

## ***Implementation Of Data Mining In Determining Promotional Strategies At Al-Amah Vocational School Cimanggung Using C4.5 Algorithm***

### **Implementasi Data Mining Untuk Menentukan Strategi Promosi Di Smk Al-Amah Cimanggung Menggunakan Algoritma C4.5**

Sofia Dewi<sup>1</sup>, Ai Kresnawati<sup>2</sup>, Syifa Hanifa Salsabil<sup>3</sup>

<sup>1,2,3</sup> Sistem Informasi, Universitas Ma'soem, Indonesia

E-Mail: <sup>1</sup>sofiadewi@masoemuniversity.ac.id, <sup>2</sup>kresnawatia@gmail.com,

<sup>3</sup>syifa.hanifa@masoemuniversity.ac.id

Makalah: Diterima 13 Desember 2024; Diperbaiki 19 Januari 2025; Disetujui 23 Maret 2025

Corresponding Author: Sofia Dewi

#### **Abstrak**

Dalam pengelolaan data dapat dilakukan penggalian informasi menggunakan *data mining*. *Data mining* dapat membantu pihak sekolah dalam pengambilan keputusan, yang dimana diharapkan dalam mengelola data mendapatkan informasi yang terkandung didalamnya menjadi sebuah pengetahuan baru. Tujuan dari penelitian ini yaitu untuk melakukan proses pengelolaan data siswa dengan metode klasifikasi dalam upaya untuk mendukung membuat keputusan dalam menentukan strategi promosi di SMK Al-Amah Cimanggung dengan metode C4.5. Analisis dan implementasi data menggunakan algoritma C4.5 dilakukan pengujian menggunakan *RapidMiner Studio*. Data yang digunakan dalam penelitian ini berjumlah 363 data yang terdiri dari data *training* 253 dan data *testing* 108. Atribut yang digunakan dalam penelitian ini diantaranya Asal sekolah, alamat, informasi PPDB, nilai akademik, dan pendapatan orang tua. Berdasarkan hasil implementasi algoritma C4.5, atribut dengan nilai gain tertinggi adalah atribut sekolah asal merupakan atribut yang paling berpengaruh dalam penelitian ini kemudian di tahap selanjutnya atribut yang memengaruhi adalah pendapatan orang tua yang berasal dari asal sekolah negeri dan informasi PPDB dari asal sekolah swasta. Data mining dengan algoritma C4.5 dapat digunakan untuk menentukan strategi promosi sekolah dengan menghasilkan *decision Tree* dengan hasil *accuracy* 89,81. Maka menunjukkan bahwa penelitian memiliki tingkat akurasi yang cukup tinggi.

**Kata Kunci** : Data Mining, C4.5, Klasifikasi, Promosi, *RapidMiner*

#### **Abstract**

*In data management, information can be extracted using data mining. Data mining can help schools in decision making, which is expected in managing data to get the information contained in it into new knowledge. The purpose of this research is to carry out the process of managing student data with classification methods in an effort to support making decisions in determining promotional strategies at SMK Al-Amah Cimanggung using the C4.5 method. Data analysis and implementation using the C4.5 algorithm are tested using RapidMiner Studio. The data used in this study amounted to 363 data consisting of 253 training data and 108 testing data. The attributes used in this study include school origin, address, PPDB information, academic scores, and parents' income. Based on the results of the implementation of the C4.5 algorithm, the attribute with the highest gain value is the school of origin attribute is the most influential attribute in this study then in the next stage the attribute that influences is the income of parents who come from public school origin and PPDB information from private school origin. Data mining with the C4.5 algorithm can be used to determine school promotion strategies by producing decision trees with an accuracy of 89.81. So it shows that the research has a fairly high level of accuracy.*

**Keywords** : Data mining, C4.5, Classifications, Promotion, *RapidMiner*.

#### **1. PENDAHULUAN**

Kemajuan teknologi informasi berkembang pesat di segala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang kompleks, mulai dari bidang industri, ekonomi, ilmu pengetahuan

dan masih banyak bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data berlimpah mengenai siswa dan proses pembelajarannya. Pada pendidikan Sekolah Menengah Atas atau Sekolah menengah kejuruan, data dapat diperoleh berdasarkan data historis, sehingga datanya terus bertambah seperti data siswa. Seperti proses penerimaan siswa baru di suatu sekolah menghasilkan data yang berlimpah berupa catatan siswa baru.[1] Data tersebut dapat diolah menggunakan sebuah teknologi, salah satunya dengan teknik data mining.[2]

