-elemen ini ditemukan di mana pola atau karakteristik lain dicari yang mungkin menghasilkan informasi yang bermanfaat bagi organisasi atau perusahaan yang memiliki basis data itu. Adapun tahapan atau penjelasan mengenai proses data mining adalah sebagai berikut:

- a. *Data Cleansing*, Proses dimana data diolah lalu dipilih data yang dianggap bisa dipakai.
- b. *Data Integration*, Proses menggabungkan data yang dianggap berulang akan digabungkan menjadi satu.
- c. *Selection*, Proses seleksi atau pemilihan data yang dianggap relevan terhadap analisis.

- d. *Data Transformation*, Proses transformasi data terpilih ke dalam bentuk mining procedure.
- e. *Data Mining*, Proses dimana dilakukan beragam teknik untuk mengekstrak pola-pola potensial menghasilkan data yang berguna.
- f. *Pattern Evolution*, Proses dimana pola-pola yang telah diidentifikasi berdasarkan measure yang diberikan.
- g. *Knowledge Presentation*, Proses paling akhir dari proses KDD, Data-data yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil Tindakan berdasarkan analisis. (Rongcai et al., n.d.)



**Gambar 2.1 Tahapan data mining**

## 2.4 Klasifikasi

Pada umumnya *classification* atau klasifikasi adalah sebuah proses pengelompokan. Ditarik pada keilmuan *data science*, pengelompokan dalam klasifikasi adalah sebuah proses membagi data menjadi beberapa kategori atau

kelas/label dengan tujuan agar lebih mudah untuk diolah dan dianalisis (Algoritma et al., 2018)

## 2.5 Decision Tree Method

*Decision tree* adalah *flowchart* seperti pohon. Dimana setiap *node* menunjukkan suatu test pada suatu atribut, tiap branch merepresentasikan hasil dari test tersebut, dan *leaf node* menunjukkan kelas-kelas atau distribusi kelas. (Han et al, 2008). Decision tree berguna untuk mengeksplorasi data yang sudah melewati tahap preprocessing dan menemukan model yang tersembunyi dari data dengan sebuah target variabel sehingga dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan record yang lebih kecil dengan memperhatikan variabel tujuannya. (Zega, 2014)

## 2.6 Algoritma C4.5

Menurut Ba'abbad et al (2021). Algoritma C4.5 merupakan teknik klasifikasi dengan menggunakan entropy dan keuntungan informasi sebagai pemisah pada pohon Keputusan. Algoritma C4.5 sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk mem- prediksi nilai atribut bertipe diskret dari *record* yang baru.

Adapun cara algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai *node* akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.

4. Ulangi proses untuk setiap cabang sampai semua
5. Kasus pada cabang memiliki kelas yang sama. (Sukardi, Syukur A. and Supriyanto C. (2014),

Untuk memilih atribut sebagai *node* akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* digunakan rumus seperti tertera dalam persamaan berikut:

$$Gain (S,A) = Entropy (S) - \sum_{i=1}^n \frac{S_i}{S} \times Entropy (S_i)$$

Keterangan :

S= Himpunan kasus

A= fitur N = jumlah partisi atribut A

|S<sub>i</sub>|= proporsi S<sub>i</sub> terhadap S

|S|= jumlah kasus dalam S

Setelah mendapatkan nilai *Gain*, ada satu hal lagi yang perlu kita lakukan perhitungan, yaitu mencari nilai *Entropy*. *Entropy* digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan keluaran atribut. Rumus dasar dari *Entropy* tersebut adalah sebagai berikut:

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

Keterangan:

S: Himpunan kasus

n: Jumlah partisi S

p<sub>i</sub>: Proporsi dari S<sub>i</sub> terhadap S (Budiarto et al., 2022).

## 2.7 Pendidikan dan Kelompok Belajar

Menurut Marimba, A. D. *Pengantar Filsafat Pendidikan Islam*. Al-Ma'arif) Pendidikan adalah bimbingan atau pertolongan yang diberikan secara sadar oleh pendidik kepada peserta didik untuk mengembangkan kepribadian dan potensi mereka agar lebih baik. Sedangkan menurut Hamalik, O. *Proses Belajar Mengajar*. Bumi Aksara.) kelompok belajar merupakan suatu unit kecil dari peserta didik yang dibentuk untuk melakukan aktivitas belajar secara kolaboratif sehingga dapat meningkatkan pemahaman dan minat belajar. (Munandar et al., 2022)

Berdasarkan pengertian kelompok belajar yang menekankan pentingnya kerja sama, interaksi, serta pembagian peran dalam proses belajar, dan pengertian pendidikan sebagai usaha sadar untuk mengembangkan potensi peserta didik, maka penentuan kelompok belajar yang tepat menjadi bagian penting dalam mendukung keberhasilan proses pembelajaran. Di SMK Muhammadiyah 16 Desa Pon, pembentukan kelompok belajar olimpiade perlu dilakukan secara terstruktur agar siswa yang memiliki kemampuan, minat, dan potensi yang sesuai dapat dikelompokkan secara efektif.

Dalam konteks inilah, algoritma C4.5 digunakan sebagai alat bantu untuk mengolah data siswa secara sistematis sehingga proses pengelompokan tidak hanya bergantung pada pengamatan subjektif, tetapi didasarkan pada analisis yang lebih objektif dan terukur. Dengan demikian, penerapan algoritma C4.5 mendukung prinsip pendidikan dan pembentukan kelompok belajar yang optimal, yaitu dengan menempatkan siswa pada kelompok yang paling sesuai untuk mengembangkan kemampuan mereka dalam kegiatan olimpiade.

## 2.8 Data

Pengertian data menurut *Webster New World Dictionary*, Data adalah things known or assumed, yang berarti bahwa data itu sesuatu yang diketahui atau dianggap. Diketahui artinya yang sudah terjadi merupakan fakta (bukti). Data dapat memberikan gambaran tentang suatu keadaan atau persoalan. Data bisa juga didefinisikan sekumpulan informasi atau nilai yang diperoleh dari pengamatan (observasi) suatu obyek, data dapat berupa angka dan dapat pula merupakan lambang atau sifat. Beberapa macam data antara lain; data populasi dan data sampel, data observasi, data primer, dan data sekunder. (Syafrizal, n.d.)

Adapun data yang akan digunakan penelitian ini sebagai implementasi algoritma C4.5 dalam klasifikasi kelompok belajar di SMK Muhammadiyah 16 Desa pon adalah:

1. Data siswa
2. Data analisis minat siswa
3. Data nilai variabel siswa
4. Data hasil analisis klasifikasi.

## 2.9 Python

Menurut (Rosmala et al), *Python* adalah bahasa pemrograman tingkat tinggi yang bersifat interpreter, interaktif, object oriented dan dapat berjalan di hampir semua platform. *Python* merupakan bahasa pemrograman dengan tingkat tinggi yang mudah dipahami berkat cara penulisannya yang jelas dan sederhana, yang lebih mirip dengan bahasa manusia dibandingkan kode komputer, serta dilengkapi dengan modul-modul yang efisien dan siap digunakan. Kode sumber *Python* akan dikonversi menjadi format *bytecode* yang dapat dilaksanakan. Program *Python*

cenderung berjalan lebih lambat jika dibandingkan dengan bahasa pemrograman yang berada pada tingkat lebih rendah.

### **2.10 Rapid Miner**

Menurut (CTI dkk., 2017) RapidMiner, merupakan *software*/perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma dalam penambangan data, RapidMiner dapat menarik pola dari data yang sangat besar dengan menggabungkan metode statistik, kecerdasan buatan, dan sistem basis data. RapidMiner memungkinkan pengguna untuk menangani data dalam jumlah besar dengan cara menggunakan operator. Operator ini berfungsi untuk melakukan modifikasi pada data. Data dihubungkan dengan node operator, dan setelah itu, cukup dengan menghubungkannya ke *node* output untuk menampilkan hasilnya. Hasil yang diperlihatkan RapidMiner pun dapat ditampilkan secara visual dengan grafik. Menjadikan RapidMiner adalah salah satu *software* pilihan untuk melakukan ekstraksi data dengan metode-metode data *mining*. (Hendry, 2021)

### **2.11 Unified Modelling Language (UML)**

Menurut (Wiyono et al., 2012) *Unified Modeling Language* (UML) merupakan bahasa pemodelan visual yang digunakan dalam rekayasa perangkat lunak untuk menggambarkan sistem secara terstruktur, baik dari aspek struktur maupun perilaku sistem yang dibangun. UML menyajikan sekumpulan diagram terstandarisasi yang membantu perancang sistem dalam memvisualisasikan desain, mendokumentasikan kebutuhan fungsional dan struktural, serta mempermudah komunikasi antar pemangku kepentingan selama proses pengembangan perangkat lunak.



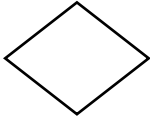



Dalam konteks pemodelan sistem, UML memfasilitasi pengembang untuk mengidentifikasi kebutuhan sistem secara grafis, serta mendokumentasikan dan merinci rancangan sistem sebelum dilakukan pembangunan kode program. Dengan bantuan UML, kompleksitas sebuah sistem dapat dibuat lebih sederhana dan mudah untuk dianalisis oleh tim pengembang maupun pihak terkait lainnya. (Wayahdi & Ruziq, 2023)

### **2.9.1 *Activity Diagram***

*Activity diagram* merupakan salah satu jenis diagram dalam *Unified Modeling Language (UML)* yang digunakan untuk menggambarkan urutan alur aktivitas atau proses kerja dalam sebuah sistem. Diagram ini memaparkan bagaimana sebuah aktivitas dimulai, berlangsung, dan selesai, termasuk pengambilan keputusan serta aktivitas yang dilakukan secara bersamaan. Menurut Pressman (2015), *Activity Diagram* digunakan untuk memodelkan alur kerja (workflow) dari suatu sistem yang menunjukkan aktivitas dan transisi antar aktivitas tersebut. Diagram ini sangat membantu dalam memahami logika proses bisnis dan alur sistem secara menyeluruh sebelum sistem diimplementasikan.

Penelitian yang dilakukan oleh Nugroho (2018) menyatakan bahwa *Activity Diagram* efektif digunakan dalam tahap analisis dan perancangan sistem karena mampu merepresentasikan proses sistem secara visual dan sistematis, sehingga memudahkan pengembang dalam mengidentifikasi kesalahan alur sejak tahap awal perancangan.

Tabel 2.1 Komponen Activity Diagram

Simbol	Nama	Keterangan
	Status awal	Sebuah diagram aktivitas dimulai dengan sebuah keadaan awal.
	Aktivitas	Aktivitas yang dilakukan oleh sistem umumnya dimulai dengan kata kerja.
	Percabangan/Decision	Percabangan di mana terdapat lebih dari satu opsi aktivitas
	Penggabungan/Join	Penggabungan yang melibatkan lebih dari satu aktivitas disatukan menjadi satu kesatuan.
	Status akhir	Status akhir yang ditentukan oleh sistem, sebuah diagram sistem menunjukkan status akhir.
	Swimlane	Swimlane membedakan bagian-bagian dalam organisasi yang memiliki


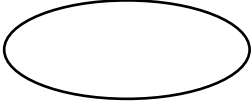

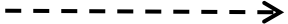


		tanggung jawab atas tindakan yang berlangsung.
--	--	--

### 2.9.2 *Use Case Diagram*

*Use Case Diagram* merupakan sebuah diagram UML yang berfungsi untuk menggambarkan interaksi antara pengguna (aktor) dan sistem. Diagram ini memperlihatkan berbagai fungsi yang disediakan oleh sistem, serta keterkaitan antara aktor dan fungsi-fungsi tersebut.

Berdasarkan penelitian oleh Pratama et al. (2019), *Use Case Diagram* sangat penting dalam tahap analisis kebutuhan karena mampu menggambarkan ruang lingkup sistem secara jelas serta membantu komunikasi antara pengembang dan pengguna.

Tabel 2.2 Komponen Use Case Diagram

Simbol	Keterangan
	<p>Aktor: Menjadi pengganti bagi individu, sistem lain, atau perangkat saat berinteraksi dengan <i>use case</i>.</p>
	<p><i>Use case</i>: Penggambaran dan hubungan antara sistem dan pihak yang terlibat.</p>
	<p><i>Associaton</i>: Abstraksi dari koneksi antara aktor dan kasus penggunaan.</p>
	<p>Generalisasi: Menunjukkan keahlian aktor agar dapat berinteraksi dengan skenario penggunaan.</p>
<p>&lt;&lt;<i>include</i>&gt;&gt;</p> 	<p>Menunjukkan bahwa sebuah <i>use case</i> sepenuhnya adalah bagian dari fungsi <i>use case</i> yang lain.</p>
<p>&lt;&lt;<i>extend</i>&gt;&gt;</p> 	<p>Mengindikasikan bahwa sebuah <i>use case</i> sepenuhnya merupakan fitur tambahan dari <i>use case</i> yang lain ketika suatu syarat tertentu terpenuhi.</p>






### 2.9.3 Flowchart

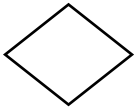



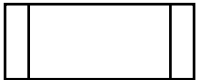

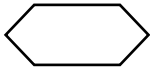
*Flowchart* atau bagan alir merupakan representasi grafis dari langkah-langkah penyelesaian suatu masalah atau proses yang disusun secara sistematis

menggunakan simbol-simbol standar. *Flowchart* digunakan untuk menggambarkan alur logika suatu proses atau algoritma.

Penelitian oleh Hidayat dan Prabowo (2020) menyatakan bahwa *Flowchart* membantu pengembang dalam memahami logika proses sistem serta mempermudah proses dokumentasi dan pemeliharaan sistem.

**Tabel 2.3 Simbol *Flowchart***

<b>Simbol</b>	<b>Nama Simbol</b>	<b>Keterangan</b>
	<i>Flow (Garis Alir)</i>	Menghubungkan satu simbol dengan simbol lainnya (connecting line).
	<i>On-Page Reference</i>	Penghubung antara langkah-langkah dalam dokumen yang sama.
	<i>Off-Page Reference</i>	Penghubung proses pada lembar kerja yang berbeda.
	<i>Terminator</i>	Mengindikasikan permulaan atau penutupan suatu program atau proses.
	<i>Process</i>	Proses yang dilakukan oleh komputer.

	<i>Decision</i>	Menampilkan situasi dengan dua opsi (iya / tidak).
	<i>Input / Output</i>	Proses input atau output.
	<i>Manual Operation</i>	Proses yang dilakukan secara manual (bukan komputer).
	<i>Document</i>	Input diambil dari bahan fisik atau hasil cetakan.
	<i>Predefined Process</i>	Subprogram atau prosedur yang telah didefinisikan.
	<i>Display</i>	Menyatakan output yang ditampilkan ke layar.
	<i>Preparation</i>	Proses persiapan atau pemberian nilai awal.

