

Implementasi Algoritma C4.5 dalam Menentukan Prediksi Pendaftaran Ulang Mahasiswa Baru di STT Payakumbuh

Rini Budiarni¹, Lusiana Eka Putri², Ranti Irsa³

Program Studi Informatika, Sekolah Tinggi Teknologi Payakumbuh, Indonesia¹²³
rinibudiarni309@gmail.com*¹, lucyzie91@gmail.com², ranti903@gmail.com³

Abstract

Re-registration is one of the self-reporting processes that must be carried out by new prospective students who have passed the selection process in order to get official status at a higher education institution. This stage is crucial for sustainability of the institution. However, not all accepted candidates do re-registration process. Therefore, it needs a predictive strategy to predict the numbers of the candidates that will do re-registration process. The objective of this study is to predict the number of new students who will complete the re-registration process at STT Payakumbuh by applying data mining techniques using the C4.5 algorithm. The variables involved in data processing include gender, age, average grade, school type, and school location. The implementation of the C4.5 algorithm resulted in a decision tree model comprising of 19 branches and 33 leaf node, with the "School Location" variable as the root of the decision tree. The decision tree model evaluation was conducted using a confusion matrix, yielding an accuracy rate 81,26%, a precision value of 82.82%, and recall value of 97,75%. These results indicates that the constructed model is effective as a predictive tool for supporting the re-registration process of new students.

Keywords: re-registration prediction, data mining, C4.5 algorithm, decision tree

Abstrak

Proses pendaftaran ulang merupakan tahapan pelaporan diri yang wajib dilaksanakan oleh calon mahasiswa baru yang telah dinyatakan lolos seleksi, guna memperoleh status resmi sebagai mahasiswa di perguruan tinggi. Proses ini merupakan tahapan penting karena berpengaruh langsung terhadap keberlangsungan institusi. Namun, tidak semua calon mahasiswa yang dinyatakan lolos seleksi melanjutkan proses ini, sehingga diperlukan strategi prediktif untuk memperkirakan jumlah mahasiswa yang akan melakukan pendaftaran ulang. Tujuan dari penelitian ini adalah untuk memprediksi pendaftaran ulang calon mahasiswa baru di STT Payakumbuh dengan menerapkan teknik data mining menggunakan algoritma C4.5. Variabel yang digunakan dalam proses pengolahan data mencakup jenis kelamin, umur, nilai rata-rata, jenis sekolah dan letak sekolah. Hasil penerapan algoritma C4.5 dalam pengolahan data menghasilkan model pohon keputusan dengan struktur 19 cabang dan 33 daun, di mana variabel "Letak Sekolah" menjadi akar dari pohon keputusan. Evaluasi model pohon keputusan dilakukan dengan menggunakan *confusion matrix* dengan yang menghasilkan tingkat akurasi sebesar 81,26 %, nilai precision sebesar 82,82% dan nilai recall sebesar 97,75%. Hasil ini menunjukkan bahwa model yang dibangun efektif digunakan sebagai alat bantu prediktif dalam mendukung proses pendaftaran ulang mahasiswa baru.

Kata kunci: prediksi, pendaftaran ulang, *data mining*, algoritma C4.5, pohon keputusan

1. Pendahuluan

Jumlah mahasiswa merupakan faktor penting yang berperan dalam menjamin keberlanjutan dan stabilitas operasional suatu perguruan tinggi. Prosedur penerimaan mahasiswa baru umumnya dimulai dari tahap pendaftaran, yang kemudian dilanjutkan dengan seleksi administrasi dan ujian masuk. Calon mahasiswa yang dinyatakan lolos dalam rangkaian seleksi tersebut harus melakukan pendaftaran ulang sebagai bentuk pengesahan status mereka sebagai mahasiswa. Pendaftaran ulang ini mencakup pemenuhan dokumen administratif dan penyelesaian kewajiban finansial dalam batas waktu yang telah ditetapkan oleh perguruan tinggi.