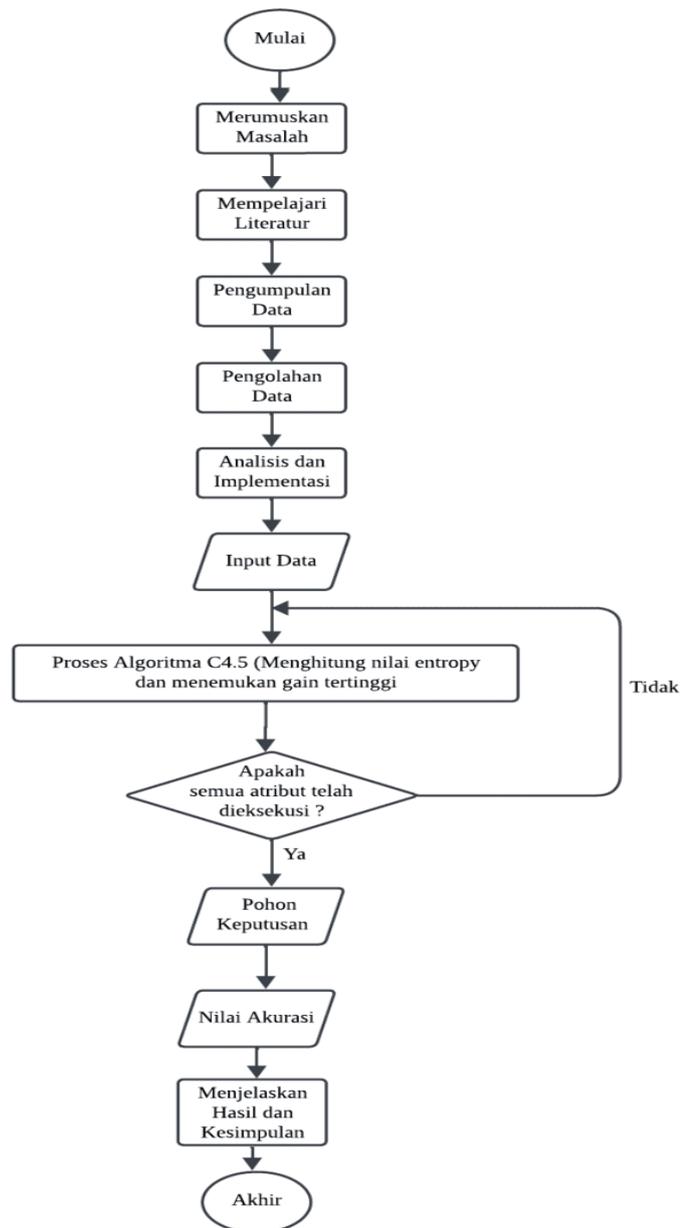
Diharapkan bahwa data mining akan membantu sekolah dalam pengambilan keputusan dan mengubah informasi yang terkandung dalam data menjadi pengetahuan baru. Data mining merupakan bagian dari proses KDD (Penemuan Pengetahuan dalam Basis Data) dan mencakup beberapa tahapan seperti pemilihan data, preprocessing, transformasi, penambahan data, dan evaluasi hasil. [3] Dalam penelitian ini digunakan algoritma C.45 yang memiliki kemampuan untuk menghasilkan pohon keputusan yang sangat baik.[4] Mampu menghilangkan perhitungan atau data yang mungkin tidak diperlukan adalah kelebihan lain dari teknik ini. Karena sampel pada penelitian ini biasanya hanya diuji menggunakan kelas atau standar tertentu. Meskipun metode ini memiliki banyak kelebihan, itu tidak berarti tidak memiliki kekurangan. Mungkin terjadinya tumpang tindih dalam hasil pohon keputusan.[5]