## 2.10 Literatur Review

Berikut ini tinjauan pustaka yang relevan dengan penelitian seperti pada table 2.1 dibawah ini

Tabel 2.4 Literatur *Review*

No.	Penulis & Tahun	Judul Penelitian	Metode & instrumen	Kelebihan & Keterbatasan
1.	Yofhanda Septi Eirlangga & Aldo Eko Syaputra (2022)	Klasifikasi Penjurusan pada Sekolah Menengah Atas (SMA) dengan Metode Algoritma C4.5	Penelitian ini menggunakan metode data mining dengan algoritma C4.5 untuk klasifikasi penjurusan siswa SMA. Data yang digunakan berasal dari 44 data rapor siswa kelas IX tahun 2022 di SMA Pertiwi 2 Padang, dengan 11 variabel mata pelajaran (IPA, IPS, Bahasa Indonesia, PKN, PAI, Matematika, Bahasa Inggris, Seni Budaya, TIK, dan Penjas). Instrumen penelitian berupa data nilai rapor siswa, yang diolah secara manual menggunakan Microsoft Excel untuk	Pohon keputusan dan rule yang jelas, sehingga mudah dipahami oleh pihak sekolah. Selain itu, hasil penelitian menunjukkan tingkat akurasi yang sangat tinggi, sehingga dapat membantu panitia PPDB dalam menentukan jurusan siswa secara objektif dan mengurangi kesalahan penjurusan akibat human error .  Jumlah data yang relatif sedikit, sehingga hasil klasifikasi masih terbatas pada satu sekolah. Selain itu,

			perhitungan entropy, gain, dan pembentukan decision tree .	variabel yang digunakan hanya berdasarkan nilai akademik, belum mempertimbangkan faktor non-akademik seperti minat dan bakat siswa secara psikologis.
2.	Dela Ratih Sanggar Wati, Gisa Restu Mahameru, Hapsah Roudhotul Jannah, Iin Fathonah Fatmalasari, Ulya Makhmudah (2024)	Efektivitas Konseling Kelompok: Peningkatan Motivasi Belajar Siswa Menengah Atas.	Metode penelitian menggunakan studi pustaka (literature review) dengan mengumpulkan data sekunder dari jurnal, buku, dan penelitian terdahulu. Analisis dilakukan secara deskriptif untuk mengevaluasi efektivitas konseling.	Penelitian ini memberikan wawasan komprehensif bagi pendidik mengenai faktor-faktor yang mempengaruhi efektivitas konseling kelompok, seperti dinamika kelompok dan peran konselor. Karena menggunakan metode studi pustaka, penelitian ini tidak melakukan uji coba langsung di lapangan untuk memvalidasi

				efektivitas intervensi pada sampel spesifik secara real-time.
3.	Tim Peneliti Technomedia Journal, 2023	Klasifikasi Tingkat Kepuasan Mahasiswa terhadap Pelayanan Akademik Menggunakan Algoritma C4.5	Penelitian ini menggunakan metode data mining klasifikasi dengan algoritma C4.5 untuk mengukur tingkat kepuasan mahasiswa. Data diperoleh dari kuesioner survei kepuasan mahasiswa Universitas Jenderal Achmad Yani (Unjani) Cimahi tahun 2017–2021. Instrumen penelitian berupa kuesioner SERVQUAL yang mencakup lima dimensi, yaitu Tangible, Empathy, Responsiveness, Reliability, dan Assurance. Proses	Kelebihan penelitian ini adalah penggunaan data dalam jumlah besar dari beberapa fakultas, sehingga hasil klasifikasi lebih representatif. Algoritma C4.5 mampu menunjukkan variabel pelayanan yang paling dominan memengaruhi kepuasan mahasiswa, dengan tingkat akurasi yang sangat tinggi (di atas 90%) . Keterbatasan penelitian ini adalah ketergantungan pada data persepsi subjektif mahasiswa melalui

			<p>pengujian dilakukan menggunakan RapidMiner Studio 9.10 untuk memperoleh nilai akurasi .</p>	<p>kuesioner. Selain itu, penelitian hanya menggunakan satu algoritma, sehingga belum membandingkan performa C4.5 dengan metode klasifikasi lain seperti Naïve Bayes atau K-NN</p>
4.	<p>Tri Lisnawati, Suroyo, Benny Agus Pribadi (2022)</p>	<p>Efektivitas Model Pembelajaran Kelompok dan Problem Based Learning pada Studi Sosial terhadap Hasil Belajar Siswa Berdasarkan Gaya Belajar Siswa Sekolah Dasar.</p>	<p>Penelitian kuantitatif dengan metode kuasi-eksperimental dan desain faktorial 2x2. Sampel berjumlah 32 siswa. Analisis data menggunakan ANOVA dua arah dan uji-t dengan tingkat signifikansi 0,05</p>	<p>Mampu membuktikan adanya interaksi antara model pembelajaran (STAD dan PBL) dengan gaya belajar siswa (introvert dan ekstrovert) terhadap hasil belajar.</p> <p>Keterbatasan: Jumlah sampel yang relatif kecil (32 siswa) dapat membatasi generalisasi hasil penelitian untuk</p>

				populasi yang lebih luas.
5.	Selpita Sari (2022).	Pemanfaatan Kelompok Belajar Siswa Dalam Mengatasi Kejenuhan Belajar Siswa di SMAN 8 Rejang Lebong.	Penelitian kualitatif dengan pendekatan deskriptif. Instrumen pengumpulan data meliputi wawancara (kepada guru BK, wali kelas, dan siswa), observasi, dan dokumentasi. Analisis data menggunakan reduksi data dan triangulasi sumber.	Memberikan gambaran mendalam mengenai penyebab kejenuhan belajar langsung dari perspektif berbagai pemangku kepentingan di sekolah.  Penulis mengakui adanya kelemahan dan keterbatasan pribadi dalam penyusunan skripsi ini serta masih banyaknya kekurangan yang perlu diperbaiki.

## **BAB III**

### **ANALISA DAN PERANCANGAN SISTEM**

#### **3.1 Analisis Permasalahan**

Identifikasi permasalahan dalam penelitian ini adalah proses penentuan kelompok belajar olimpiade di SMK Muhammadiyah 16 Desa Pon yang masih dilakukan secara manual. Penentuan tersebut didasarkan pada penilaian guru, observasi umum, serta nilai rapor siswa, sehingga berpotensi menimbulkan ketidaktepatan dalam pengelompokan siswa. Proses manual ini juga memungkinkan terjadinya subjektivitas dalam pengambilan keputusan serta kesulitan dalam mengolah data siswa yang jumlahnya cukup besar dan memiliki banyak kriteria penilaian.

Permasalahan berikutnya adalah belum optimalnya pemanfaatan data siswa yang sebenarnya telah tersedia secara lengkap dan terstruktur. Data tersebut meliputi nilai mata pelajaran olimpiade, nilai rata-rata rapor, minat siswa. Namun, data-data tersebut belum diolah secara sistematis untuk menghasilkan informasi yang dapat digunakan sebagai dasar penentuan kelompok belajar olimpiade secara objektif dan akurat.

Selain itu, perbedaan kemampuan akademik dan minat siswa menyebabkan perlunya klasifikasi yang tepat ke dalam kategori kelompok belajar olimpiade, seperti kelompok olimpiade matematika, kelompok olimpiade sains komputer, serta siswa yang tidak mengikuti kelompok olimpiade. Apabila proses klasifikasi tidak dilakukan dengan metode yang tepat, maka pembinaan olimpiade menjadi kurang efektif dan potensi siswa tidak dapat dikembangkan secara maksimal.

## 3.2 Algoritma Sistem

Penelitian ini menggunakan pendekatan kuantitatif dengan bantuan bahasa pemrograman Python dalam proses pengolahan dan analisis data. Pendekatan kuantitatif dipilih karena penelitian ini mengolah data siswa dalam bentuk angka, seperti nilai akademik dan kriteria pendukung lainnya, yang dianalisis secara matematis menggunakan algoritma C4.5 untuk menentukan kelompok belajar olimpiade.

Dalam pengembangan sistem, penelitian ini menerapkan metode *System Development Life Cycle* (SDLC) dasar untuk menciptakan perangkat lunak. Pendekatan SDLC diterapkan agar proses pengembangan sistem dilakukan melalui langkah-langkah yang terencana dan teratur, mulai dari analisis kebutuhan hingga pengujian sistem.

### 3.2.1 Pengumpulan Data

Pengumpulan data dilakukan dengan mengakses dan mengumpulkan data nilai murid SMK 16 Muhammadiyah Desa Pon. Data yang dikumpulkan mencakup informasi data siswa, data variabel analisis penilaian, data nilai variabel siswa, data hasil analisis klasifikasi. Proses pengumpulan data dilakukan sesuai dengan batasan masalah yang telah ditetapkan, yakni periode 2024-2025 selama satu semester.

Menurut (Septiani & Dahria, 2021) dan (Dwika Asrani et al., 2024), dalam menentukan siswa yang akan mengikuti olimpiade matematika dan sains, diperlukan data yang relevan, objektif, dan terukur sebagai dasar dalam proses pengambilan klasifikasi. Adapun data tersebut meliputi sebagai berikut:

**Tabel 3.1 Data Penelitian**

<b>No.</b>	<b>Data Siswa</b>	<b>Kategori</b>
1.	Nilai Matematika	- Tinggi - Sedang - Rendah
2	Nilai Teknik (Sains Komputer)	- Tinggi - Sedang - Rendah
3	Nilai rata-rata semester	- Tinggi - Sedang - Rendah
4	Pemahaman	- Sangat paham - Paham - Cukup paham
5	Perilaku	- Sangat baik - Baik

Kategori paham dibagi mejadi 3 kategori yang penilaiannya berdasarkan, Selalu menjawab dan menyimpulkan tanpa melihat catatan (sangat paham), Kadang kadang menjawab dan menyimpulkan tanpa melihat catatan (paham), Tidak pernah menjawab dan menyimpulkan dengan melihat catatan (cukup paham).

### **3.2.2 Pengolahan Data**

Berdasarkan langkah-langkah penelitian, informasi dikumpulkan untuk diolah. Pengolahan informasi melibatkan pengaturan data agar bisa dimanfaatkan untuk analisis, pengambilan keputusan, atau penyampaian informasi yang berguna. Misalkan ada contoh data *sample* seperti dibawah ini:

Tabel 3.2 Contoh data sample nilai rekap siswa

<b>NAMA SISWA</b>	<b>MU</b>	<b>TKDJ</b>	<b>Pemahaman</b>	<b>Prilaku</b>	<b>Nilai Rata-rata 1 Semester</b>	<b>Kelas</b>
ALDI SAPUTRA	88	81	Cukup paham	Baik	86	Tidak ikut olimp
ALVIN KURNIAWAN	88	83	Paham	Baik	86	Tidak ikut olimp
ALYA SILVANA	90	81	Cukup paham	Baik	85	Tidak ikut olimp
ARFAH NADIA LUDFI	90	86	Paham	Sangat baik	88	Sains
ASYIFA KEYSA SABILLA	91	86	Paham	Baik	88	Matematika
AULIA RAMADANI	91	86	Paham	Baik	88	Sains
DAFINA ANDINI PUTRI	92	83	Paham	Sangat Baik	87	Matematika
DINA KARINA BR SITANGGANG	90	81	Cukup paham	Sangat Baik	87	Tidak ikut olimp
DINDA LAILA	90	83	Paham	Baik	86	Matematika
DWI AJENG PRATIWI	91	85	Paham	Sangat Baik	88	Sains
EKA JAINI ANWAR	91	85	Paham	Sangat Baik	88	Matematika
FANY HAZANIA	91	83	Paham	Baik	87	Matematika
FIKA FADILLA	90	82	Sangat paham	Sangat Baik	88	Matematika
HASANAH	89	83	Paham	Baik	87	Tidak ikut olimp
INA PUTRI SUSANTI	90	83	Sangat paham	Sangat Baik	88	Tidak ikut olimp

ISMA TRI CAHYA	89	81	Cukup paham	Sangat Baik	87	Tidak ikut olimp
KHUSNUL KHATIMAH	90	80	Cukup paham	Sangat Baik	83	Tidak ikut olimp
LISA ANANDA	90	81	Cukup paham	Baik	87	Tidak ikut olimp
LISA MERRY ANDANI	89	84	Paham	Baik	88	Tidak ikut olimp
LISNAWATI	92	86	Sangat paham	Sangat Baik	89	Matematika
LUTHVI ALHABIP	89	86	Paham	Baik	88	Sains
MELATI SUKMA	90	83	Paham	Sangat Baik	87	Tidak ikut olimp
MIRA TARI	90	83	Paham	Sangat baik	87	Tidak ikut olimp
Muhammad Fauzan Hilmy	89	82	Paham	Baik	87	Tidak ikut olimp
mulisa	92	90	Sangat paham	Sangat Baik	89	Sains
NABIL AZEHARI	89	82	Cukup paham	Baik	87	Tidak ikut olimp
NUR JANAHA	89	82	Cukup Paham	Baik	86	Tidak ikut olimp
RAIHAN MAULANA	88	80	Cukup Paham	Baik	87	Tidak ikut olimp
Reva rialdo	88	81	Cukup Paham	Baik	87	Tidak ikut olimp
SITI ZAHRA	90	86	Sangat paham	Sangat Baik	88	Sains
SUKMA ULFA HAIRANI	90	83	Paham	Baik	87	Tidak ikut olimp
TASYA AULIA	91	85	Sangat paham	Sangat Baik	88	Matematika

WIDI	90	86	Sangat paham	Baik	88	Sains
YESI APRIANI	88	84	Cukup paham	Baik	87	Tidak ikut olimp
ZULFAN	88	81	Paham	Sangat Baik	88	Tidak ikut olimp

Mu pada tabel diatas mengartikan atribut nilai matematika, sedangkan TKDJ mengartikan atribut nilai sains computer.

### 3.2.3 Transformasi Data

Data numerik berupa nilai matematika, nilai Teknik (sains computer), dan nilai rata-rata semester didiskretisasi menjadi tiga kategori, yaitu Tinggi, Sedang, dan Rendah. Proses diskretisasi ini dilakukan untuk memudahkan penerapan algoritma C4.5 dalam pembentukan pohon keputusan. Adapun skema diskretisasi yang dipakai adalah

**Tabel 3.3 Skema Diskretisasi**

Nilai	Kategori
$\geq 90$	Tinggi
85-89	Sedang
$< 85$	Rendah

Ini adalah contoh hasil dari pengolahan data nilai siswa yang telah ditransformasikan:

**Tabel 3.4 Rekap Nilai yang telah ditransformasi**

NAMA SISWA	MU	TKDJ	Pemahaman	Prilaku	Nilai Rata-rata 1 Semester	Kelas
ALDI SAPUTRA	Sedang	rendah	Cukup paham	Baik	Sedang	Tidak ikut olimp
ALVIN KURNIAWAN	Sedang	rendah	Paham	Baik	Sedang	Tidak ikut olimp

ALYA SILVANA	Tinggi	rendah	Cukup paham	Baik	Sedang	Tidak ikut olimp
ARFAH NADIA LUDFI	Tinggi	sedang	Paham	Sangat baik	Sedang	Sains
ASYIFA KEYSA SABILLA	Tinggi	sedang	Cukup paham	Baik	Sedang	Tidak ikut olimp
AULIA RAMADANI	Tinggi	sedang	Paham	Baik	Sedang	Sains
DAFINA ANDINI PUTRI	Tinggi	rendah	Paham	Sangat Baik	Sedang	Matemati ka
DINA KARINA BR SITANGGA NG	Tinggi	rendah	Cukup paham	Sangat Baik	Sedang	Tidak ikut olimp
DINDA LAILA	Tinggi	rendah	Paham	Baik	Sedang	Matemati ka
DWI AJENG PRATIWI	Tinggi	sedang	Paham	Sangat Baik	Sedang	Sains
EKA JAINI ANWAR	Tinggi	sedang	Paham	Sangat Baik	Sedang	Matemati ka
FANY HAZANIA	Tinggi	rendah	Paham	Baik	Sedang	Matemati ka
FIKA FADILLA	Tinggi	rendah	Sangat paham	Sangat Baik	Sedang	Matemati ka
HASANA H	Sedang	rendah	Paham	Baik	Sedang	Tidak ikut olimp
INA PUTRI SUSANTI	Tinggi	rendah	Sangat paham	Sangat Baik	Sedang	Tidak ikut olimp
ISMA TRI CAHYA	Sedang	rendah	Cukup paham	Sangat Baik	Sedang	Tidak ikut olimp
KHUSNUL KHATIMAH	Tinggi	rendah	Cukup paham	Sangat Baik	Rendah	Tidak ikut olimp
LISA ANANDA	Tinggi	rendah	Cukup paham	Baik	Sedang	Tidak ikut olimp
LISA MERRY ANDANI	Sedang	rendah	Paham	Baik	Sedang	Tidak ikut olimp
LISNAWATI	Tinggi	sedang	Sangat paham	Sangat Baik	Sedang	Matemati ka
LUTHVI ALHABIP	Sedang	sedang	Paham	Baik	Sedang	Sains
MELATI SUKMA	Tinggi	rendah	Paham	Sangat Baik	Sedang	Tidak ikut olimp
MIRA TARI	Tinggi	rendah	Paham	Sangat baik	Sedang	Tidak ikut olimp
Muhammad Fauzan Hilmy	Sedang	rendah	Paham	Baik	Sedang	Tidak ikut olimp
mulisa	Tinggi	tinggi	Sangat paham	Sangat Baik	Sedang	Sains

