



Implementasi Data Mining Prediksi Perminatan Jurusan Siswa pada SMK Negeri 1 Waikabubak dengan Metode Algoritma C4.5

Marten Sudi^{1*}, Gergorius Kopong Pati², Lidia Lali Momo³

¹⁻³Universitas Stella Maris Sumba, Indonesia

Email: martensudi@gmail.com¹, gregkopong80@gmail.com², momo-lidia@yahoo.co.id³

Korespondensi penulis: martensudi@gmail.com*

Abstract. Admission of new students to an educational institution is an activity that is always carried out every new academic year, where prospective new students always increase from year to year (Muwardah and Pramunendar, 2015). Admission of students can be held from elementary to middle school, from middle school to high school / vocational school. The focus of this research is the registration of new students at SMK. As is known, SMK is a Vocational High School or abbreviated as (SMK) and where there are many majors provided which ultimately makes prospective new students confused about which major is right for them because will take a long time.. Based on C4.5 as a Classification Algorithm: C4.5 is a popular algorithm for building decision trees. It works by dividing a dataset into smaller subsets based on attribute values, thus forming an easy-to-understand tree structure. Classification results using decision trees provide a clear visualization of the decision-making process and the variables that contribute to student choices.

Keywords: Data Mining, C4.5 Algorithm Method, Student

Abstrak. Penerimaan siswa baru pada suatu institusi pendidikan merupakan sebuah kegiatan yang selalu dilaksanakan setiap tahun ajaran baru, yang dimana calon siswa baru tersebut selalu meningkat dari tahun ke tahun (Muwardah dan Pramunendar, 2015). Penerimaan siswa ini bisa diadakan dari jenjang SD ke SMP, dari SMP ke SMA / SMK. Fokus penelitian ini adalah pendaftaran siswa baru di SMK.. Seperti yang diketahui bahwa SMK adalah Sekolah Menengah Kejuruan atau di singkat dengan (SMK) dan dimana disediakan banyak sekali jurusan yang akhirnya membuat para calon siswa baru bingung untuk memilih jurusan mana yang tepat untuk mereka karena akan berlangsung lama.. Berdasarkan C4.5 sebagai Algoritma Klasifikasi: C4.5 adalah algoritma yang populer untuk membangun pohon keputusan. Ia bekerja dengan membagi dataset menjadi subset yang lebih kecil berdasarkan nilai atribut, sehingga membentuk struktur pohon yang mudah dipahami. Hasil klasifikasi menggunakan pohon keputusan memberikan visualisasi yang jelas tentang proses pengambilan keputusan dan variabel-variabel yang berkontribusi pada pilihan siswa.

Kata Kunci: Data Mining , Metode Algoritma C4.5, Siswa

1. PENDAHULUAN

Penerimaan siswa baru pada suatu institusi pendidikan merupakan sebuah kegiatan yang selalu dilaksanakan setiap tahun ajaran baru, yang dimana calon siswa baru tersebut selalu meningkat dari tahun ke tahun (Muwardah dan Pramunendar, 2015). Penerimaan siswa ini bisa diadakan dari jenjang SD ke SMP, dari SMP ke SMA / SMK. Fokus penelitian ini adalah pendaftaran siswa baru di SMK.. Seperti yang diketahui bahwa SMK adalah Sekolah Menengah Kejuruan atau di singkat dengan (SMK) dan dimana disediakan banyak sekali jurusan yang akhirnya membuat para calon siswa baru bingung untuk memilih jurusan mana yang tepat untuk mereka karena akan berlangsung lama. Untuk mengatasi masalah tersebut maka peneliti memberi solusi dengan cara menerapkan data mining . Melihat masalah yang dihadapi sekolah dalam menentukan jurusan untuk setiap siswa baru, maka perlu diterapkan

suatu metode untuk menyelesaikan masalah tersebut. Salah satunya metode yang cocok dalam melakukan prediksi peminatan jurusan yang diminati adalah penerapan data mining dengan menggunakan algoritma C4.5. Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai data base besar (Swastina, 2013). Data mining disisi lain adalah kegiatan meliputi pengumpulan, pemakaian data.