SMK Al-Amah Cimanggung merupakan sekolah kejuruan yang beralamatkan di Jl. Parakan Muncang-Sindulang Km.02 Desa Cimanggung Kec. Cimanggung Kab.Sumedang dibawah naungan Yayasan AL AMAH yang pada tahun 2006 diresmikan, di SMK Al-Amah Cimanggung memiliki 4 jurusan, yaitu : jurusan Bisnis Daring dan Pemasaran (BDP), Akuntansi dan Keuangan Lembaga, Teknik Kendaraan Ringan (TKR), dan Teknik Komputer Jaringan (TKJ). Data yang dikumpulkan dari pendaftaran ini sangat besar, jadi dapat digunakan dan diolah untuk mendapatkan informasi yang bermanfaat bagi sekolah untuk digunakan saat membuat rencana untuk meningkatkan strategi promosi penerimaan siswa baru pada tahun berikutnya. Berdasarkan data yang diperoleh dari bagian Staff Tata Usaha SMK Al-Amah Cimanggung data siswa PPDB tahun 2021, 2022 dan 2023.

Pada kegiatan promosi yang dilakukan SMK Al-Amah Cimanggung yaitu dengan cara online marketing seperti penggunaan Instagram lalu ada kegiatan offline yaitu melalui pemasangan spanduk, penyebaran brosur dan kunjungan ke sekolah dengan agenda melakukan demo di sekolah tingkat SMP untuk memperkenalkan jurusan yang ada. Salah satu penyebab terjadinya penurunan jumlah siswa baru pada PPDB tahun 2021, 2022 dan 2023 adalah kurangnya dilakukan pengolahan data siswa secara tepat berdasarkan data historis di SMK Al-Amah Cimanggung, padahal pengolahan data siswa dapat menjadi aset yang berharga dengan mengelola data yang tepat sehingga dapat membantu dalam menentukan promosi sekolah yang tepat sasaran sehingga tidak terjadi penurunan jumlah siswa pada tahun berikutnya.[6] Adapun indikator variabel dalam penelitian ini antara lain: Asal Sekolah, Alamat, Informasi PPDB, Nilai Akademik dan Pendapatan Orang Tua. Penelitian ini untuk merekomendasikan strategi promosi yang sesuai dengan metode C4.5. Dengan menggunakan metode klasifikasi yang menggunakan algoritma C4.5, penelitian ini dapat membantu sekolah melakukan strategi promosi yang tepat untuk menarik siswa baru setiap tahunnya.

## **2. METODE DAN BAHAN**

Pada tahapan penelitian ini, prosedur kerja dilakukan seperti yang ditunjukkan dalam gambar berikut:



**Gambar 1.** Prosedur Kerja

Sumber : Dewi & Oktaviawati, 2022[7]

## 2.1 Strategi Promosi

Strategi promosi adalah suatu kegiatan terencana yang menggunakan berbagai variabel promosi sebagai alat yang bertujuan untuk menginformasikan, membujuk, dan mengingatkan konsumen.[8] Dalam pelaksanaan strategi promosi sangatlah berpengaruh untuk berlangsungnya sebuah bisnis ataupun sebuah organisasi untuk menarik perhatian individu atau kelompok. Strategi promosi adalah rencana untuk penggunaan yang optimal dari elemen-elemen promosi: periklanan, hubungan masyarakat, penjualan pribadi, dan promosi penjualan.[9]

## 2.2 Data Mining

*Data mining* adalah proses menemukan pengetahuan, mengekstrak pengetahuan, menganalisis data atau pola, mengumpulkan informasi, dan menganalisis statistik data. Data mining juga dikenal sebagai *knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain.[10] Penerapan

*data mining* salah satunya adalah pada bidang pendidikan. Di bidang pendidikan memiliki bermacam-macam data dengan jumlah yang besar. Data mining dikenal dengan istilah Knowledge Discovery in Database (KDD) yang berarti mengekstraksi pengetahuan dari database. KDD melibatkan pengumpulan dan penggunaan data historis untuk menemukan pola atau hubungan dalam set data dalam kumpulan data yang lebih besar.[11] Knowledge Discovery In Database (KDD) adalah pendekatan untuk mendapatkan pengetahuan dari database yang memiliki banyak tabel yang saling berhubungan.[12]