NABIL AZEHARI	Sedang	rendah	Cukup paham	Baik	Sedang	Tidak ikut olimp
NUR JANAHAH	Sedang	rendah	Cukup Paham	Baik	Sedang	Tidak ikut olimp
RAIHAN MAULANA	Sedang	rendah	Cukup Paham	Baik	Sedang	Tidak ikut olimp
Reva rialdo	Sedang	rendah	Cukup Paham	Baik	Sedang	Tidak ikut olimp
SITI ZAHRA	Tinggi	sedang	Sangat paham	Sangat Baik	Sedang	Sains
SUKMA ULFA HAIRANI	Tinggi	rendah	Paham	Baik	Sedang	Tidak ikut olimp
TASYA AULIA	Tinggi	sedang	Sangat paham	Sangat Baik	Sedang	Matemati ka
WIDI	Tinggi	sedang	Sangat paham	Baik	Sedang	Sains
YESI APRIANI	Sedang	rendah	Cukup paham	Baik	Sedang	Tidak ikut olimp
ZULFAN	Sedang	rendah	Paham	Sangat Baik	Sedang	Tidak ikut olimp

### 3.2.4 Penerapan Algoritma C4.5

Implementasi algoritma C4. 5 melibatkan serangkaian langkah, seperti memilih atribut utama, memisahkan dataset, menghitung entropi, dan menghitung gain. Atribut utama ditentukan sebagai dasar untuk memisahkan dataset, dan langkah ini diulang sampai terbentuk pohon keputusan. Data dipisahkan berdasarkan nilai dari atribut, dan entropi dihitung untuk mengevaluasi keragaman data. Gain dihitung untuk setiap atribut, dan atribut dengan nilai gain tertinggi akan dipilih. Proses ini terus berlangsung hingga terbentuk pohon keputusan yang bisa digunakan untuk klasifikasi atau prediksi.

#### a. Perhitungan Entropy

Tahap ini bertujuan untuk mengukur seberapa beragam data dalam setiap subset. Entropi adalah ukuran ketidakpastian atau ketidakteraturan dari data. Entropi semakin tinggi jika data semakin heterogen, dan semakin rendah jika data semakin homogen. Entropi dapat dihitung dengan rumus berikut:

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

di mana E(S) merujuk pada entropi dari sekumpulan S, n menunjukkan total kelas yang terdapat dalam data, dan  $p_i$  adalah bagian dari data yang tergolong dalam kelas  $i$ . Contoh penghitungan nilai *entropy* dan *gain* terhadap data *sample* adalah sebagai berikut:

1) *Entropy* Total

$$\begin{aligned} Entropy \text{ [Total]} &: \left(-\frac{21}{35} \times \log_2 \frac{21}{35}\right) + \left(-\frac{7}{35} \times \log_2 \frac{7}{35}\right) + \left(-\frac{7}{35} \times \log_2 \frac{7}{35}\right) \\ &: 1,37095 \end{aligned}$$

2) *Entropy* MUKategori

$$\begin{aligned} - \text{ Tinggi} &: \left(-\frac{9}{22} \times \log_2 \frac{9}{22}\right) + \left(-\frac{7}{22} \times \log_2 \frac{7}{22}\right) + \left(-\frac{6}{22} \times \log_2 \frac{6}{22}\right) \\ &: 1,56441 \end{aligned}$$

$$\begin{aligned} - \text{ Sedang} &: \left(-\frac{12}{13} \times \log_2 \frac{12}{13}\right) + \left(-\frac{0}{13} \times \log_2 \frac{0}{13}\right) + \left(-\frac{1}{13} \times \log_2 \frac{1}{13}\right) \\ &: 0 \end{aligned}$$

$$\begin{aligned} - \text{ Rendah} &: \left(-\frac{0}{0} \times \log_2 \frac{0}{0}\right) + \left(-\frac{0}{0} \times \log_2 \frac{0}{0}\right) + \left(-\frac{0}{0} \times \log_2 \frac{0}{0}\right) \\ &: 0 \end{aligned}$$

3) *Entropy* TKDJKategori

$$\begin{aligned} - \text{ Tinggi} &: \left(-\frac{0}{1} \times \log_2 \frac{0}{1}\right) + \left(-\frac{0}{1} \times \log_2 \frac{0}{1}\right) + \left(-\frac{1}{1} \times \log_2 \frac{1}{1}\right) \\ &: 0 \end{aligned}$$

$$\begin{aligned} - \text{ Sedang} &: \left(-\frac{1}{10} \times \log_2 \frac{1}{10}\right) + \left(-\frac{3}{10} \times \log_2 \frac{3}{10}\right) + \left(-\frac{6}{10} \times \log_2 \frac{6}{10}\right) \\ &: 1,29546 \end{aligned}$$

$$\begin{aligned} - \text{ Rendah} &: \left(-\frac{21}{24} \times \log_2 \frac{21}{24}\right) + \left(-\frac{3}{24} \times \log_2 \frac{3}{24}\right) + \left(-\frac{0}{24} \times \log_2 \frac{0}{24}\right) \\ &: 0 \end{aligned}$$

4) *Entropy* RatarataKategori

$$\begin{aligned} - \text{ Tinggi} & : (-\frac{0}{0} \times \log_2 \frac{0}{0}) + (-\frac{0}{0} \times \log_2 \frac{0}{0}) + (-\frac{0}{0} \times \log_2 \frac{0}{0}) \\ & : 0 \end{aligned}$$

$$\begin{aligned} - \text{ Sedang} & : (-\frac{20}{34} \times \log_2 \frac{20}{34}) + (-\frac{7}{34} \times \log_2 \frac{7}{34}) + (-\frac{7}{34} \times \log_2 \frac{7}{34}) + \\ & : 1,38918 \end{aligned}$$

$$\begin{aligned} - \text{ Rendah} & : (-\frac{1}{1} \times \log_2 \frac{1}{1}) + (-\frac{0}{1} \times \log_2 \frac{0}{1}) + (-\frac{0}{1} \times \log_2 \frac{0}{1}) \\ & : 0 \end{aligned}$$

5) *Entropy* Pemahaman :

$$\begin{aligned} - \text{ Sangat Paham} & : (-\frac{1}{7} \times \log_2 \frac{1}{7}) + (-\frac{3}{7} \times \log_2 \frac{3}{7}) + (-\frac{3}{7} \times \log_2 \frac{3}{7}) \\ & : 1,44882 \end{aligned}$$

$$\begin{aligned} - \text{ Paham} & : (-\frac{10}{18} \times \log_2 \frac{10}{18}) + (-\frac{4}{18} \times \log_2 \frac{4}{18}) + (-\frac{4}{18} \times \\ & \log_2 \frac{4}{18}) \\ & : 1,43552 \end{aligned}$$

$$\begin{aligned} - \text{ Cukup Paham} & : (-\frac{10}{10} \times \log_2 \frac{10}{10}) + (-\frac{0}{10} \times \log_2 \frac{0}{10}) + (-\frac{0}{10} \times \\ & \log_2 \frac{0}{10}) \\ & : 0 \end{aligned}$$

6) *Entropy* Perilaku:

$$\begin{aligned} - \text{ Sangat Baik} & : (-\frac{7}{16} \times \log_2 \frac{7}{16}) + (-\frac{5}{16} \times \log_2 \frac{5}{16}) + (-\frac{4}{16} \times \\ & \log_2 \frac{4}{16}) \\ & : 1,54618 \end{aligned}$$

$$\begin{aligned}
 - \text{Baik} & : \left(-\frac{14}{19} \times \log_2 \frac{14}{19}\right) + \left(-\frac{2}{19} \times \log_2 \frac{2}{19}\right) + \left(-\frac{3}{19} \times \log_2 \frac{3}{19}\right) \\
 & : 1,08699
 \end{aligned}$$

b. Perhitungan *Gain*

Tahap ini bertujuan untuk menilai seberapa banyak pengurangan entropi yang bisa diperoleh dengan memilih atribut yang digunakan untuk membagi data. Keuntungan merupakan perbedaan antara entropi sebelum dan sesudah data dipisahkan. Keuntungan dapat dihitung dengan menggunakan rumus berikut:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{S_i}{S} \times \text{Entropy}(S_i)$$

di mana  $G(S,A)$  merupakan keuntungan dari subset  $S$  dengan memanfaatkan atribut  $A$ ,  $|S_i|$  adalah bagian dari  $S$  yang memiliki nilai  $i$  untuk atribut  $A$ , dan  $|S|$  adalah jumlah keseluruhan data dalam subset  $S$ .

Contoh penghitungan nilai *gain* dan *gain* terhadap data *sample* adalah

1) MUKategori

$$\begin{aligned}
 & = 1,37095 - \left(\left(\frac{22}{35}\right) \times 1,56441\right) + \left(\left(\frac{13}{35}\right) \times 0\right) + \left(\left(\frac{0}{35}\right) \times 0\right) \\
 & = 0,387607
 \end{aligned}$$

2) TKDJKategori

$$\begin{aligned}
 & = 1,37095 - \left(\left(\frac{1}{35}\right) \times 0\right) + \left(\left(\frac{10}{35}\right) \times 1,29546\right) + \left(\left(\frac{24}{35}\right) \times 0\right) \\
 & = 1,00082
 \end{aligned}$$

3) RatarataKategori

$$= 1,37095 - \left(\left(\frac{0}{35}\right) \times 0\right) + \left(\left(\frac{34}{35}\right) \times 1,38918\right) + \left(\left(\frac{1}{35}\right) \times 0\right)$$

$$= 0,0214609$$

4) PemahamanKategori

$$= 1,37095 - \left(\frac{7}{35}\right) \times 1,44882 + \left(\frac{18}{35}\right) \times 1,43552 + \left(\frac{10}{35}\right) \times 0$$

$$= 0,342919$$

5) PerilakuKategori

$$= 1,37095 - \left(\frac{16}{35}\right) \times 1,54618 + \left(\frac{19}{35}\right) \times 1,08699$$

$$= 0,0740446$$

Untuk menentukan atribut yang dijadikan akar, dilakukan berdasarkan nilai gain tertinggi dari atribut yang ada. Entropy dimanfaatkan dalam proses perhitungan nilai gain. Tabel di bawah ini menyajikan hasil dari penghitungan nilai *entropy* dan *gain*:

**Tabel 3.5 Hasil Perhitungan *Gain* dan *Entropy***

Atribut	Keterangan	Total Kasus	Kelas Olimp MTK	Kelas Olimp Sains	Tidak Ikut Kelas	Entropy	Gain
Total MUKategori		35	8	7	20	1,37095	<b>0,387607</b>
	Tinggi	22	9	7	6	1,56441	
	Sedang	13	12	0	1	0	
	Rendah	0	0	0	0	0	
TKDJKategori							<b>1,00082</b>
	Tinggi	1	0	0	1	0	
	Sedang	10	1	3	6	1,29546	
	Rendah	24	21	3	0	0	
RatarataKategori							<b>0,0214609</b>
	Tinggi	0	0	0		0	
	Sedang	34	20	7	7	1,38918	
	Rendah	1	1	1	1	0	
PemahamanKategori							<b>0,342919</b>
	Sangat Paham	7	1	3	2	1,44882	
	Paham	18	10	4	4	1,43552	

Kategori Perilaku	Cukup Paham	10	10	0	0	0	<b>0,0740446</b>
	Sangat Baik	16	7	5	4	1,54618	
	Baik	19	14	2	3	1,08699	

Mengacu pada perhitungan yang terdapat dalam tabel sebelumnya, atribut "kategori" menunjukkan nilai keuntungan paling tinggi. Karena itu, atribut ini bisa dimanfaatkan sebagai simpul utama. Nilai keuntungan paling tinggi juga sesuai dengan atribut "TKDJKategori", yang akan dipilih sebagai simpul utama.

c. Pembentukan pohon Keputusan

Tahap ini ditujukan untuk menyusun pohon keputusan yang menggambarkan keterkaitan antara atribut dan kategori data. Pohon keputusan terbuat dari simpul dan cabang. Simpul menggambarkan atribut yang digunakan untuk memisahkan data, sementara cabang menunjukkan nilai-nilai dari atribut tersebut. Simpul yang tidak memiliki cabang dikenal sebagai simpul daun dan menggambarkan kategori yang terdapat dalam data. Pohon keputusan dapat dimanfaatkan untuk mengklasifikasikan atau memprediksi data baru dengan mengikuti cabang yang sesuai dengan nilai atribut dari data tersebut.

### 3.3 Pemodelan dan Perancangan Sistem

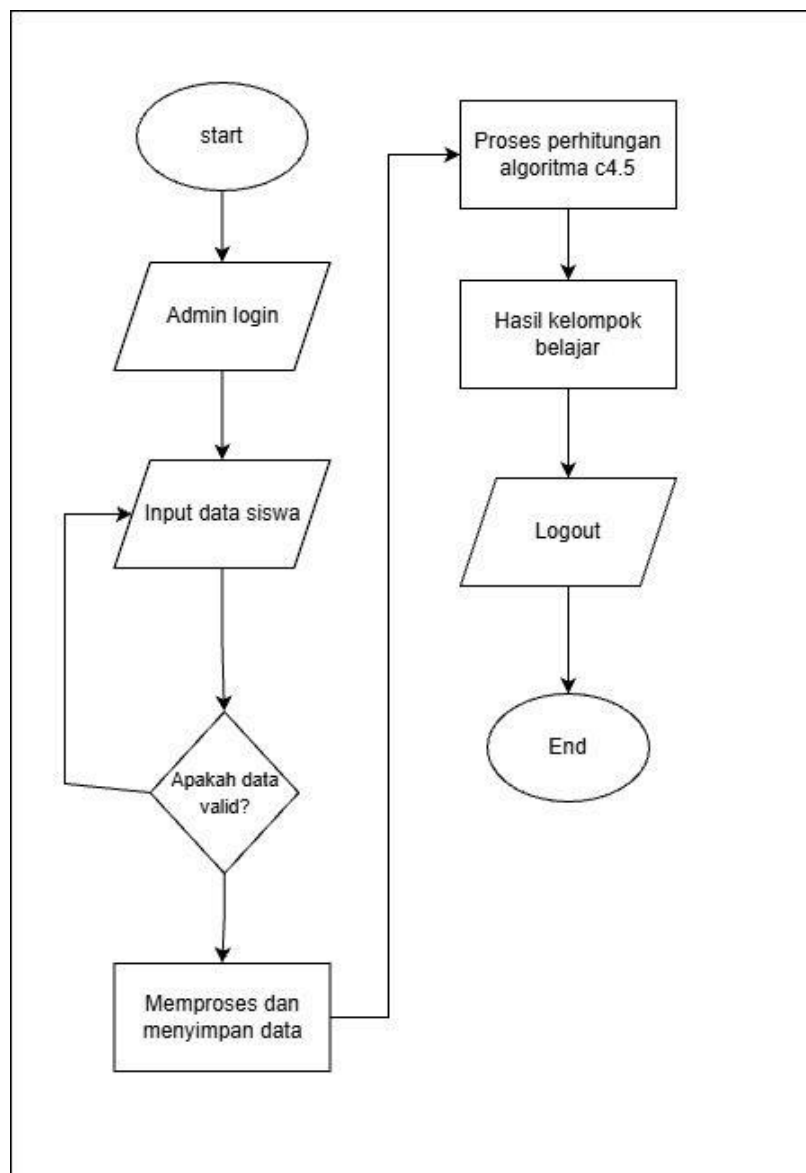
Perancangan sistem dilakukan untuk menggambarkan alur kerja sistem dalam menentukan kelompok belajar olimpiade menggunakan algoritma C4.5. Sistem ini dirancang untuk mengolah data siswa yang meliputi nilai MU, TKDJ, nilai rata-rata raport, dan minat siswa. Data numerik terlebih dahulu didiskretisasi agar dapat

diproses menggunakan algoritma C4.5. Hasil akhir dari sistem ini berupa aturan (*rule*) dan rekomendasi kelompok belajar olimpiade siswa.