Data mining dan penggunaan algoritma C4.5 sebagai salah satu teknik pembelajaran berbasis induksi dalam analisis data sangat tepat. Berikut adalah penjelasan yang lebih mendalam tentang data mining, pembelajaran berbasis induksi, dan algoritma C4.5. Penggunaan data mining dan algoritma C4.5 dalam konteks analisis data dan pengambilan keputusan menunjukkan potensi besar dalam mengekstrak pengetahuan dari data yang kompleks. Pembelajaran berbasis induksi seperti ini dapat memberikan wawasan yang berharga dan mendukung pengambilan keputusan yang lebih baik di berbagai bidang, termasuk pendidikan.

2. METODE PENELITIAN

Metode Pengembangan Sistem

Penting untuk membuat struktur yang jelas agar pelaksanaan penelitian ini lebih mudah. Untuk memastikan bahwa tujuan penelitian selesai sesuai rencana, kerangka penelitian menjabarkan proses yang menunjukkan banyak tahapan kegiatan yang harus dilakukan sepanjang proses tersebut. Kami akan menerapkan fase selanjutnya dari kerangka penelitian :

1 Studi Literatur

Untuk memperkuat konsep dan teori serta memastikan landasan akademis yang kuat dan dapat diterima, kini dilakukan pencarian landasan teori yang diambil dari berbagai buku dan sumber internet.

2 Pengumpulan data

Pengumpulan data yang dilakukan untuk mendapatkan informasi yang diperlukan. Oleh karena itu, penulis menggunakan metode pengumpulan data, termasuk observasi dan wawancara dengan partisipan penelitian terkait.

3 Analisis data

Setelah proses pengumpulan data selesai, informasi yang diperoleh dari berbagai sumber ilmiah yang telah dilakukan sebelumnya dianalisis. Tujuan dari prosedur ini adalah untuk mengklasifikasikan data yang diperlukan untuk pengembangan sistem di masa depan, sehingga memudahkan analisis penulis.

Metode Pengumpulan Data

Sejumlah strategi pengumpulan data digunakan untuk mendapatkan informasi guna keperluan penyusunan skripsi yaitu:

1. Observasi

Untuk mendapatkan informasi yang dapat membantu proses perancangan sistem, diperlukan pengumpulan data observasi langsung di SMK N 1 Waikabubak yang meliputi prosedur pembelian obat, entri data, persediaan obat dan laporan.

2. Wawancara

Wawancara langsung kepada pimpinan dilakukan dengan menggunakan teknik wawancara ini guna memperoleh informasi yang diperlukan untuk penelitian.

3. Studi Pustaka

Buku, jurnal, dan referensi yang memaparkan gagasan studi kasus terkait pembuatan sistem informasi pengelolaan inventaris obat di Puskesmas, serta topik terkait lainnya diperiksa untuk melakukan tinjauan pustaka.

3. HASIL DAN PEMBAHASAN

Data

Data Siswa SMK N 1 Waikabubak" yang berisi 247 record dan 5 atribut. Anda dapat mengisi detail lebih lanjut berdasarkan informasi yang tersedia:

Tabel 1 Dataset siswa SMK N 1 Waikabubak

NO	No. Tes	Nilai	Minat Siswa	Rekom BK	JURUSAN
1	YOHANIS TOGU LAWA	89	AKUNTANSI 1	PEMASARAN	PEMASARAN
2	GRACELA KAMURI	84	AKUNTANSI 1	PEMASARAN	PEMASARAN
3	MARLINCE TALLU	84	PEMASARAN	PEMASARAN	PEMASARAN
4	APLIANA MARA	85	PEMASARAN	PEMASARAN	PEMASARAN
5	APLIANA GOLE	85	PEMASARAN	PEMASARAN	PEMASARAN
6	MARLINCE JOWA	86	PEMASARAN	PEMASARAN	PEMASARAN
7	ADRIANA PIGE	84	PEMASARAN	PEMASARAN	PEMASARAN
8	JAMES ARIYANTO TALLU	88	PEMASARAN	PEMASARAN	AD. PERKANTORAN 1
9	LUKAS BATE UMBU LELE	86	AD. PERKANTORAN 1	PEMASARAN	AD. PERKANTORAN 1
10	SOLEMAN SAIRO	85	PEMASARAN	AD. PERKANTORAN 1	AD. PERKANTORAN 1
...
246	MARTHA MAGI BEWA	75	UPW	PHT 1	UPW
247	MARTHA BELA MAWO	76	UPW	PHT 1	UPW

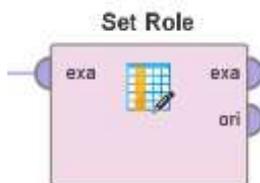
Penggunaan operator Read Excel di RapidMiner untuk mengimpor data dari file Excel "Data Siswa SMK N 1 Waikabubak.xlsx" ke dalam repositori RapidMiner. Dengan menggunakan operator **Read Excel** di RapidMiner, Anda dapat dengan mudah mengimpor dataset "Data Siswa SMK N 1 Waikabubak.xlsx" ke dalam repositori. Setelah diimpor, dataset ini dapat digunakan untuk berbagai analisis dan pemodelan untuk mendukung keputusan berbasis data.