### **2.3 Algoritma C4.5**

Dalam membuat *decision Tree* membutuhkan suatu algoritma, algoritma C4.5 adalah salah satu alat bantu yang dapat digunakan untuk mengolah dan menyederhanakan data numerik dan atribut C4.5 adalah algoritma berbasis rasio gen, jadi tahapan diperlukan untuk menghasilkan *decision Tree*. [13]

### **2.4 Metode Klasifikasi**

Klasifikasi adalah proses menemukan pola atau fungsi yang dapat digunakan untuk menjelaskan kelas atau konsep yang ada dalam data.[14] Dalam prakteknya metode dalam data mining terdapat beberapa teknik diantaranya yaitu metode klasifikasi. Klasifikasi adalah salah satu pendekatan pembelajaran yang diawasi yang bertujuan untuk mengidentifikasi hubungan antara atribut masukan dan atribut target. Tujuan dari metode ini adalah untuk meningkatkan keandalan hasil yang diperoleh dari data.[15]

## **3. HASIL DAN PEMBAHASAN**

### **3.1 Pengumpulan Data**

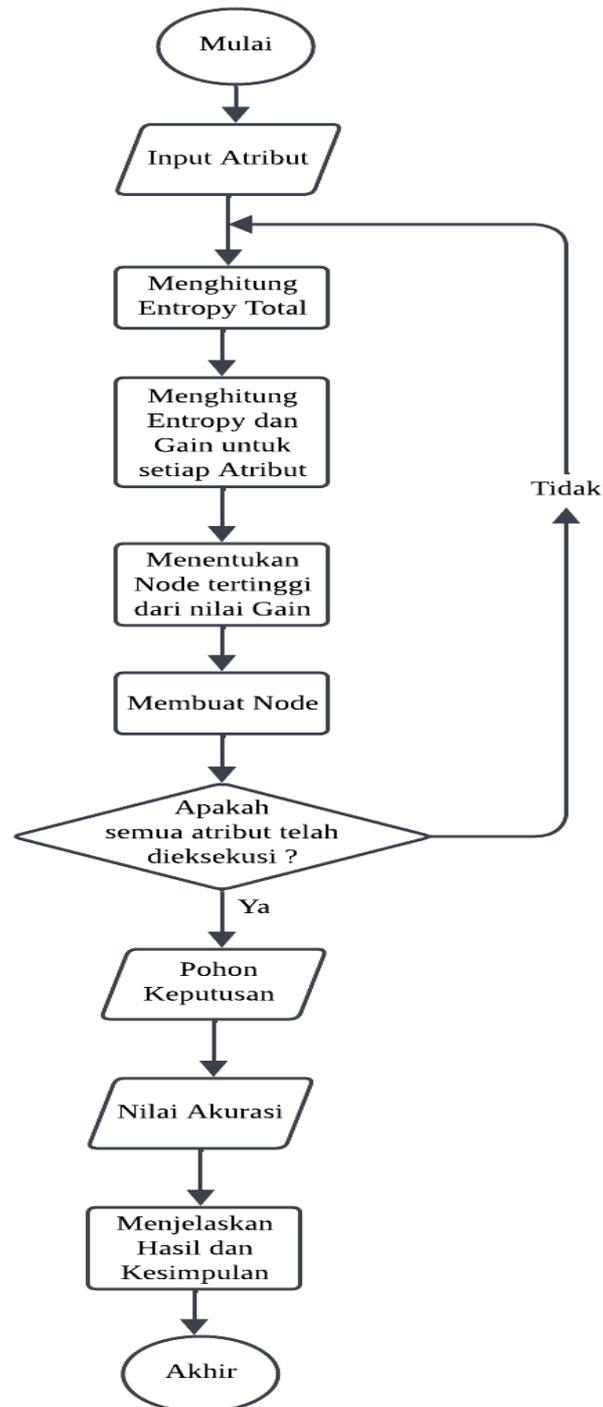
Untuk mendapatkan data-data yang dibutuhkan dalam penelitian ini, maka penulis menggunakan beberapa metode pengumpulan data diantaranya wawancara dan peneliti mengumpulkan data-data yang berhubungan dengan penelitian, yaitu seperti data siswa yang terdaftar pada tahun ajaran 2023-2024 di sekolah, data ini dapat menjadi dokumen yang dapat mendukung proses penelitian. Peneliti menggunakan objek penelitian mengenai strategi promosi sekolah di SMK Al-Amah Cimanggung yang dimana data yang diambil yaitu data pada PPDB tahun 2021, 2022 dan 2023 semua jurusan sebagai data yang dibutuhkan dalam penelitian ini. Sumber data berupa file excel dengan format .xlsx.