### 3.3.1 Pemodelan sistem

Pemodelan sistem digunakan untuk mempermudah pemahaman alur proses dan interaksi dalam sistem. Pemodelan dilakukan menggunakan *Flowchart*, *Use Case Diagram*, dan *UML Diagram*.

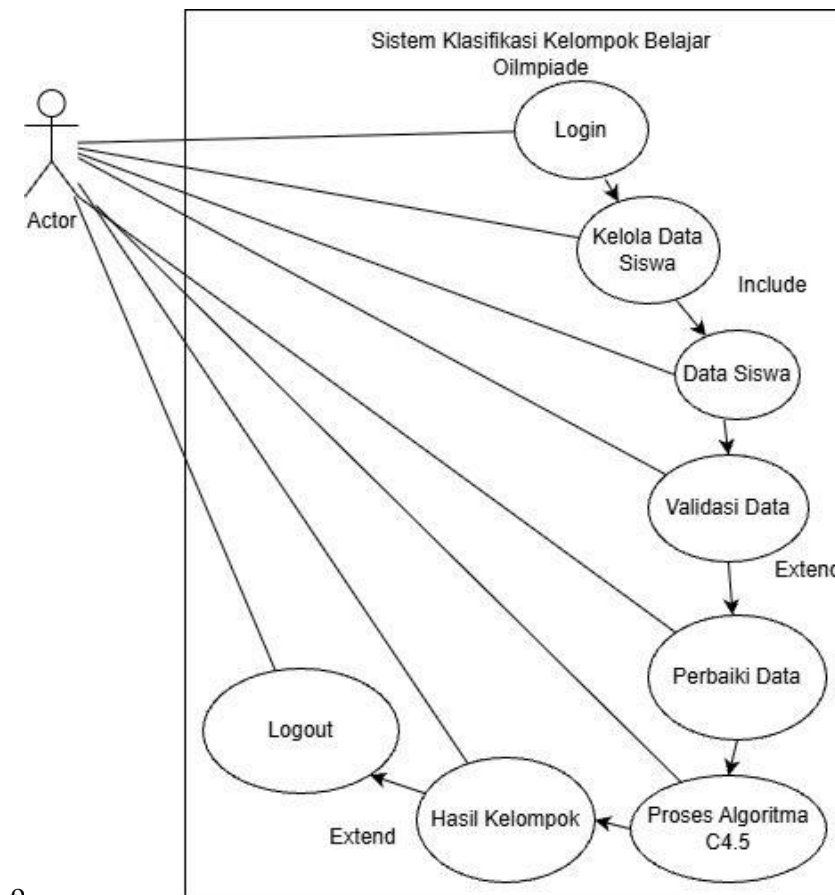
#### a. *Flowchart* Sistem Klasifikasi Kelompok Belajar



**Gambar 3.1** *Flowchart* Sistem Klasifikasi Kelompok Belajar

Diagram alur sistem dimulai dengan tahap *login*, kemudian administrator mengakses sistem. Setelah mendapatkan akses, administrator menyimpan informasi siswa. Selanjutnya, sistem menilai apakah informasi yang dimasukkan benar. Apabila informasi tersebut tidak benar, sistem mengulang proses ke langkah entri informasi siswa agar diperbaiki. Jika informasi tersebut benar, sistem akan memproses dan menyimpannya, kemudian melanjutkan ke tahap perhitungan menggunakan algoritma C4. 5. Sistem menampilkan hasil dari perhitungan C4. 5, yaitu pengelompokan siswa ke dalam kelompok belajar mereka. Setelah hasil diperoleh, admin melakukan logout dan proses sistem diakhiri pada tahap end.

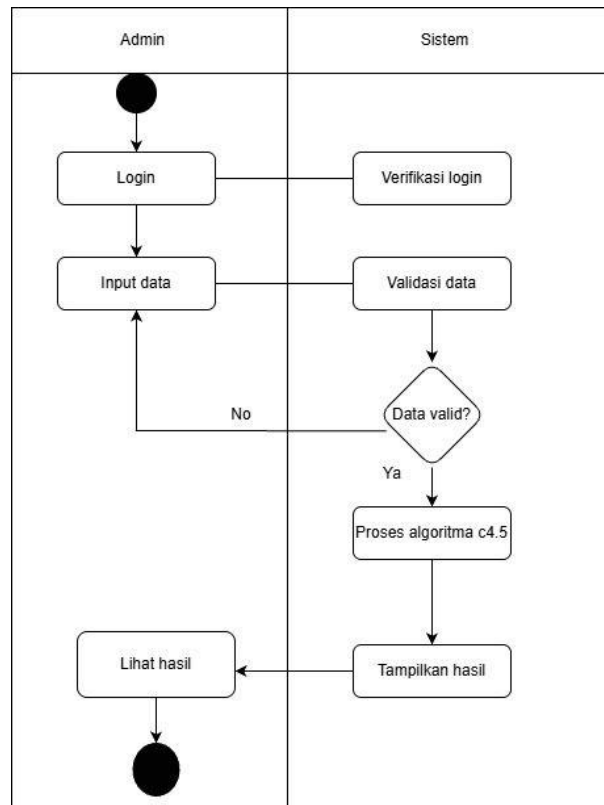
b. *Use Case* Sistem Klasifikasi Kelompok Belajar



**Gambar 3.2** *Use Case* Sistem Klasifikasi Kelompok Belajar

*Use Case Diagram* pada sistem ini melibatkan aktor Admin yang berinteraksi dengan sistem untuk mengelola data siswa dan menjalankan algoritma C4.5. Proses login merupakan use case yang selalu disertakan (*<<include>>*) pada setiap aktivitas utama. Pada pengelolaan data siswa, proses input dan validasi data merupakan use case yang wajib dijalankan, sedangkan perbaikan data merupakan use case tambahan (*<<extend>>*) yang terjadi apabila data tidak valid. Dalam proses algoritma C4.5, diskretisasi data, perhitungan entropy, dan information gain merupakan use case yang selalu disertakan (*<<include>>*). Setelah hasil kelompok belajar ditampilkan, admin dapat melakukan logout sebagai proses tambahan yang bersifat opsional (*<<extend>>*)

c. *Activity Diagram* Sistem Klasifikasi Kelompok Belajar

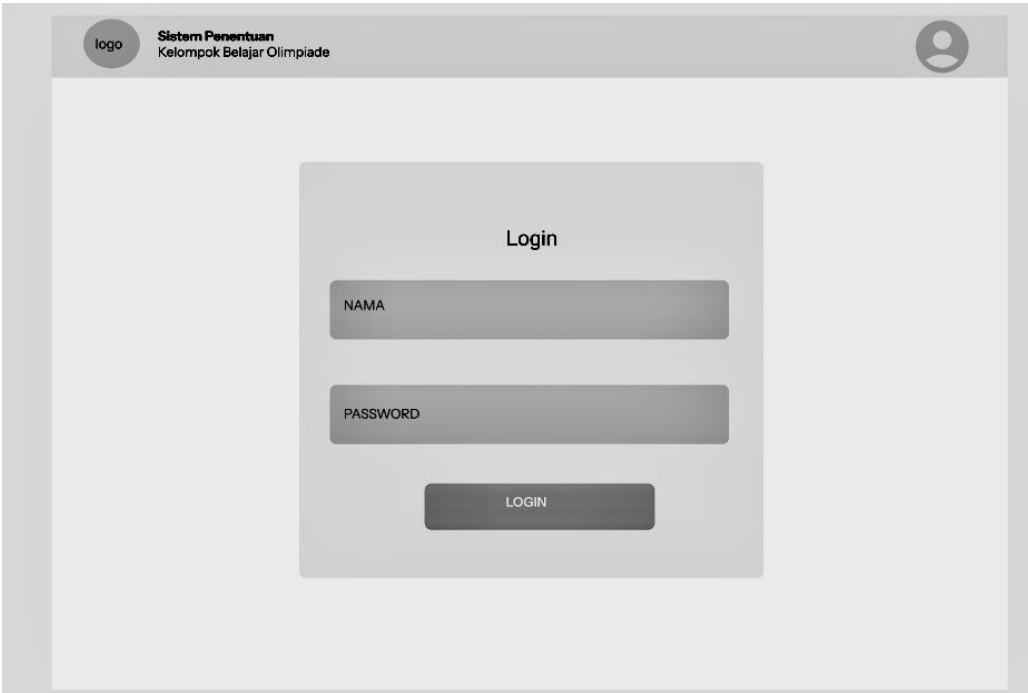


**Gambar 3.3 Activity Diagram Sistem Klasifikasi Kelompok Belajar**

*Activity Diagram* menggambarkan interaksi antara Admin dan Sistem dalam menentukan kelompok belajar siswa. Proses diawali oleh admin dengan melakukan login ke sistem, kemudian sistem melakukan verifikasi login. Setelah berhasil, admin menginput data siswa yang selanjutnya divalidasi oleh sistem. Apabila data tidak valid, sistem akan memberikan pesan kesalahan dan mengarahkan admin untuk melakukan perbaikan data. Jika data valid, sistem menyimpan data dan melanjutkan proses diskretisasi serta perhitungan menggunakan algoritma C4.5. Hasil pengelompokan ditampilkan kepada admin, kemudian admin melakukan logout dan proses sistem berakhir.

### 3.4 Perancangan Antarmuka

#### 1. Halaman *Login*

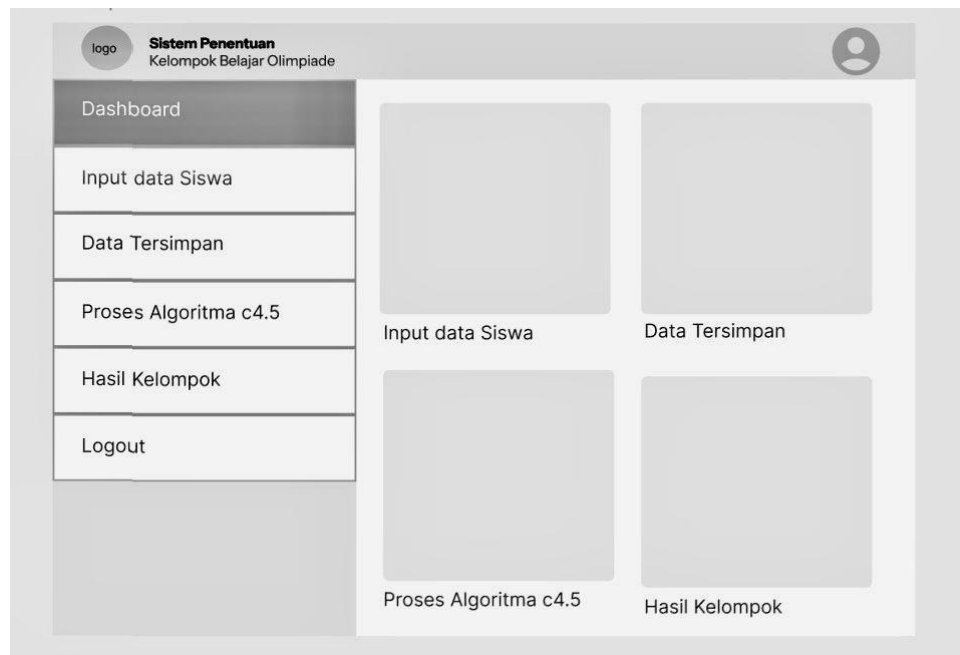


The image shows a web application interface for a login page. At the top left, there is a circular logo with the word "logo" inside. To its right, the text "Sistem Penentuan Kelompok Belajar Olimpiade" is displayed. At the top right, there is a user profile icon. The main content area is a light gray box containing a darker gray box with the title "Login". Below the title are two input fields: the first is labeled "NAMA" and the second is labeled "PASSWORD". Below these fields is a button labeled "LOGIN".

**Gambar 3.4 Halaman Login**

Halaman *login* adalah tampilan pertama dari sistem yang berfungsi untuk mengamankan akses aplikasi. Di halaman ini, admin diminta untuk menginput nama pengguna dan kata sandi yang sudah terdaftar. Sistem akan menjalankan proses autentikasi untuk memastikan hanya pengguna yang berhak yang dapat memasuki sistem penentuan kelompok belajar olimpiade. Apabila data login benar, admin akan diarahkan ke halaman utama, sementara jika tidak cocok, sistem akan menolak akses.

## 2. Halaman *Dashboard*



**Gambar 3.5 Halaman Dashboard**

Halaman dasbor berperan sebagai halaman utama setelah admin berhasil masuk. Di halaman ini, terdapat menu navigasi yang mencakup Input Data Siswa, Data yang Disimpan, Proses Algoritma C4. 5, Hasil Klasifikasi, dan Logout. Dasbor memudahkan admin untuk mengakses semua fitur dalam sistem serta memberikan gambaran umum tentang cara kerja sistem secara keseluruhan.

## 3. Halaman Input Data Siswa

Halaman input data siswa digunakan oleh admin untuk memasukkan data yang diperlukan dalam proses penentuan kelompok belajar. Data yang diinput meliputi nama siswa, nilai mata pelajaran Matematika, nilai mata pelajaran lainnya, rata-rata nilai rapor, serta minat siswa. Setelah data diisi, admin dapat menyimpan data dengan menekan tombol simpan. Data yang telah disimpan akan digunakan sebagai dataset pada proses perhitungan algoritma C4.5.

**Sistem Penentuan Kelompok Belajar Olimpiade**

Dashboard

**Input data Siswa**

Data Tersimpan

Proses Algoritma c4.5

Hasil Kelompok

Logout

### Input Data Siswa

Nama Siswa

Nilai Matematika

Rata-rata Raport

Nilai Sains Komputer

Pemahaman

Perilaku

Simpan Data

**Gambar 3.6 Halaman Input Data Siswa**

#### 4. Halaman Data Tersimpan

**Sistem Penentuan Kelompok Belajar Olimpiade**

Dashboard

Input data Siswa

**Data Tersimpan**

Proses Algoritma c4.5

Hasil Kelompok

Logout

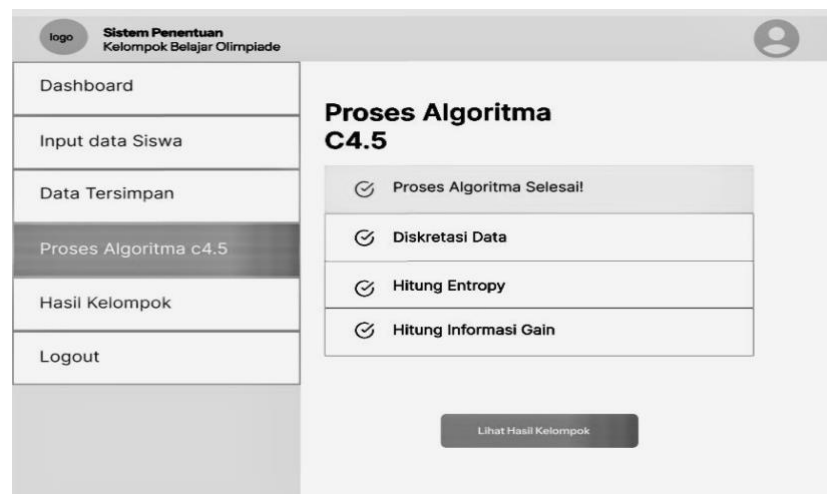
### Data Tersimpan

Nama Siswa	Nilai Matematika	Nilai Sains Komputer	Rata-rata Raport	Pemahaman	Perilaku
					<input type="checkbox"/>

**Gambar 3.7 Halaman Data Tersimpan**

Halaman Data Disimpan berfungsi untuk menampilkan semua informasi siswa yang sudah berhasil dimasukkan ke dalam sistem. Di halaman ini, data disajikan dalam format tabel yang mencakup nama siswa, nilai Matematika, nilai Sains Komputer, nilai rata-rata rapor, dan minat siswa. Halaman ini juga berperan sebagai sarana untuk memeriksa data sebelum proses penghitungan menggunakan algoritma C4.5, sehingga administrator dapat memastikan bahwa informasi yang digunakan sudah lengkap dan akurat.