Gambar 1 Operator *retrive* pada rapidminer

Selection

Dengan menggunakan operator Set Role, Anda dapat menetapkan peran pada atribut yang relevan dalam dataset "Data Siswa SMK N 1 Waikabubak". Ini membantu dalam mempersiapkan data untuk analisis lebih lanjut dan meningkatkan pemahaman mengenai faktor-faktor yang mempengaruhi penentuan jurusan siswa. Dengan atribut yang telah dipilih, proses analisis dapat dilakukan untuk menghasilkan rekomendasi yang lebih akurat bagi siswa.



Gambar 2 Operator *set role* pada rapidminer

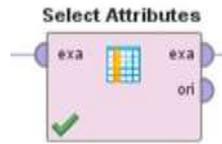
Penggunaan operator Set Role di RapidMiner sangat penting untuk memastikan bahwa atribut yang berfungsi sebagai ID dapat dikenali dengan jelas dalam analisis data. Dengan menetapkan role ini, Anda dapat mengelola dan memproses dataset dengan lebih efisien, serta meningkatkan keakuratan dalam analisis yang dilakukan Tabel 2 Parameter atribut yang dipilih pada operator *set role*

Tabel 2 Parameter atribut yang dipilih pada operator set role

NO	Parameter	Isi
1	Nama	Id
2	Jurusan	Label

Preprocessing

Pembersihan data merupakan langkah penting dalam proses data mining untuk memastikan bahwa data yang digunakan berkualitas dan sesuai dengan kebutuhan algoritma yang akan digunakan. Dengan mengikuti langkah-langkah di atas di RapidMiner, Anda dapat menyiapkan dataset yang bersih dan siap untuk dianalisis.



Gambar 3 Operator *select* atribut

Parameter pada operator *select attribute* yang digunakan tampak pada tabel dibawah ini.

Tabel 3 *Select atribut*

<i>Atribut</i>	<i>Select Atribut</i>
No	Nama
	Nilai
	Minat Siswa
	Rekom BK
	Jurusan

Dari hasil operator *select* atribut didapat informasi sebagai berikut.

Tabel 4 Hasil operator *select* atribut

<i>Atribut</i>
Nama
Nilai
Minat Siswa
Rekom BK
Jurusan

Transformation

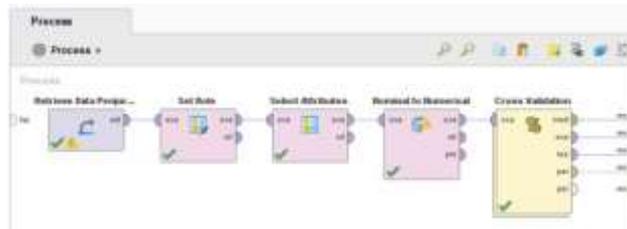
Tahap transformasi sangat penting untuk memastikan bahwa data dapat diproses dengan benar menggunakan algoritma yang dipilih. Dengan menggunakan operator **Nominal to Numerical** di RapidMiner, Anda dapat mengonversi atribut nominal menjadi numerik, memungkinkan algoritma C4.5 untuk berfungsi dengan baik. Pastikan untuk memverifikasi hasil konversi agar akurasi analisis tetap terjaga.



Gambar 4 Operator *nominal to numerical*

Data Mining

Tahap modeling adalah langkah krusial dalam proses data mining, di mana algoritma decision tree digunakan untuk klasifikasi data. Dengan mengikuti langkah-langkah di atas dalam RapidMiner, Anda dapat membangun dan menguji model dengan baik, sehingga menghasilkan prediksi yang akurat berdasarkan data yang telah dipersiapkan. Pastikan untuk memantau hasil cross-validation untuk memahami performa model yang dihasilkan..