### **3.2 Preprocessing Data**

Pemrosesan data adalah langkah di mana data terpilih mengalami berbagai langkah pemrosesan untuk mempersiapkan data yang akan dianalisis. Proses ini meliputi pengecekan duplikasi, pembersihan, dan memperbaiki kesalahan pada penulisan. Fungsi utama pengolahan data adalah untuk memastikan bahwa data tersedia dan dapat diandalkan, serta meminimalkan kesalahan yang mungkin terjadi selama analisis, dalam proses ini data tersisa berjumlah 361 data.

### **3.3 Pemodelan**

Algoritma C4.5 adalah salah satu alat bantu yang dapat digunakan.. Metode ini kemudian dapat mengolah dan menyederhanakan data atribut/numerik. Langkah-langkah yang dilakukan yang dapat dideskripsikan dalam gambar berikut :



**Gambar 2.** Prosedur C4.5

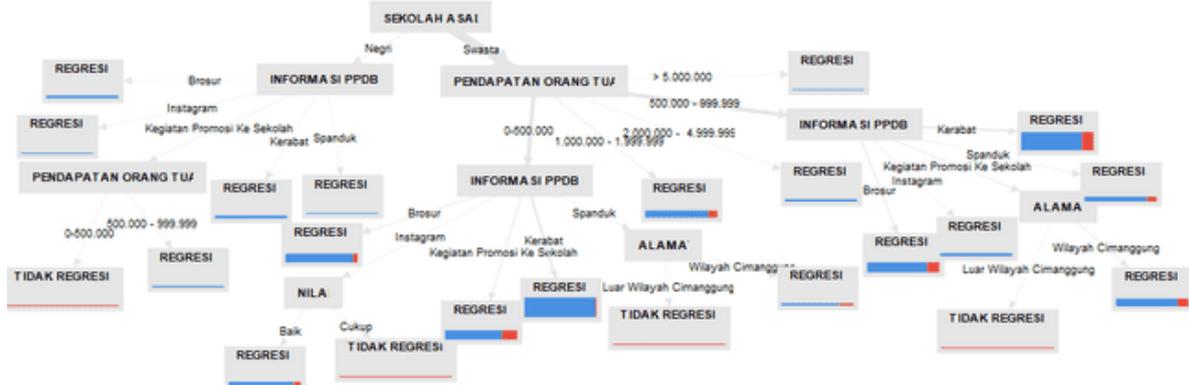
Sumber : Dewi & Oktaviawati, 2022

### 3.4 Implementasi

Berdasarkan penerapan data mining di SMK Al-Amah Cimanggung untuk menentukan strategi promosi sekolah dengan data sebanyak 363 data, dengan melalui beberapa tahapan yaitu *Data Selection*, *Preprocessing*, *Transformation Data Mining* dan *Interpretasi / Evaluasi*. Kemudian data tersebut di olah menggunakan *tools RapidMiner*. *RapidMiner studio* adalah alat penambangan data yang kuat yang memungkinkan pembuatan

model prediktif dengan cepat. Ini adalah alat multifungsi yang mencakup ratusan algoritma pembelajaran mesin dan persiapan data untuk mendukung proyek data mining.[16]

Dalam tahapan penerapan algoritma sendiri yaitu dimana dengan menghitung nilai *entropy* dan nilai *gain* tertinggi, sehingga dalam penelitian ini mendapatkan atribut Asal Sekolah sebagai node akar memiliki nilai *gain* sebesar 0,51072228, dan untuk kategori negeri dan swasta memerlukan atribut lain untuk dapat menentukan hasil kategori siswa dari atribut Asal Sekolah itu sendiri, sehingga dalam proses perhitungan gain diawali oleh node 1.0 dan berakhir di node 1.2.1. Sehingga dalam penelitian ini atribut yang paling berpengaruh yaitu asal sekolah kemudian di susul oleh atribut informasi PPDB dan Pendapatan orang tua sebagai atribut penting dalam menentukan strategi promosi.