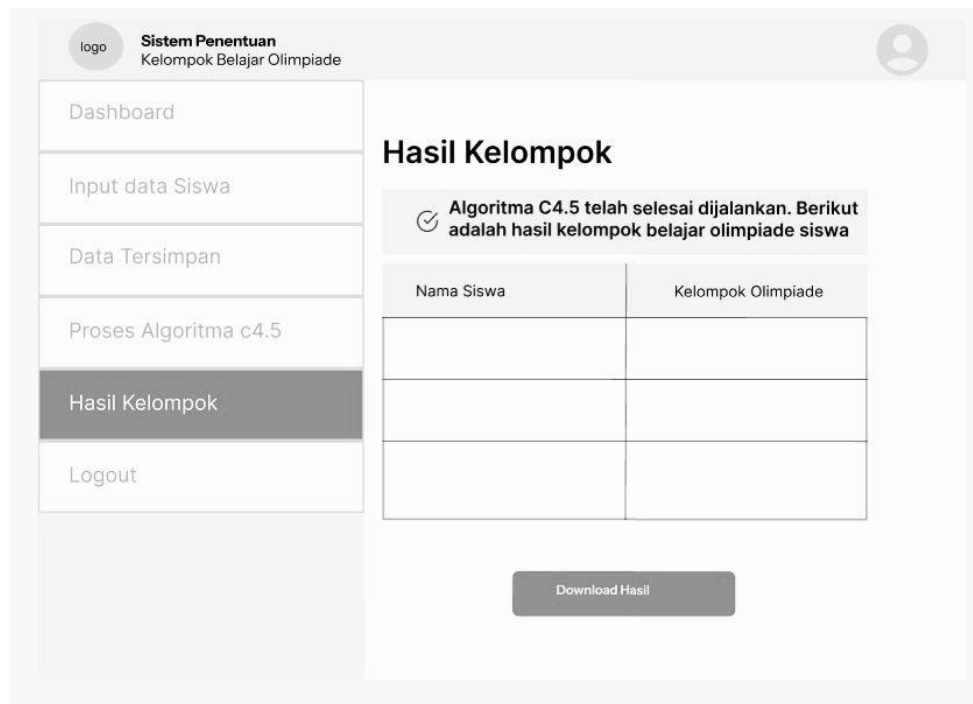
#### 5. Halaman Proses Algoritma C4.5



**Gambar 3.8 Halaman Proses Algoritma C4.5**

Halaman proses algoritma C4.5 digunakan untuk menjalankan tahapan perhitungan data mining. Pada halaman ini, sistem menampilkan tahapan proses yang dilakukan, mulai dari diskretisasi data, perhitungan entropy, hingga perhitungan information gain. Setelah seluruh proses selesai, sistem memberikan notifikasi bahwa proses algoritma C4.5 telah berhasil dijalankan dan admin dapat melanjutkan untuk melihat hasil pengelompokan.

## 6. Halaman Hasil Kelompok



**Gambar 3.9 Halaman Hasil Kelompok**

Halaman Hasil Kelompok menampilkan hasil akhir dari proses perhitungan algoritma C4.5 dalam menentukan kelompok belajar olimpiade siswa. Sistem menampilkan notifikasi bahwa proses algoritma C4.5 telah selesai dijalankan, kemudian menyajikan hasil pengelompokan dalam bentuk tabel yang berisi nama siswa dan kelompok olimpiade yang direkomendasikan. Selain itu, pada halaman ini juga tersedia tombol unduh hasil, yang memungkinkan admin untuk menyimpan hasil pengelompokan sebagai bahan dokumentasi atau laporan.



## BAB IV

### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1 Data Perhitungan Algoritma C4.5

Data yang didapatkan berasal dari tempat penelitian. Informasi yang terkumpul meliputi nilai dan tingkah laku siswa selama masa pembelajaran, termasuk tahun ajaran 2024-2025, khususnya untuk satu semester. Informasi ini mencakup elemen seperti nama, nilai matematika, nilai komputer, pemahaman, perilaku, serta rata-rata semester dan kelas Olimpiade. Informasi ini selanjutnya digunakan sebagai landasan untuk melatih dan memverifikasi model klasifikasi dengan algoritma C4. 5. Rincian informasi tersebut dapat dilihat pada tabel berikut ini:

**Tabel 4. 1 Data Siswa**

NO	NAMA SISWA	MU	TKDJ	Pemahaman	Prilaku	Nilai Rata-rata 1 Semester	Kelas
1	ALDI SAPUTRA	Sedang	rendah	Cukup paham	Baik	sedang	Tidak ikut olimp
2	ALVIN KURNIAWAN	Sedang	tinggi	Sangat paham	Sangat baik	tinggi	Sains
3	ALYA SILVANA	Tinggi	rendah	Cukup paham	Baik	sedang	Tidak ikut olimp
4	ARFAH NADIA LUDFI	Tinggi	tinggi	Paham	Sangat baik	sedang	Sains
5	ASYIFA KEYSA SABILLA	Tinggi	sedang	Paham	Baik	sedang	Tidak ikut olimp
6	AULIA RAMADANI	Tinggi	tinggi	Paham	Baik	tinggi	Sains
7	DAFINA ANDINI PUTRI	Tinggi	sedang	Paham	Sangat Baik	tinggi	Matematika



102	ZULPA	rendah	sedang	Cukup	Baik	sedang	Tidak ikut
	AULIA			paham			olimp

Data yang telah dikumpulkan selanjutnya akan melewati tahap persiapan awal dalam sistem web yang telah dirancang. Tujuan dari tahap ini adalah untuk memastikan data siap dengan format dan struktur yang diperlukan untuk proses pengelompokan menggunakan algoritma C4.5. Jumlah keseluruhan data yang digunakan dalam analisis ini mencapai 102 data. Pembagian data dalam setiap kategori sasaran (label klasifikasi) adalah sebagai berikut:

- Tidak ikut olimpiade = 65
- Sains = 19
- Matematika = 18

Total = 102 siswa

#### 4.1.1 Perhitungan Algoritma C4.5

##### a. Perhitungan Entropy

###### 1. Entropy Total

Distribusi kelas: (tidak ikut:65, mtk:19, sains:18)

$$: \left(-\frac{63}{102} \times \log_2 \frac{63}{102}\right) + \left(-\frac{20}{102} \times \log_2 \frac{20}{102}\right) + \left(-\frac{19}{102} \times \log_2 \frac{19}{102}\right)$$

$$: 1.34$$

###### 2. Entropy MUKategori

###### - Tinggi

Distribusi kelas: (tidak ikut:3, mtk:13, sains:7)

$$: \left(-\frac{3}{23} \times \log_2 \frac{3}{23}\right) + \left(-\frac{13}{23} \times \log_2 \frac{13}{23}\right) + \left(-\frac{7}{23} \times \log_2 \frac{7}{23}\right)$$

$$: 1.365$$

- Sedang

Distribusi kelas: (tidak ikut:25, mtk:4, sains:3)

$$: \left(-\frac{25}{41} \times \log_2 \frac{25}{41}\right) + \left(-\frac{6}{41} \times \log_2 \frac{6}{41}\right) + \left(-\frac{10}{41} \times \log_2 \frac{10}{41}\right)$$

: 1.337

- Rendah

Distribusi kelas: (tidak ikut:47, mtk:1, sains:0)

$$: \left(-\frac{35}{38} \times \log_2 \frac{35}{38}\right) + \left(-\frac{1}{38} \times \log_2 \frac{1}{38}\right) + \left(-\frac{2}{38} \times \log_2 \frac{2}{38}\right)$$

: 0.474

### 3. Entropy TKDJKategori

- Tinggi

Distribusi kelas: (tidak ikut:5, mtk:7, sains:11)

$$: \left(-\frac{5}{23} \times \log_2 \frac{5}{23}\right) + \left(-\frac{7}{23} \times \log_2 \frac{7}{23}\right) + \left(-\frac{11}{23} \times \log_2 \frac{11}{23}\right)$$

: 1.506

- Sedang

Distribusi kelas: (tidak ikut:34, mtk:10, sains:6)

$$: \left(-\frac{18}{34} \times \log_2 \frac{18}{34}\right) + \left(-\frac{10}{34} \times \log_2 \frac{10}{34}\right) + \left(-\frac{6}{34} \times \log_2 \frac{6}{34}\right)$$

: 1.448

- Rendah

Distribusi kelas: (tidak ikut:40, mtk:3, sains:3)

$$: \left(-\frac{40}{45} \times \log_2 \frac{40}{45}\right) + \left(-\frac{3}{45} \times \log_2 \frac{3}{45}\right) + \left(-\frac{2}{45} \times \log_2 \frac{2}{45}\right)$$

: 0.618

### 4. Entropy RatarataKategori

- Tinggi

Distribusi kelas: (tidak ikut:0, mtk:6, sains:1)

$$:(-\frac{0}{7} \times \log_2 \frac{0}{7}) + (-\frac{6}{7} \times \log_2 \frac{6}{7}) + (-\frac{1}{7} \times \log_2 \frac{1}{7})$$

: 0.937

- Sedang

Distribusi kelas: (tidak ikut:38, mtk:14, sains:8)

$$:(-\frac{38}{62} \times \log_2 \frac{38}{62}) + (-\frac{14}{62} \times \log_2 \frac{14}{62}) + (-\frac{8}{62} \times \log_2 \frac{8}{62}) +$$

: 1.295

- Rendah : tidak ikut:25, mtk:0, sains:0)

$$:(-\frac{25}{25} \times \log_2 \frac{25}{25}) + (-\frac{0}{25} \times \log_2 \frac{0}{25}) + (-\frac{0}{25} \times \log_2 \frac{0}{25})$$

: 0

##### 5. Entropy Pemahaman:

- Sangat Paham

Distribusi kelas: (tidak ikut:2, mtk:12, sains:13)

$$:(-\frac{2}{27} \times \log_2 \frac{2}{27}) + (-\frac{12}{27} \times \log_2 \frac{12}{27}) + (-\frac{13}{27} \times \log_2 \frac{13}{27})$$

: 1.302

- Paham

Distribusi kelas: (tidak ikut:15, mtk:7, sains:6)

$$:(-\frac{15}{28} \times \log_2 \frac{15}{28}) + (-\frac{7}{28} \times \log_2 \frac{7}{28}) + (-\frac{6}{28} \times \log_2 \frac{6}{28})$$

: 1.459

- Cukup Paham

Distribusi kelas: (tidak ikut:46, mtk:1, sains:0)

$$:(-\frac{46}{47} \times \log_2 \frac{46}{47}) + (-\frac{1}{47} \times \log_2 \frac{1}{47}) + (-\frac{0}{47} \times \log_2 \frac{0}{47})$$

: 0.147

6. *Entropy* Perilaku:

- Sangat Baik

Distribusi kelas: (tidak ikut:8, mtk:12, sains:13)

$$: \left(-\frac{8}{33} \times \log_2 \frac{8}{33}\right) + \left(-\frac{12}{33} \times \log_2 \frac{12}{33}\right) + \left(-\frac{13}{33} \times \log_2 \frac{13}{33}\right)$$

$$: 1.557$$

- Baik

Distribusi kelas: (tidak ikut:55, mtk:8, sains:6)

$$: \left(-\frac{55}{69} \times \log_2 \frac{55}{69}\right) + \left(-\frac{8}{69} \times \log_2 \frac{8}{69}\right) + \left(-\frac{6}{69} \times \log_2 \frac{6}{69}\right)$$

$$: 0.929$$

b. Perhitungan *gain*

- MUKategori

$$= 1.34 - \left(\left(\frac{23}{102}\right) \times 1.365\right) + \left(\left(\frac{41}{102}\right) \times 1.337\right) + \left(\frac{38}{102} \times 0.474\right)$$

$$= 0.320$$

- TKDJKategori

$$= 1.34 - \left(\left(\frac{23}{102}\right) \times 1.506\right) + \left(\left(\frac{34}{102}\right) \times 1.448\right) + \left(\frac{45}{102} \times 0.618\right)$$

$$= 0.247$$

- RatarataKategori

$$= 1.34 - \left(\left(\frac{17}{102}\right) \times 0.937\right) + \left(\left(\frac{60}{102}\right) \times 1.295\right) + \left(\frac{25}{102} \times 0\right)$$

$$= 0.424$$

- PemahamanKategori

$$= 1.34 - \left(\left(\frac{27}{102}\right) \times 1.302\right) + \left(\left(\frac{28}{102}\right) \times 1.459\right) + \left(\frac{47}{102} \times 0.148\right)$$

$$= 0.527$$

- PerilakuKategori

$$= 1.34 - \left(\left(\frac{33}{102}\right) \times 1.557\right) + \left(\left(\frac{68}{102}\right) \times 0.929\right)$$

$$= 0.209$$

Berdasarkan hasil kalkulasi nilai entropi dan *informasion gain* yang diperoleh dengan menggunakan algoritma C4.5 pada data siswa, didapatkan nilai gain untuk setiap atribut yaitu Pemahaman sebesar 0,527, Nilai Rata-rata sebesar 0,424, MU sebesar 0,320, TKDJ sebesar 0,247, dan Perilaku sebesar 0,209. Nilai gain ini menunjukkan tingkat kemampuan masing-masing atribut dalam memisahkan data ke dalam kelas yang lebih homogen.

Atribut yang memiliki nilai *gain* tertinggi adalah Pemahaman dengan nilai 0,527. Ini mengindikasikan bahwa atribut Pemahaman memberikan dampak terbesar dalam proses klasifikasi, sehingga dipilih sebagai dasar (*root*) untuk membangun pohon keputusan. Tingginya nilai gain pada atribut ini menunjukkan bahwa tingkat pemahaman siswa sangat menentukan dalam pengambilan keputusan apakah siswa akan mengikuti olimpiade atau tidak, serta penentuan bidang olimpiade yang sesuai.

Atribut dengan nilai gain terbesar kedua adalah Nilai Rata-rata dengan nilai sebesar 0,424. Hal ini menunjukkan bahwa performa akademik siswa juga memiliki pengaruh yang cukup signifikan dalam proses klasifikasi, meskipun masih berada di bawah atribut Pemahaman. Oleh karena itu, atribut ini berpotensi digunakan sebagai node lanjutan dalam pohon keputusan setelah atribut Pemahaman.

Selanjutnya, atribut MU memiliki nilai *gain* sebesar 0,320. Nilai ini menunjukkan bahwa motivasi siswa cukup berpengaruh dalam menentukan klasifikasi, namun tidak sekuat atribut Pemahaman dan Nilai Rata-rata. Dengan

demikian, atribut MU dapat digunakan sebagai atribut pendukung dalam proses percabangan berikutnya pada pohon keputusan.