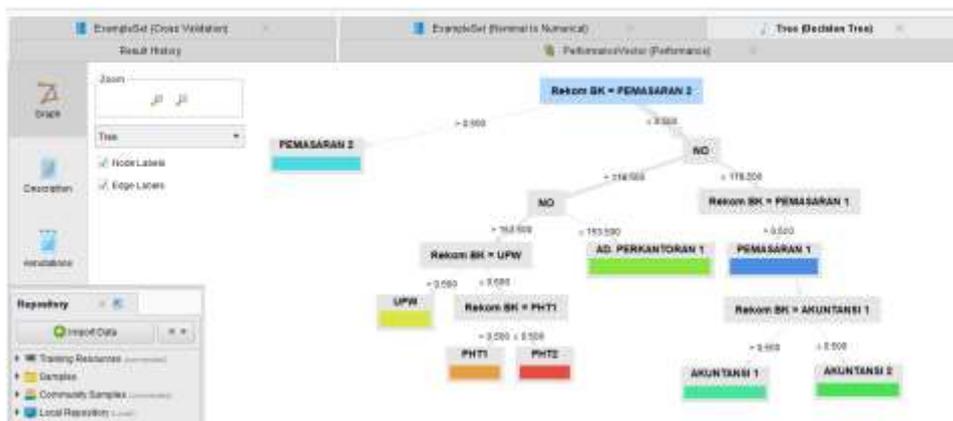


Gambar 5 Model proses langkah data mining

Dengan menggunakan teknik decision tree dan operator-operator dalam RapidMiner, Anda dapat membangun dan mengevaluasi model klasifikasi secara efektif. Penggunaan Cross Validation memastikan bahwa model diuji secara komprehensif, memberikan wawasan yang lebih baik tentang kinerja model pada data baru.



Gambar 6 Kumpulan operator cross validation



Gambar 7. decision tree

4. KESIMPULAN DAN SARAN

Berdasarkan uji yang telah dilakukan, penelitian ini menunjukkan bahwa data penjurusan siswa baru dapat berhasil diklasifikasikan menggunakan algoritma C4.5 melalui RapidMiner. Hasil klasifikasi dengan menggunakan pohon keputusan memberikan gambaran yang jelas mengenai variabel-variabel yang mempengaruhi pilihan yang dibuat oleh siswa saat memilih program studi. Variabel-variabel ini, seperti nilai ujian penjurusan dan rekomendasi dari bimbingan konseling (BK), berperan penting dalam proses pengambilan keputusan siswa.

DAFTAR PUSTAKA

- Ainurrohmah. (2021). Akurasi algoritma klasifikasi pada software RapidMiner dan Weka. *Prisma*, 4, 493–499. Retrieved from <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Aulia, D., & Budayawan, K. (2023). Penerapan algoritma C4.5 untuk klasifikasi jurusan siswa SMA Negeri 6 Padang. *Jurnal Vocational Teknik Elektronika dan Informatika*, 11(4), 1–9. Retrieved from <http://ejournal.unp.ac.id/index.php/voteknika/index>
- Hijrah, M., & Adri. (2022). Analisis RapidMiner dan Weka dalam memprediksi kualitas kinerja karyawan menggunakan metode algoritma C4.5. *Jurnal Teknik Informatika dan Sistem Informasi*, 9(2), 1655–1665.
- Jufani, M. N., Zahro, H. Z., & Achmadi, S. (2023). Pengembangan penentuan sistem pendukung keputusan penjurusan siswa di SMAN 1 Sanggar menggunakan metode Analytical Hierarchy Process (AHP) dan Technique for Order Preference By Similarity To Ideal Solution (TOPSIS). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 945–952. <https://doi.org/10.36040/jati.v6i2.5405>
- Nanang. (2022). Analisis perbandingan algoritma data mining metode decision tree C4.5 dengan Naïve Bayes dalam penjurusan siswa (studi kasus MAN 1 Kota Tangerang Selatan). *Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat*, 2(1), 44–61.
- Paul, V. M., Indra, G. M., Damanik, B. E., Parlina, I., & Saputra, W. (2021). Dalam menentukan kelayakan penerimaan bantuan bedah rumah pada Desa Tiga Dolok. *STIKOM Tunas Bangsa Pematangsiantar*, 1, 396–409.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan machine learning dalam berbagai bidang: Review paper. *Indonesian Journal of Computer and Information Technology*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Soleh, M., Andani, S. R., & Qurniawan, H. (2021). Algoritma C4.5 dalam mengklasifikasi tindak kejahatan yang sering terjadi di Kecamatan Siantar Barat Pematangsiantar. *Jurnal Ilmiah*, 2(5), 335–340.