Permasalahan yang sering dijumpai baik adalah tidak semua calon mahasiswa yang dinyatakan lolos seleksi melanjutkan ke tahap pendaftaran ulang. Fenomena ini menjadi perhatian penting bagi institusi pendidikan tinggi karena dapat mempengaruhi perencanaan akademik dan anggaran operasional. Untuk mengatasi ketidakpastian tersebut, analisis terhadap data historis pendaftaran mahasiswa menjadi langkah strategis. Dengan memanfaatkan teknik data mining, khususnya algoritma C4.5, perguruan tinggi dapat membangun model prediktif yang berfungsi untuk memperkirakan jumlah mahasiswa baru yang berpotensi melakukan pendaftaran ulang.

Prediksi merupakan suatu pendekatan yang sistematis yang digunakan untuk mengestimasi kondisi yang kemungkinan besar akan terjadi di masa mendatang dengan memanfaatkan informasi yang ada pada saat ini dan informasi pada masa lalu. Langkah-langkah dalam proses prediksi dilakukan secara terstruktur guna mengurangi potensi kesalahan dalam hasil perkiraan [1]. Hasil prediksi dapat memberikan gambaran kondisi yang mendekati peristiwa nyata yang akan terjadi. Terdapat berbagai metode untuk melakukan prediksi terhadap kejadian di masa yang akan datang, salah satunya adalah dengan teknik data mining yang menawarkan analisis berbasis pola dari data yang telah ada.

Data mining merupakan suatu pendekatan analitis yang bertujuan untuk mengekstraksi informasi penting yang terdapat pada himpunan data yang tersedia melalui proses pengumpulan dan pengolahan data secara sistematis. Aktivitas ini dapat dilakukan dengan menerapkan metode statistik dan matematika, serta memanfaatkan perangkat lunak dan teknologi berbasis kecerdasan buatan. Pendekatan data mining berfokus pada pengidentifikasian pola, hubungan, dan kecenderungan tersembunyi dalam data melalui teknik pengenalan pola, sehingga memungkinkan terciptanya wawasan yang relevan untuk pengambilan keputusan [2]. Pola-pola tersebut diidentifikasi dari beragam jenis basis data, termasuk basis data relasional, gudang data, data transaksi, serta data berorientasi objek. Penerapan teknik data mining memiliki potensi untuk mendukung pengambilan keputusan institusional yang efisien dan akurat [3].

Salah satu metode dalam data mining yang berfungsi sebagai alat prediksi terhadap suatu kondisi di masa mendatang adalah algoritma C4.5. Proses pengolahan data dengan algoritma C4.5 menghasilkan sebuah model berbentuk pohon keputusan (*decision tree*) yang menyajikan skema pengambilan keputusan secara visual melalui pengidentifikasian alternatif, kondisi-kondisi yang relevan, serta kemungkinan dan konsekuensi dari setiap pilihan alternatif [4]. Pohon keputusan ini memungkinkan pembagian dataset yang besar menjadi subset yang lebih kecil berdasarkan serangkaian aturan keputusan.

Penelitian ini dilakukan dengan tujuan untuk mengetahui prediksi mahasiswa baru yang melakukan proses pendaftaran ulang di STT Payakumbuh melalui implementasi algoritma C4.5. Beberapa studi sebelumnya yang menjadi referensi dalam penelitian ini antara lain adalah penelitian oleh Linda Monizah Fitriani dan Andik Setyono (2018) yang menerapkan algoritma C4.5 dalam proses seleksi penerimaan peserta didik baru pada SD Islam Terpadu Permata Bunda Demak, dan memperoleh tingkat akurasi sebesar 90,50% melalui pengujian menggunakan RapidMiner [5]. Penelitian lain oleh Ratna Puspita Sari Putri dan Indra Waspada (2018) yang memanfaatkan algoritma C4.5 untuk memprediksi masa kelulusan mahasiswa dengan tingkat akurasi 62,44% [6]. Selain itu, penelitian oleh Siska Narulita, Andreas Tigor Oktaga dan Ika Susanti (2021) juga menggunakan metode ini untuk menguji akurasi model klasifikasi dalam menentukan peminatan peserta didik dengan tingkat akurasi sebesar 86,84% dengan nilai AUC sebesar 0,752, yang termasuk kategori klasifikasi wajar (*fair classification*) [7].