Atribut TKDJ memiliki nilai *gain* sebesar 0,247, yang menunjukkan bahwa kemampuan dasar siswa juga memiliki kontribusi dalam proses klasifikasi, namun pengaruhnya lebih kecil dibandingkan atribut sebelumnya. Atribut ini tetap relevan untuk digunakan sebagai atribut tambahan dalam pembentukan cabang pohon keputusan yang lebih spesifik.

Terakhir, atribut Perilaku memiliki nilai *gain* paling rendah yaitu sebesar 0,209. Hal ini menunjukkan bahwa perilaku siswa memiliki pengaruh yang relatif kecil dalam menentukan keikutsertaan siswa dalam olimpiade dibandingkan atribut lainnya. Oleh karena itu, atribut ini tidak dipilih sebagai atribut utama dalam pembentukan pohon keputusan, namun tetap dapat digunakan sebagai atribut pelengkap pada tingkat percabangan tertentu.

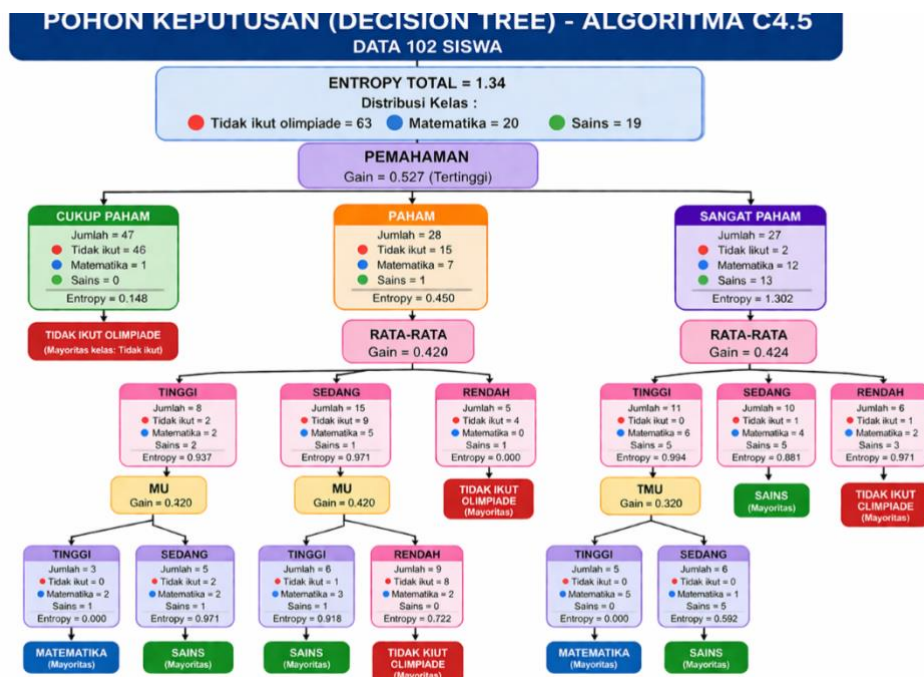
Berdasarkan keseluruhan hasil perhitungan tersebut, dapat disimpulkan bahwa atribut Pemahaman merupakan atribut paling dominan dalam proses klasifikasi, diikuti oleh Nilai Rata-rata, MU, TKDJ, dan Perilaku. Urutan nilai *gain* ini menjadi dasar dalam pembentukan struktur pohon keputusan menggunakan algoritma C4.5, di mana atribut dengan nilai *gain* tertinggi akan dipilih terlebih dahulu sebagai node utama, kemudian diikuti oleh atribut dengan nilai *gain* berikutnya pada tingkat percabangan selanjutnya.

#### **4.1.2 Pohon Keputusan**

Dalam penelitian ini, setelah dilakukan proses perhitungan nilai *entropy* dan *information gain* pada setiap atribut, langkah selanjutnya adalah membentuk model

klasifikasi menggunakan algoritma C4.5 dalam bentuk pohon keputusan (*decision tree*).

Pohon keputusan ini digunakan untuk menggambarkan proses pengambilan keputusan dalam menentukan kategori siswa, yaitu Tidak Ikut Olimpiade, Sains, dan Matematika, berdasarkan atribut yang telah dianalisis sebelumnya.



**Gambar 4. 1 Pohon Keputusan Sistem**

Berdasarkan hasil perhitungan, atribut dengan nilai *gain* tertinggi dipilih sebagai akar pohon, yaitu atribut Pemahaman. Selanjutnya dilakukan pemecahan cabang berdasarkan atribut dengan nilai gain berikutnya hingga seluruh data dapat terklasifikasi dengan baik.

Model pohon keputusan yang dihasilkan diharapkan mampu memberikan aturan (*rule*) yang jelas dan mudah dipahami dalam proses pengambilan keputusan penentuan siswa olimpiade. Adapun rule dari hasil perhitungan adalah seperti berikut:

IF Pemahaman = Cukup Paham

THEN Kelas = Tidak Ikut Olimpiade

IF Pemahaman = Paham AND Rata-rata = Tinggi AND MU = Tinggi

THEN Kelas = Matematika

IF Pemahaman = Paham AND Rata-rata = Tinggi AND MU = Sedang

THEN Kelas = Sains

IF Pemahaman = Paham AND Rata-rata = Sedang AND MU = Tinggi

THEN Kelas = Sains

IF Pemahaman = Paham AND Rata-rata = Sedang AND MU = Rendah

THEN Kelas = Tidak Ikut Olimpiade

IF Pemahaman = Paham AND Rata-rata = Rendah

THEN Kelas = Tidak Ikut Olimpiade

IF Pemahaman = Sangat Paham AND Rata-rata = Tinggi AND MU = Tinggi

THEN Kelas = Matematika

IF Pemahaman = Sangat Paham AND Rata-rata = Tinggi AND MU = Sedang

THEN Kelas = Sains

IF Pemahaman = Sangat Paham AND Rata-rata = Sedang

THEN Kelas = Sains

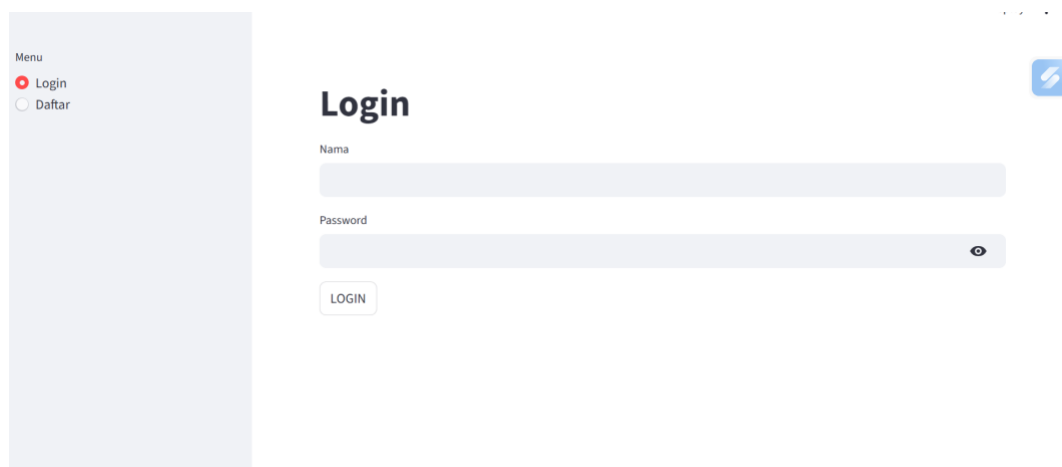
IF Pemahaman = Sangat Paham AND Rata-rata = Rendah

THEN Kelas = Tidak Ikut Olimpiade

### 4.3 Implementasi Algoritma C4.5 Berbasis Web dalam Penentuan Siswa Olimpiade

Pada tahap ini, hasil dari proses perhitungan menggunakan algoritma C4.5 diimplementasikan ke dalam sebuah sistem berbasis web. Sistem ini dirancang untuk mempermudah proses pengolahan data serta membantu dalam pengambilan keputusan penentuan siswa olimpiade secara lebih cepat, objektif, dan terstruktur.

#### 4.3.1 Halaman *Login*



**Gambar 4. 2 Tampilan Halaman *Login***

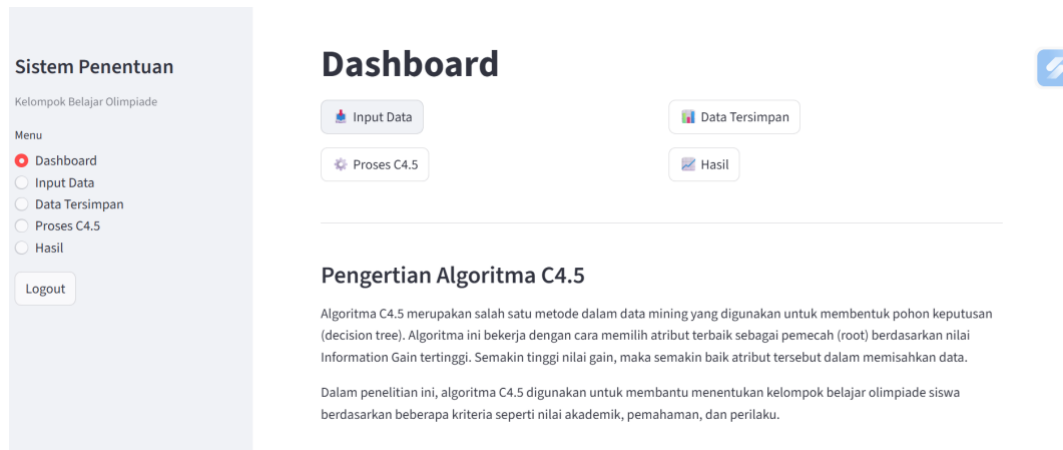
Halaman *login* adalah tampilan pertama yang digunakan oleh pegawai sekolah untuk mengakses sistem. Halaman ini berperan sebagai tahapan autentikasi agar hanya pengguna yang memiliki akun yang dapat mengakses sistem.

Di halaman ini, terdapat *form* yang memuat Nama dan Kata Sandi yang harus diisi oleh pegawai sesuai dengan akun yang telah didaftarkan. Selain itu, terdapat fitur untuk menampilkan atau menyembunyikan kata sandi yang membantu pengguna untuk memastikan bahwa data yang dimasukkan benar.

Tombol *Login* digunakan untuk memproses data yang telah diinput. Jika data sesuai, maka pengguna akan diarahkan ke halaman utama sistem. Jika tidak, maka sistem akan menolak akses dan meminta pengisian ulang data.

Halaman ini dirancang sederhana dan mudah digunakan sehingga dapat membantu staff sekolah dalam mengakses sistem secara efektif.

### 4.3.2 Tampilan *Dashboard*

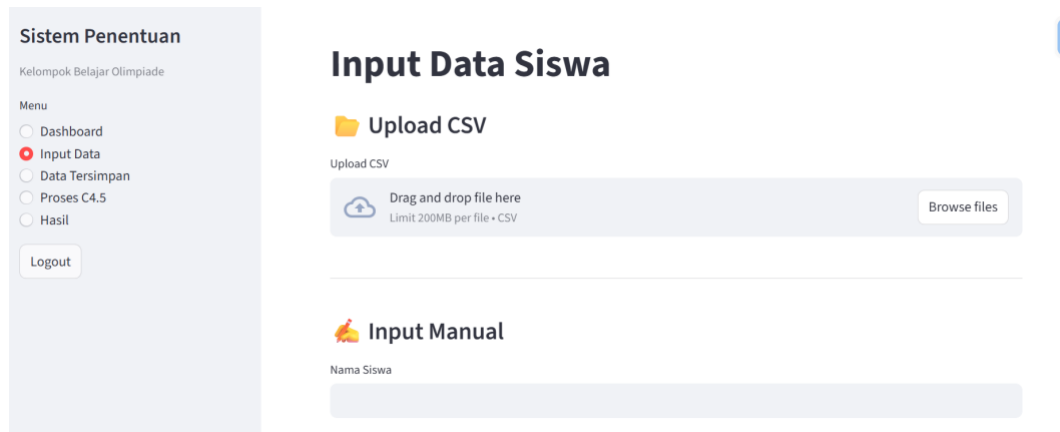


**Gambar 4. 3 Tampilan Halaman *Dashboard***

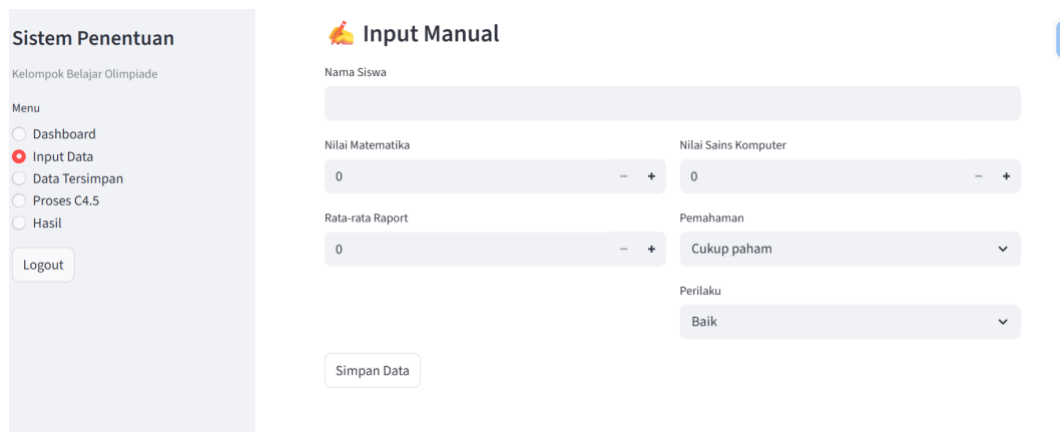
Halaman dashboard adalah halaman yang muncul pertama kali setelah staf sekolah berhasil masuk ke dalam sistem. Halaman ini berfungsi sebagai pusat untuk menavigasi dan mengakses semua fitur yang ada dalam sistem.

Di dalam halaman dashboard terdapat beberapa pilihan utama, yaitu Input Data, Data Tersimpan, Proses C4.5, dan Hasil. Menu Input Data berfungsi untuk menambahkan data siswa, menu Data Tersimpan untuk menampilkan data yang sudah dimasukkan, menu Proses C4.5 untuk melakukan perhitungan dengan menggunakan algoritma C4.5, dan menu Hasil untuk melihat klasifikasi siswa. Selain itu, pada halaman dashboard juga ditampilkan informasi mengenai pengertian algoritma C4.5 dan profil pengembang sistem.

### 4.3.3 Halaman Input Data Siswa



**Gambar 4. 4 Halaman Input**



**Gambar 4. 5 Halaman Input Manual**

Halaman untuk memasukkan data siswa adalah halaman yang dipakai oleh staff sekolah untuk menginput informasi siswa yang akan dikelola dalam sistem. Informasi yang diinput kemudian akan dimanfaatkan sebagai fondasi dalam proses pengelompokan menggunakan algoritma C4.5.

Pada halaman ini terdapat dua metode input data, yaitu upload file CSV dan input manual. *Fitur upload CSV* memungkinkan pengguna untuk memasukkan data siswa dalam jumlah banyak secara sekaligus, sehingga lebih efisien. Sedangkan fitur input manual digunakan untuk memasukkan data siswa satu per satu melalui *form* yang telah disediakan.

Form *input* manual terdiri dari beberapa *field*, seperti nama siswa, nilai akademik, rata-rata raport, tingkat pemahaman, dan perilaku. Setiap data yang diinputkan harus sesuai dengan kondisi sebenarnya agar hasil klasifikasi yang diperoleh lebih akurat.