Gambar 3. Decision Tree Strategi Promosi

Dari gambar diatas dapat dijelaskan bahwa atribut sekolah asal merupakan atribut yang paling berpengaruh penelitian ini atribut ini menjadi sebuah gain tertinggi dari atribut lainnya kemudian di tahap selanjutnya atribut yang memengaruhi adalah pendapatan orang tua yang berasal dari asal sekolah negeri dan informasi PPDB dari asal sekolah swasta.

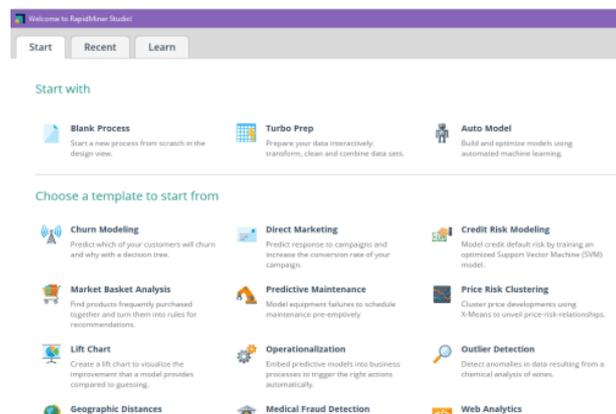
### Validasi Performance

Penelitian ini membutuhkan validasi performance yaitu menggunakan RapidMiner studio yang dimana tools ini untuk mengetahui seberapa akurasi dalam perhitungan C4.5 untuk menentukan strategi promosi. Yang dimana data set telah dibagi menjadi 2 data yaitu data testing dan data training.

### Menguji Performance dari Decision Tree

Setelah menyelesaikan proses implementasi menggunakan *decision Tree* dalam penelitian ini untuk selanjutnya melakukan validasi *performance*. Berikut merupakan implementasi Validasi *performance* menggunakan RapidMiner Studio :

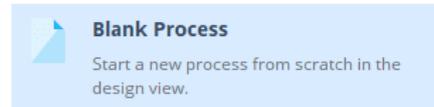
1. Membuka Aplikasi RapidMiner Studio
- 2.



Gambar 4. Halaman RapidMiner Studio

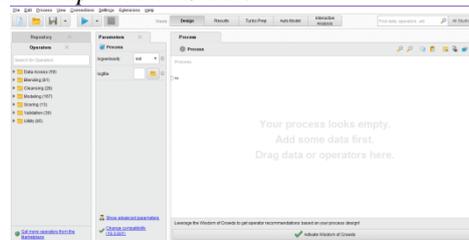
- Pilih Blank Process untuk melakukan proses baru

Start with



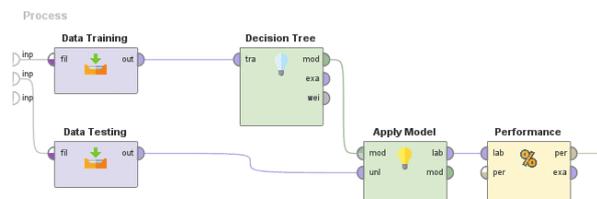
**Gambar 5.** Menu Membuat Proses Baru *RapidMiner* Studio

- Tampilan Halaman utama pada *RapidMiner* Studio



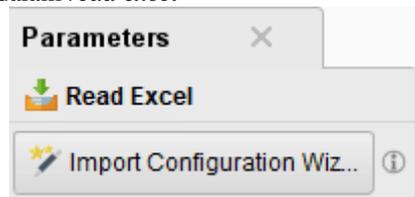
**Gambar 6.** Halaman kerja *RapidMiner* Studio

- Menambahkan *Operator* yang dibutuhkan : *Read Excel*, *Decision Tree*, *Apply model* dan *Performance* (semua *operator* ini berfungsi untuk menghitung *performance* akurasi dalam penelitian ini) dan sambungkan atau gabungkan seluruh *operator* yang ada.