2. Metode Penelitian

Dalam penelitian ini diterapkan pendekatan kuantitatif, yang memanfaatkan data numerik terukur sebagai dasar analisis. Pengolahan data dilakukan menggunakan model statistik untuk memperoleh kesimpulan yang objektif [8]. Variabel penelitian terdiri atas sejumlah atribut yang merepresentasikan faktor-faktor prediktif dalam keputusan pendaftaran ulang mahasiswa baru. Atribut tersebut mencakup jenis kelamin, jenis sekolah, letak sekolah, umur, dan nilai rata-rata ujian sekolah bagi lulusan SMA/ sederajat atau Indeks Prestasi Kumulatif (IPK) untuk mahasiswa transfer dari perguruan tinggi lain.

Setiap atribut memiliki sejumlah kriteria, yang dalam algoritma pohon keputusan C4.5 berperan sebagai simpul (*node*), sedangkan kriteria atau nilai spesifiknya berfungsi sebagai cabang (*branch*) maupun daun (*leaf*). Proses penelitian mengikuti siklus kerja standar dalam data mining, dimulai dari tahap pengumpulan data, dilanjutkan dengan persiapan data, pengolahan data, dan diakhiri dengan evaluasi hasil.



Gambar 1. Tahapan penelitian

2.1 Pengumpulan data (*datasheet*)

Data yang digunakan dalam penelitian ini merupakan rekam jejak pendaftaran calon mahasiswa baru di STT Payakumbuh selama kurun waktu tujuh tahun terakhir. Informasi tersebut diperoleh melalui bagian akademik kampus dan selanjutnya diseleksi serta diorganisir sesuai dengan kebutuhan analisis untuk menunjang validitas penelitian.

2.2 Persiapan data (*preprocessing data*)

Tahapan persiapan data dilakukan untuk memastikan bahwa data memenuhi kriteria teknik yang dibutuhkan dalam proses data mining. Aktivitas utama dalam tahap ini mencakup pembersihan data (*data cleaning*) dan reduksi data (*data reduction*). Proses pembersihan data melibatkan penghapusan entri yang bersifat duplikat dan penanganan terhadap data kosong atau tidak lengkap. Sementara itu, reduksi data dilakukan dengan meminimalkan ukuran dataset melalui penggabungan data yang sejenis, mengeliminasi fitur yang dianggap tidak relevan serta melakukan pengelompokan variabel yang berkaitan dengan pendaftaran ulang mahasiswa.

2.3 Pengolahan data dengan teknik data mining

Setelah dataset yang relevan tersedia, langkah berikutnya adalah pengolahan data menggunakan metode data mining berbasis algoritma C4.5 untuk menghasilkan model berupa pohon keputusan. Menurut Larose (2005) [9], proses awal pembentukan pohon keputusan dimulai dari penyusunan data latih (*training data*), yang kemudian diikuti dengan perhitungan nilai entropy guna mengukur tingkat ketidakpastian dengan menggunakan rumus berikut:

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

dimana S merupakan jumlah himpunan kasus, n merupakan jumlah partisi himpunan kasus (S) dan p_i merupakan proporsi dari S_i terhadap S. Setelah nilai entropy diperoleh, dilakukan perhitungan nilai *information gain* untuk masing-masing atribut dengan menggunakan rumus berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

dimana A merupakan atribut, n merupakan jumlah partisi atribut (A), dan $|S_i|$ adalah proporsi terhadap S_i . Langkah berikutnya adalah menghitung nilai *split info* dari setiap atribut dengan menggunakan persamaan (3) dan menghitung *gain ratio* dengan menggunakan persamaan (4)

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(S, A)} \quad (4)$$

Atribut yang mempunyai nilai *gain ratio* paling tinggi akan dijadikan sebagai akar (*root*) pohon keputusan [10], sementara cabang-cabang pohon (*node*) ditentukan berdasarkan nilai *gain ratio* tertinggi dari atribut yang tersisa. Proses perhitungan ini dilakukan secara iteratif hingga semua atribut terpakai atau hingga kondisi berhenti terpenuhi.