Halaman ini dibuat dengan antarmuka yang simpel dan mudah dioperasikan, sehingga dapat mendukung staf sekolah dalam mengelola serta memasukkan informasi siswa dengan cara yang efektif dan efisien.

### 4.3.4 Halaman Data Tersimpan

**Sistem Penentuan**  
Kelompok Belajar Olimpiade

Menu

- Dashboard
- Input Data
- Data Tersimpan
- Proses C4.5
- Hasil

Logout

## Data Tersimpan

### Daftar Data Siswa

No	Nama	MU	TKDJ	Pemahaman	Perilaku	Nilai	Aksi
1	ALVIN KURNIAWAN	Sedang	tinggi	Sangat paham	Sangat baik	tinggi	
2	ALYA SILVANA	Tinggi	rendah	Cukup paham	Baik	sedang	
3	ARFAH NADIA LUDFI	Tinggi	tinggi	Paham	Sangat baik	sedang	
4	ASYIFA KEYSA SABILLA	Tinggi	sedang	Paham	Baik	sedang	
5	AULIA RAMADANI	Tinggi	tinggi	Paham	Baik	tinggi	

Deploy

43	Dwi Pratiwi	rendah	rendah	Cukup paham	Baik	Rendah	
44	EZZI DANIEL	Sedang	tinggi	Sangat paham	Sangat baik	tinggi	
45	FAJAR ABIMANYU	rendah	rendah	Cukup paham	Sangat baik	Rendah	
46	FIONA FENITA	Sedang	tinggi	Sangat paham	Sangat baik	tinggi	
47	INGGRI CLAURA INTAN	rendah	rendah	Cukup paham	Baik	Rendah	
48	IZZATA SHIDQI FADHILLAH	rendah	sedang	Paham	Sangat baik	sedang	
49	KAWIRA TRIFATANSYA	Sedang	tinggi	Paham	Baik	tinggi	
50	KAYLA ANGGUN PURBA	Sedang	tinggi	Paham	Sangat baik	sedang	

Deploy

97	VANI AULIA	rendah	sedang	Paham	Sangat baik	sedang	
98	WAHYU	rendah	sedang	Sangat paham	Sangat baik	sedang	
99	ZAVAZA ZADAWALAM	rendah	sedang	Cukup paham	Sangat baik	sedang	
100	ZELNA AYU AGUSTIN NASUTION	rendah	sedang	Cukup paham	Sangat baik	sedang	
101	ZULPA AULIA	rendah	sedang	Cukup paham	Baik	sedang	

Hapus Semua Data

Download CSV

**Gambar 4. 6 Halaman Data Tersimpan**

Halaman data tersimpan merupakan halaman yang digunakan oleh staff sekolah untuk melihat seluruh data siswa yang telah diinput ke dalam sistem. Data

ditampilkan yaitu nama siswa, nilai MU, TKDJ, pemahaman, perilaku, serta nilai rata-rata, yang nantinya digunakan dalam proses pengelompokkan menggunakan algoritma C4.5.

Data ditampilkan dalam bentuk tabel sehingga memudahkan pengguna dalam membaca dan mengelola informasi siswa. Pada setiap baris data, terdapat fitur hapus yang dapat digunakan untuk menghapus data tertentu. Selain itu, terdapat juga fitur hapus semua data untuk mengosongkan seluruh data dalam sistem, serta fitur download CSV yang memungkinkan pengguna untuk mengunduh data yang telah tersimpan.

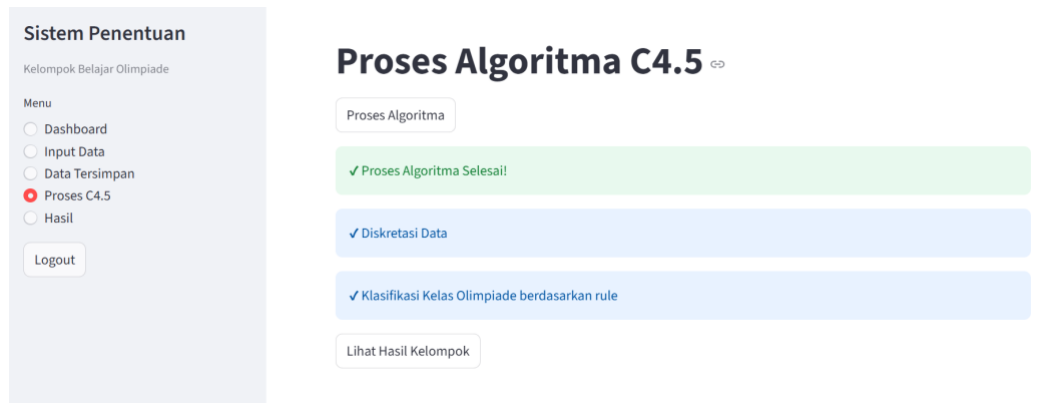
Halaman ini dirancang untuk memudahkan staff sekolah dalam melakukan pengelolaan data secara terstruktur sebelum dilakukannya proses perhitungan menggunakan algoritma C4.5.

```
def kategori_nilai(n):  
    if n >= 86:  
        return "tinggi"  
    elif n >= 85:  
        return "sedang"  
    else:  
        return "rendah"
```

**Gambar 4. 7 Code Konversi Nilai**

Sebelum memulai pengklasifikasian dengan algoritma C4.5, informasi data numerik yang berupa nilai angka terlebih dahulu diubah ke dalam bentuk kategori (diskretisasi). Proses ini dilakukan agar data dapat diolah oleh algoritma C4.5 yang bekerja dengan atribut kategorikal.

### 4.3.5 Halaman Proses Algoritma



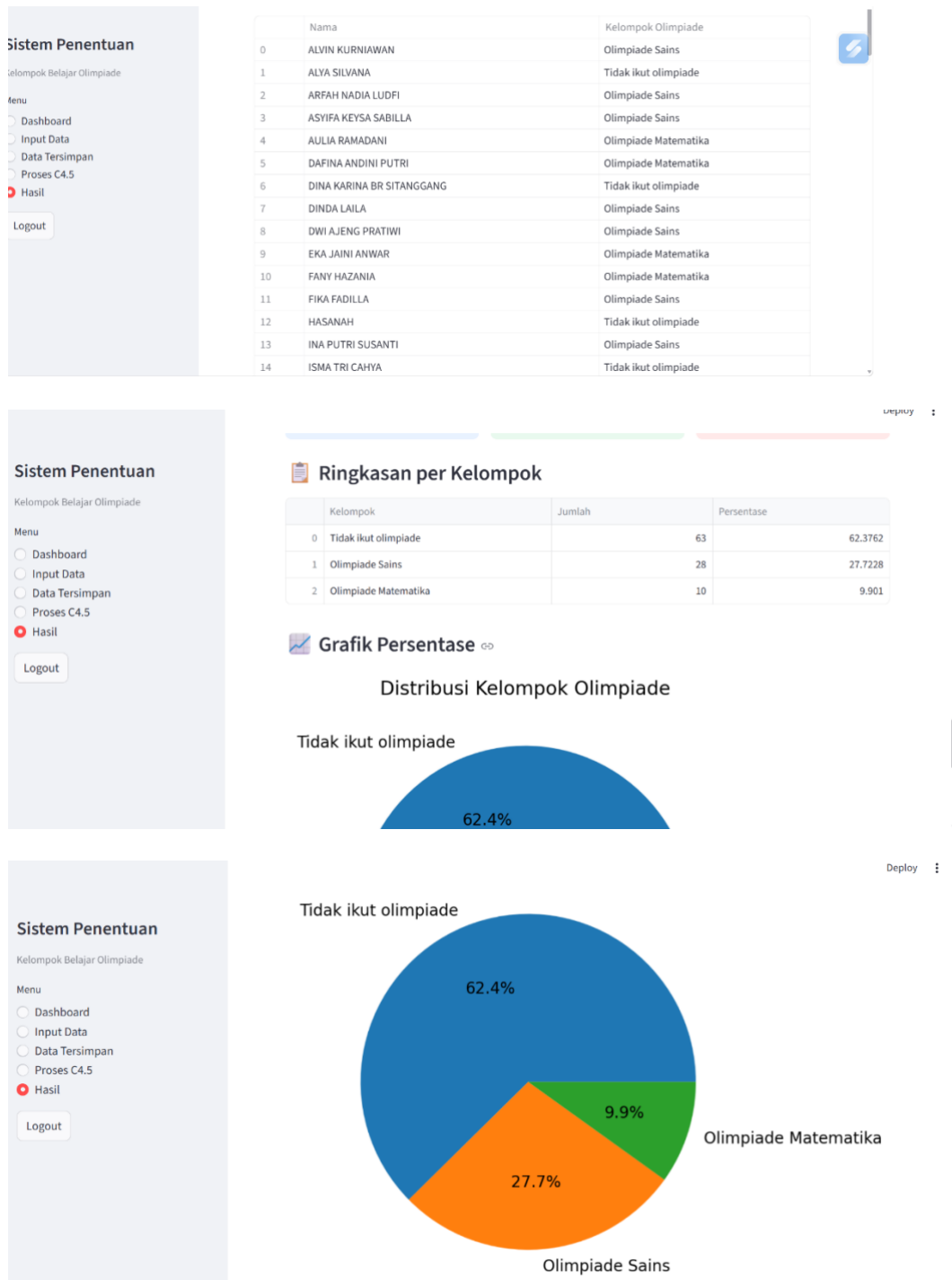
**Gambar 4. 8 Halaman Proses Algoritma**

Halaman diatas merupakan bagian utama dalam sistem yang digunakan untuk menjalankan proses pengolahan data siswa menggunakan algoritma C4.5. Pengguna dapat memulai proses dengan menekan tombol “Proses Algoritma”, kemudian sistem akan menampilkan notifikasi bahwa proses telah selesai.

Pada halaman ini ditampilkan tahapan yang dilakukan, yaitu diskretisasi data dan klasifikasi berdasarkan *rule*. Diskretisasi digunakan untuk mengubah data numerik menjadi kategori (rendah, sedang, tinggi), sedangkan proses klasifikasi dilakukan berdasarkan aturan yang dihasilkan dari algoritma C4.5 untuk menentukan kelompok olimpiade siswa.

Selain itu, tersedia tombol “Lihat Hasil Kelompok” untuk menampilkan hasil akhir klasifikasi. Halaman ini membantu pengguna memahami bahwa proses penentuan dilakukan secara sistematis dan berbasis data.

### 4.3.6 Halaman Hasil Kelompok



**Gambar 4. 9 Halaman Hasil Kelompok**

Halaman hasil kelompok merupakan tahap akhir dari proses penerapan algoritma C4.5 pada sistem. Halaman ini menampilkan hasil klasifikasi siswa ke

dalam kelompok belajar olimpiade berdasarkan model pohon keputusan yang telah terbentuk sebelumnya.

Pada bagian atas halaman ditampilkan notifikasi bahwa proses algoritma telah berhasil dijalankan. Selanjutnya, sistem menampilkan tabel hasil yang berisi nama siswa dan kelompok olimpiade yang direkomendasikan, seperti Olimpiade Matematika, Olimpiade Sains, maupun tidak mengikuti olimpiade. Selain itu, sistem juga menyajikan laporan statistik yang mencakup jumlah total siswa, jumlah siswa yang mengikuti dan tidak mengikuti olimpiade beserta persentasenya, serta ringkasan data dalam bentuk tabel dan visualisasi grafik untuk memudahkan analisis hasil. Hasil yang ditampilkan merupakan output dari proses klasifikasi berdasarkan atribut-atribut yang telah dianalisis, yaitu nilai akademik (MU dan TKDJ), tingkat pemahaman, perilaku, serta nilai rata-rata. Setiap siswa akan masuk ke dalam kelompok yang sesuai berdasarkan aturan (*rule*) yang dihasilkan dari algoritma C4.5.

```
def klasifikasi(row):
    pemahaman = row["Pemahaman"].strip().lower()
    MU = row["MU"].strip().lower()
    nilai = row["Nilai"].strip().lower()

    # RULE 1
    if pemahaman == "cukup paham":
        return "Tidak ikut olimpiade"

    # Pemahaman = Paham
    elif pemahaman == "paham":
        if nilai == "tinggi":
            if MU == "tinggi":
                return "Olimpiade Matematika"
            elif MU == "sedang":
                return "Olimpiade Sains"
            else:
                return "Tidak ikut olimpiade"

        elif nilai == "sedang":
            if MU == "tinggi":
                return "Olimpiade Sains"
            else:
                return "Tidak ikut olimpiade"
```

**Gambar 4. 10 Code Rule Pengelompokkan**

Selain itu, sistem ini juga memiliki opsi “Download Hasil” yang memberi kesempatan kepada pengguna, khususnya staff sekolah, untuk mengunduh hasil klasifikasi dalam bentuk file. Fitur ini memudahkan proses dokumentasi dan pengambilan keputusan lebih lanjut oleh pihak sekolah.

## 4.2 Hasil Pengujian Sistem

Pada tahap ini dilakukan pengujian terhadap sistem yang telah dibangun untuk memastikan bahwa seluruh fitur dan fungsi dapat berjalan dengan baik sesuai dengan kebutuhan pengguna. Pengujian dilakukan menggunakan metode Blackbox Testing serta evaluasi terhadap model yang dihasilkan oleh algoritma C4.5.

### 4.2.1 Hasil Pengujian Blackbox Testing

Pengujian Blackbox Testing dilakukan untuk menilai fungsi-fungsi sistem tanpa memeriksa kode program. Proses pengujian ini berorientasi pada kesesuaian antara input dan output yang dihasilkan oleh sistem. Berikut adalah hasil dari pengujian sistem:

**Tabel 4. 2 Tabel Blackbox Testing**

No	Fitur yang Diuji	Skenario Pengujian	Input	Output yang Diharapkan	Hasil
1	Login	Mengisi username dan password yang benar	Nama & Password valid	Sistem menampilkan halaman dashboard	Berhasil

2	Login	Mengisi username atau password salah	Nama / Password salah	Sistem menampilkan pesan error	Berhasil
3	Input Data Manual	Mengisi seluruh data siswa dengan benar	Data lengkap	Data berhasil disimpan	Berhasil
4	Input Data Manual	Mengosongkan salah satu field	Data tidak lengkap	Sistem menampilkan peringatan	Berhasil
5	Upload CSV	Upload file CSV sesuai format	File .csv valid	Data berhasil diupload dan disimpan	Berhasil
6	Upload CSV	Upload file bukan CSV	File selain .csv	Sistem menampilkan pesan error	Berhasil
7	Data Tersimpan	Menampilkan data siswa	-	Data tampil pada tabel	Berhasil
8	Hapus Data	Menghapus data siswa	Klik tombol hapus	Data terhapus dari sistem	Berhasil
9	Proses C4.5	Menjalankan proses algoritma	Klik proses	Muncul notifikasi	Berhasil

				proses selesai	
10	Diskretisasi Data	Mengubah data numerik menjadi kategori	Nilai angka	Data berubah menjadi kategori	Berhasil
11	Hasil Klasifikasi	Menampilkan hasil kelompok	Data sudah diproses	Tampil kelompok olimpiade siswa	Berhasil
12	Download Hasil	Mengunduh hasil klasifikasi	Klik download	File berhasil diunduh	Berhasil
13	Logout	Keluar dari sistem	Klik logout	Kembali ke halaman login	Berhasil

#### 4.2.2 Evaluasi Model Sistem

Evaluasi model dilakukan untuk menilai kinerja algoritma C4.5 dalam mengklasifikasikan siswa ke dalam kelompok belajar olimpiade. Model yang dihasilkan berupa pohon keputusan yang dibentuk berdasarkan nilai entropy dan information gain dari setiap atribut.

Berdasarkan hasil perhitungan, atribut yang memiliki nilai information gain tertinggi dipilih sebagai akar (root) dalam pohon keputusan. Dalam penelitian ini, atribut yang dominan dalam menentukan klasifikasi adalah nilai MU dan TKDJ, yang kemudian diikuti oleh atribut lain seperti pemahaman, perilaku, dan nilai rata-rata.