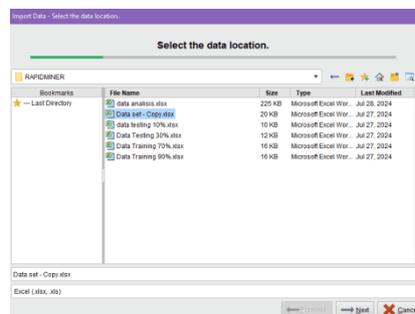
**Gambar 7.** *Operator RapidMiner* Studio

- Menambahkan/ *import file* kedalam *read excel*



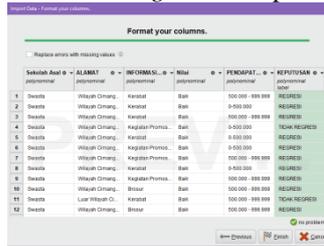
**Gambar 8.** Menu *import RapidMiner* Studio

- Pilih *File Data training* dan *testing* untuk di setiap *import read excel* untuk melakukan implementasi menghitung *performance*.



**Gambar 9.** Tampilan *Import Data RapidMiner* Studio

8. Pilih semua data yang akan digunakan dan *setting* kolom keputusan menjadi label.



Gambar 10. Tampilan Pemilihan Data *RapidMiner Studio*

9. Kemudian Klik *button Run*



Gambar 11. *Button Run/ Jalankan RapidMiner Studio*

10. Hasil Akurasi

Akurasi data *training* dan data *testing* pada *RapidMiner* didapat sebesar 89,81%.

accuracy: 89.81%

|                     | true REGRESI | true TIDAK REGRESI | class precision |
|---------------------|--------------|--------------------|-----------------|
| pred. REGRESI       | 97           | 11                 | 89.81%          |
| pred. TIDAK REGRESI | 0            | 0                  | 0.00%           |
| class recall        | 100.00%      | 0.00%              |                 |

Gambar 12. Hasil Akurasi

Berdasarkan hasil akurasi dalam penelitian diperoleh akurasi sebesar 89,81. Maka menunjukkan bahwa penelitian memiliki tingkat akurasi yang cukup tinggi dalam mengklasifikasikan data, dalam tabel standarisasi akurasi termasuk kedalam akurasi sangat baik.

Tabel 1. Standarisasi Akurasi

|                             |                    |
|-----------------------------|--------------------|
| <i>Accuracy</i> 0,90 - 1,00 | Sangat Baik Sekali |
| <i>Accuracy</i> 0,80 - 0,90 | Sangat Baik        |
| <i>Accuracy</i> 0,70 - 0,80 | Baik               |
| <i>Accuracy</i> 0,60 - 0,70 | Cukup              |
| <i>Accuracy</i> 0,50 - 0,60 | Sedang             |

Sumber : Dewi & Oktaviawati, 2022

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil implementasi Algoritma C4.5 dengan mengklasifikasikan atribut asal sekolah, alamat, informasi PPDB, nilai akademik dan pendapatan orang tua dari hasil menghitung entropy dan gain sehingga menghasilkan nilai gain tertinggi dapat disimpulkan untuk menunjang strategi promosi diantaranya yaitu:
  - a. Atribut yang pengaruh paling tinggi dalam menentukan promosi sekolah ini yaitu asal sekolah dengan kategori swasta memiliki nilai dominan dibandingkan dengan negeri. kemudian di tahap selanjutnya atribut yang memengaruhi adalah pendapatan orang tua yang berasal dari asal sekolah negeri dan informasi PPDB dari asal sekolah swasta. Kemudian atribut alamat menjadi atribut penting dalam keputusan ini dikarenakan siswa yang beralamat di cimanggung lebih dominan regresi dari pada yang diluar cimanggung.
  - b. Interpretasi hasil penelitian ini mengidentifikasi bahwa atribut dan kategori yang perlu di pertimbangkan oleh pihak sekolah yaitu lebih fokus meningkatkan promosi ke sekolah negeri dan ditingkatkan lebih baik lagi dari promosi saat ini bagi sekolah swasta. Selanjutnya promosi sebaiknya

disebarkan melalui media sosial dikarenakan minat dari siswa luar wilayah Cimanggung masih sangat rendah.

2. Hasil akurasi dari penggunaan RapidMiner Studio menggunakan algoritma C4.5 dengan metode klasifikasi dalam menentukan strategi promosi adalah 89,81%. Dengan ini dinyatakan bahwa standarisasi akurasi penelitian ini ada dalam peringkat sangat baik.