Model yang dihasilkan mampu mengelompokkan siswa ke dalam tiga kategori utama, yaitu:

- a. Olimpiade Sains
- b. Olimpiade Matematika
- c. Tidak Mengikuti Olimpiade

Selain itu, aturan (*rules*) yang terbentuk dari pohon keputusan menunjukkan bahwa siswa dengan nilai tinggi pada atribut akademik cenderung masuk ke kelompok olimpiade. Siswa dengan nilai sedang atau rendah cenderung tidak mengikuti olimpiade. Kombinasi antara nilai, pemahaman, dan perilaku juga mempengaruhi hasil klasifikasi. Dengan demikian, model yang dihasilkan oleh algoritma C4.5 dinilai mampu memberikan keputusan yang logis, konsisten, dan sesuai dengan kondisi data yang digunakan.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai implementasi algoritma C4.5 dalam penentuan kelompok belajar olimpiade siswa berbasis web, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Proses pengolahan data dilakukan melalui beberapa tahapan, yaitu diskretisasi data, perhitungan entropy, serta perhitungan information gain untuk menentukan atribut terbaik sebagai akar pohon keputusan. Dengan adanya proses tersebut, sistem mampu mengelompokkan siswa ke dalam kategori Olimpiade Sains, Olimpiade Matematika, dan tidak mengikuti olimpiade secara sistematis dan berbasis data.
2. Berdasarkan hasil perhitungan dan pembentukan pohon keputusan, atribut yang memiliki pengaruh paling besar dalam proses klasifikasi adalah nilai akademik seperti MU (nilai matematika) dan TKDJ nilai komputer sains, yang kemudian didukung oleh atribut lainnya seperti tingkat pemahaman, perilaku, dan nilai rata-rata raport. Adapun *gain* terbesar yaitu kategori pemahaman sebesar 0,527, lalu disusul nilai *gain* kategori rata-rata raport sebesar 0,424, *gain* kategori matematika sebesar 0,320, gain kategori sains komputer sebesar 0,247, dan *gain* kategori perilaku sebesar 0,209.
3. Kombinasi antara kemampuan akademik dan aspek non-akademik memiliki peran penting dalam menentukan kelayakan siswa untuk mengikuti program olimpiade. Dari data yang telah diproses diketahui bahwa dari 101 data yang diproses terdapat pengelompokkan olimpiade

4. yaitu, siswa mengikuti kelas 10 olimpiade matematika, 28 siswa mengikuti kelas olimpiade sains, dan 63 siswa tidak mengikuti kelas olimpiade.

## **5.2 Saran**

Adapun saran yang dapat diberikan untuk pengembangan sistem pada penelitian selanjutnya adalah sebagai berikut:

1. Sistem dapat dikembangkan dengan menambahkan lebih banyak atribut atau variabel yang relevan, seperti minat siswa, prestasi sebelumnya, atau hasil tes khusus, sehingga hasil klasifikasi yang dihasilkan menjadi lebih akurat dan representatif terhadap kondisi sebenarnya.
2. Perlu dilakukan evaluasi model secara kuantitatif dengan menggunakan metode pengukuran akurasi seperti confusion matrix, precision, recall, dan accuracy, agar kinerja algoritma C4.5 dapat dianalisis secara lebih mendalam dan terukur.
3. Pengembangan sistem selanjutnya dapat diarahkan pada peningkatan tampilan antarmuka (user interface) agar lebih menarik, interaktif, dan mudah digunakan oleh berbagai kalangan pengguna, khususnya staff sekolah.
4. Sistem juga dapat dikembangkan menjadi platform yang terintegrasi dengan sistem akademik sekolah, sehingga proses input data dapat dilakukan secara otomatis tanpa harus dilakukan secara manual, serta meningkatkan efisiensi dalam pengelolaan data.



## DAFTAR PUSTAKA

- Afandi, E., Syahputra, T., Boy, A. F., Setiawan, D., Amrullah, A., Rizky, F., & Marnoko. (2023). Pembuatan dan Pendampingan Sistem Informasi Portal Desa Tadukan Raga. *Abdi Sabha (Jurnal Pengabdian Kepada Masyarakat)*, 4(1), 40–47.
- Algoritma, K., Tree, D., & Naive, A. (2018). *Klasifikasi Algoritma Decision Tree, Algoritma Naive*.
- Amrullah, & Syahril, M. (2018). Sistem Pakar Mendiagnosa Kerusakan Dinamo Start Mobil Menggunakan Metode Teorema Bayes Pada Lubis Dinamo Medan. *J-Sisko Tech*, 1(2), 1–13. <https://ojs.trigunadharma.ac.id/index.php/jsk/article/download/26/11>
- Budiarto, E., Rino, R., Hariyanto, S., & Susilawati, D. (2022). Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4.5. *Algor*, 3(2), 23–34. <https://doi.org/10.31253/algor.v3i2.1019>
- C, M. A. (2024). *SWADHARMA (JRIS)*. 1–10.
- Dwika Asrani, Telaumbanua, D. M., Aldi Chandra Maulana, & Rima Tamara Aldisa. (2024). Penerapan Metode WP dan ROC dalam Pemilihan Siswa Peserta Olimpiade Sains. *ADA Journal of Information System Research*, 1(2), 53–58. <https://doi.org/10.64366/adajisr.v1i2.33>
- For, M. (2008). *N Utrition a Nd T He*. 24(3), 171–179.
- Hendry. (2021). Data mining Prediksi Data Tingkat Malas Siswa Di Sekolah SMA. *Aplikasi Dan Analisis Literatur Fasilkom UI*, m(1998), 7–34. <http://elib.unikom.ac.id/files/disk1/655/jbptunikompp-gdl-supriadini-32740-6-12.unik-i.pdf>
- Informatika, T., Teknik, F., Nusantara, U., & Kediri, P. (2022). *Berbasis Web Di Smk Intensif Baitussalam*. 221–225.
- Saputri, I., Saputra, Y. H., Arifin, M., Karim, S. M., & Azikin, N. (2025). PENERAPAN SISTEM INFORMASI MANAJEMEN DATA ( SIMDATA ) BERBASIS WEB UNTUK MENINGKATKAN PELAYANAN ADMINISTRASI WARGA RT 004 RW 003 DESA PULAU JAMBU , KECAMATAN CERENTI ,. 1(2), 40–46.
- Laudon, K. C., & Laudon, J. P. (n.d.). *Sistemas de información gerencial*.
- Laudon, K. C., & Laudon, J. P. (n.d.). *Sistemas de información gerencial*.
- Mahardika, D. D. K. (2024). *No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title (Issue February)*.
- Munandar, S. A., Fitriani, A., Karlina, Y., Yumriani, & BP, A. R. (2022). Pengertian pendidikan, ilmu pendidikan, dan unsur-unsur pendidikan. *Al Urwatul Wutsqa: Kajian Pendidikan Islam*, 2(1), 1–8. [https://www.academia.edu/download/91021639/7757\\_24249\\_1\\_PB\\_2\\_.pdf](https://www.academia.edu/download/91021639/7757_24249_1_PB_2_.pdf)
- Putra, I. C. O. (2024). *Penerapan Algoritma C4. 5 Dalam Data Mining Untuk Identifikasi Faktor Risiko Stroke Pada Dataset Medis*. <https://repositori.buddhidharma.ac.id/2475/>
- Riset, U. (n.d.). *ANALISIS DATA*.
- Rongcai, R. E. N., Guoxiong, W. U., & Ming, C. A. I. (n.d.). *No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title*.

- Saputri, I., Saputra, Y. H., Arifin, M., Karim, S. M., & Azikin, N. (2025). *PENERAPAN SISTEM INFORMASI MANAJEMEN DATA ( SIMDATA ) BERBASIS WEB UNTUK MENINGKATKAN PELAYANAN ADMINISTRASI WARGA RT 004 RW 003 DESA PULAU JAMBU , KECAMATAN CERENTI ,. 1(2), 40–46.*
- Septiani, D., & Dahria, M. (2021). Sistem Pendukung Keputusan Untuk Menentukan Kelayakan Siswa Yang Akan Mengikuti Olimpiade Matematika Nasional Pada SMA Angkasa 1 Lanud Soewondo Medan Dengan Metode Additive Ratio Assessment ( ARAS ). *Jurnal CyberTech*, 4(7), 1–14.
- Syahdan, S. Al, & Sindar, A. (2018). Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 1(2). <https://doi.org/10.32672/jnkti.v1i2.771>
- Wayahdi, M. R., & Ruziq, F. (2023). Pemodelan Sistem Penerimaan Anggota Baru dengan Unified Modeling Language (UML) (Studi Kasus: Programmer Association of Battuta). *Jurnal Minfo Polgan*, 12(1), 1514–1521. <https://doi.org/10.33395/jmp.v12i1.12870>
- Wiyono, T. S., Saputra, R., & Sarwoko, E. A. (2012). Pengembangan Sistem Informasi Pendaftaran. *Journal of Informatics and Tecnology*, 1(1), 1–20.
- Zega, S. A. (2014). Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiwa Berdasarkan Jalur Masuk Kuliah. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, 7–13.

# **LAMPIRAN**



# UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

## FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BAN-PT/IAK/Pg/PT/11/2024

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

[www.umsu.ac.id](http://www.umsu.ac.id)

[ds@umsu.ac.id](mailto:ds@umsu.ac.id)

[umsumedan](#)

[umsumedan](#)

[umsumedan](#)

[umsumedan](#)

### FORMULIR PERBAIKAN UJIAN SKRIPSI

Pada hari ini, Sabtu 18 April 2026 telah dilaksanakan Ujian Skripsi bagi mahasiswa Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Sbb:

Nama Mahasiswa : Diajeng Vesha Ramadhini  
 NPM : 2209010046  
 Program Studi : Sistem Informasi  
 Judul Proposal : Penerapan Algoritma C4.5 Pada Smk Muhammadiyah 16 Desa Pon Untuk Menentukan Kelompok Belajar Olimpiade

Materi/Point yang Diperbaiki :

		Paraf
Amrullah, S.Kom., M.Kom		
Mahardika Prawira Tanjung, S.Kom.,M.Kom.	Buat laporan pada hasil akhirnya, Cari landasan Indikator & yang menjadi acuan untuk penentuan kelompok besar olimpiade.	
Dr. Marah Dolly Nst, M.Si		

Berita acara ini ditandatangani setelah skripsi diperbaiki sesuai petunjuk/arahan dari Pembimbing dan Penguji/Pembahas.

# PENERAPAN ALGORITMA C4.5 PADA SMK MUHAMMADIYAH 16 DESA PON UNTUK MENENTUKAN KELOMPOK BELAJAR OLIMPIADE

## ORIGINALITY REPORT

<b>24%</b> SIMILARITY INDEX	<b>22%</b> INTERNET SOURCES	<b>13%</b> PUBLICATIONS	<b>12%</b> STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

## PRIMARY SOURCES

<b>1</b>	<b>journal.thamrin.ac.id</b> Internet Source	<b>3%</b>
<b>2</b>	<b>repository.umsu.ac.id</b> Internet Source	<b>1%</b>
<b>3</b>	<b>sis.binus.ac.id</b> Internet Source	<b>1%</b>
<b>4</b>	<b>repository.polman-babel.ac.id</b> Internet Source	<b>1%</b>
<b>5</b>	<b>repositori.unsil.ac.id</b> Internet Source	<b>1%</b>
<b>6</b>	<b>mafiadoc.com</b> Internet Source	<b>1%</b>
<b>7</b>	<b>123dok.com</b> Internet Source	<b>1%</b>
<b>8</b>	<b>docplayer.info</b> Internet Source	<b>1%</b>
<b>9</b>	<b>repository.upp.ac.id</b> Internet Source	<b>1%</b>
<b>10</b>	<b>ejurnal.swadharma.ac.id</b> Internet Source	<b>1%</b>
<b>11</b>	<b>www.kaskus.co.id</b>	

## LAMPIRAN LOA JURNAL

**KOHESI**  
JURNAL SAINS DAN TEKNOLOGI

### CAHAYA ILMU BANGSA INSTITUTE

Biro Penelitian, Publikasi, dan Pengabdian Kepada Masyarakat  
KEMENKUMHAM AHU-0018912-AH.0114  
Perum Puri Kartika Asri Blok 2 A2 Malang  
e-mail: [admin@cahayailmubangsa.institute](mailto:admin@cahayailmubangsa.institute)



# LETTER OF ACCEPTANCE

NO: 2026/CIB013/LOA613

### *Assalamualaikum Wr. Wb.*

Bersama surat ini, kami menerangkan bahwa artikel dengan keterangan naskah berikut

Judul	PENERAPAN ALGORITMA C4.5 PADA SMK MUHAMMADIYAH 16 DESA PON UNTUK MENENTUKAN KELOMPOK BELAJAR OLIMPIADE
Author	Diajeng Vesha Ramadhini, Amrullah
Instansi	Universitas Muhammadiyah Sumatera Utara
Korespondensi	<a href="mailto:diajengvesharamadhini@gmail.com">diajengvesharamadhini@gmail.com</a>
Jurnal	Kohesi: Jurnal Sains dan Teknologi
Volume	Vol. 10 No. 12 (2026)
Link Terbitan	<a href="https://cibangsa.com/index.php/kohesi/article/view/9764">https://cibangsa.com/index.php/kohesi/article/view/9764</a>

Berstatus ACCEPTED untuk dipublish. Keputusan ini dibuat sebagai tanda bahwa naskah yang bersangkutan telah lolos plagiarism checker. Dan LoA ini dibuat sebagai bukti bahwa author telah menyelesaikan APC yang telah ditetapkan oleh pengelola jurnal. LOA Berlaku jika dilengkapi link dan pdf publish. Hubungi kami di [admin\\_jurnal@cahayailmubangsa.institute](mailto:admin_jurnal@cahayailmubangsa.institute) jika ada pertanyaan lebih lanjut, terima kasih.

Malang, 11 May 2026



Dr. Umam Rofiq, M.Pd., Ph.D  
Director



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

**UMSU**

Unggul | Cerdas | Terpercaya

Estimasi sumbu di atas ditunjukkan nama dan tanggalnya

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019  
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fkip.umsu.ac.id>

[fkip@umsu.ac.id](mailto:fkip@umsu.ac.id)

[umsu.medan](#)

[umsu.medan](#)

[umsu.medan](#)

[umsu.medan](#)

**Berita Acara Pembimbingan Proposal**

Nama Mahasiswa : Diajeng Veshs Ramadhini Program Studi : Sistem Informasi  
 NPM : 2209010046 Judul Penelitian : PENERAPAN ALGORITMA C4.5 PADA SMK MUHAMMADIYAH 16 DESA PON UNTUK MENENTUKAN KELOMPOK BELAJAR OLIMPIADE

Nama Dosen Pembimbing : Amrullah S.Kom., M.Kom

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
23/2/26	Perbaiki hasil seminar proposal sesuai dengan arahan dosen pembimbing	
27/2/26	Bab 4 sesuaikan tampilan bab 3 yang berbeda	
9/3/26	Sesuaikan isi data dan laporan pada aplikasi dan ss di bab 4	
13/3/26	Perbaiki bahasa di bab 4. Lanjut bab 5	
31/3/26	Bab 4 acc. Perbaiki point Kesimpulan dan saran	
07/4/26	Acc bab dan exb. Lengkapi berkas persyaratan sidang	
09/04/26	Acc daftar siding skripsi	

Medan, 09 April 2026

Diketahui oleh :

Ketua Program Studi

(Mardika Abdi

Prawira Tanjung S.Kom.,M.kom)

Disetujui oleh :

Dosen Pembimbing

(Amrullah S.Kom.,M.Kom)