## REFERENSI

- [1] A. Y. Sari and E. Supriatna, "Penerapan Data Mining Menggunakan Metode Algoritma Naive Bayes Classifier untuk Mendukung Strategi Promosi," *J. Dimamu*, vol. 3, no. 1, pp. 18–28, 2023, doi: 10.32627/dimamu.v3i1.837.
- [2] F. Santoso, Sunardi, and H. Z. Lukman, "Implementasi Data Mining dengan Metode Naive Bayes Untuk Memprediksi Penerimaan Siswa Baru di MTS NU Islamiyah Asembagus," *G-Tech J. Teknol. Terap.*, vol. 7, no. 4, pp. 1355–1366, 2023, doi: 10.33379/gtech.v7i4.3086.
- [3] I. Ismawati, "Kajian Data Mining Profil Siswa Baru Dalam Penentuan Strategi Promosi Dengan Metode Two Step Clustering," *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 3, pp. 47–59, 2019, doi: 10.33197/jitter.vol5.iss3.2019.302.
- [4] R. Fardiani and Fitriyani, "Penerapan algoritma c4.5 untuk prediksi penerima beasiswa siswa berprestasi di mi al-ishlah ciganitri," vol. 8, no. 4, pp. 402–410, 2024.
- [5] K. Handoko, P. Simanjuntak, E. Hutabri, and E. Erlin, "Penerapan algoritma c4.5 untuk penentuan jurusan siswa sekolah menengah atas," *J. TEKINKOM*, vol. 6, no. 1, pp. 153–157, 2023, doi: 10.37600/tekinkom.v6i1.788.
- [6] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, and W. Gata, "Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree ( Studi Kasus Smk Multicomp Depok )," *Swabumi*, vol. 9, no. 1, pp. 9–18, 2021, doi: 10.31294/swabumi.v9i1.8363.
- [7] S. Dewi and O. Oktaviawati, "Penerapan Algoritma C4.5 untuk Pehamanan Siswa SMK Pada Pelajaran Kompetensi Keahlian," *Intern. (Information Syst. Journal)*, vol. 5, no. 2, pp. 116–125, 2022, doi: 10.32627/internal.v5i2.626.
- [8] K. Martowinangun, D. J. Sri Lestari, and K. Karyadi, "Pengaruh Strategi Promosi Terhadap Peningkatan Penjualan Di Cv. Jaya Perkasa Motor Rancaekkek Kabupaten Bandung," *J. Co Manag.*, vol. 2, no. 1, pp. 139–152, 2019, doi: 10.32670/comanagement.v2i1.162.
- [9] M. L. Hedynata and W. E. D. Radianto, "Strategi Promosi dalam Meningkatkan Penjualan Luscious Chocolate Potato Snack," *Strateg. Promosi*, vol. 1, no. April, pp. 1–10, 2016.
- [10] M. Arhani and M. Nasir, *Data Mining, Algoritma Dan Implementasi*. Andi, 2020.
- [11] M. L. Dalafranka, "Penerapan Algoritma Id3 untuk Mendukung Sosialisasi Penerimaan Mahasiswa Baru yang Tepat Sasaran di UIN Raden Fatah Palembang," *Teknomatika*, vol. 10, no. 2, pp. 141–150, 2020, [Online]. Available: <http://ojs.palcomtech.com/index.php/teknomatika/article/view/501>
- [12] H. Astuti, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Pengelompokkan Data Pelanggan (Studi Kasus : PT. Pinus Merah Abadi)," *J. Web Inform. Teknol.*, vol. 4, no. 1, p. 9, 2020.
- [13] A. Rufiyanto, M. Rochcham, and A. Rohman, *Penerapan Algoritma C45.5 Untuk Prediksi Kepuasan Mahasiswa*. Deepublish, 20202.
- [14] Mulaab, *Data Mining Konsep dan Aplikasi*. Media Nusa Creative, 2017.
- [15] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [16] D. Jollyta, M. Siddik, H. Mawengkang, and S. Efendi, *Teknik Evaluasi Cluster Solusi menggunakan Python dan RapidMiner*. Deepublish, 2021.