2.4 Evaluasi dan Validasi

Tahapan akhir dari penelitian ini adalah evaluasi dan validasi terhadap pohon keputusan yang terbentuk. Evaluasi dilakukan dengan mengukur tingkat akurasi model prediksi menggunakan *confusion matrix*, yakni matriks klasifikasi yang menampilkan kombinasi antara nilai dugaan (prediksi) dan nilai sebenarnya (aktual). Confusion matrix merupakan representasi tabular yang digunakan untuk mengilustrasikan distribusi klasifikasi data uji, baik yang sesuai maupun yang tidak sesuai dengan labelnya [11]. Empat kategori evaluasi yang digunakan yaitu *True Positive* (TP), *True Negatif* (TN), *False Positif* (FP) serta *False Negatif* (FN). Berdasarkan hasil tersebut, dilakukan perhitungan nilai akurasi, presisi dan *recall* guna menilai kinerja model secara kuantitatif dan menyeluruh [12].

3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan cara mengumpulkan data histori pendaftaran calon mahasiswa baru STT Payakumbuh selama 7 tahun terakhir. Dari 720 data pendaftaran, 110 data tidak ditemukan secara fisik, sehingga data yang dapat digunakan berjumlah 610 data.

3.1. Data Training

Data training didapatkan dari data *sheet* yang telah melalui proses seleksi data, pembersihan data dan reduksi data.

Tabel 1. Data training

Jenis Kelamin	Nilai	Umur	Jurusan Sekolah	Letak Sekolah	Status
L	Sedang	Fresh Graduate	SMK Teknik	Cluster 1	IYA
L	Sedang	Medium Graduate	SMK Teknik	Cluster 1	IYA
P	Sedang	Fresh Graduate	SMA IPS	Cluster 1	IYA
P	Sedang	Fresh Graduate	SMA IPA	Cluster 3	IYA
L	Rendah	Fresh Graduate	SMK Non Teknik	Cluster 2	IYA
P	Tinggi	Fresh Graduate	SMA IPA	Cluster 1	IYA
P	Sedang	Fresh Graduate	SMA IPA	Cluster 2	IYA
L	Sedang	Fresh Graduate	SMA IPA	Cluster 2	IYA
P	Sedang	Fresh Graduate	SMA IPA	Cluster 4	IYA
P	Sedang	Medium Graduate	SMK Non Teknik	Cluster 1	IYA
P	Tinggi	Fresh Graduate	SMA IPA	Cluster 1	IYA
P	Tinggi	Old Graduate	Transfer PT	Cluster 1	IYA
P	Tinggi	Old Graduate	Transfer PT	Cluster 1	TIDAK
P	Sedang	Fresh Graduate	SMK Teknik	Cluster 1	IYA
L	Sedang	Medium Graduate	SMA IPS	Cluster 1	TIDAK
P	Sedang	Fresh Graduate	SMA IPS	Cluster 1	TIDAK
L	Sedang	Fresh Graduate	SMA IPA	Cluster 4	IYA
L	Sedang	Fresh Graduate	SMA IPS	Cluster 4	TIDAK
L	Sedang	Fresh Graduate	SMK Teknik	Cluster 1	IYA
...
...

3.2. Pembentukan Pohon Keputusan dengan Algoritma C4.5

Tahap awal dalam proses pembentukan pohon keputusan menggunakan algoritma C4.5 dimulai dengan identifikasi akar (*root*) pohon keputusan melalui perhitungan nilai entropy total. Nilai ini menjadi dasar untuk menentukan *information gain* dari masing-masing atribut yang selanjutnya digunakan dalam perhitungan nilai *split info*. Gabungan dari kedua nilai tersebut menghasilkan nilai *gain ratio* yang menjadi indikator utama dalam pemilihan atribut terbaik. Atribut dengan nilai *gain ratio* tertinggi akan ditempatkan sebagai simpul akar (*root*) pada pohon keputusan. Proses ini dilakukan secara iteratif, dan hasil perhitungan pada iterasi pertama ditampilkan dalam Tabel 2.

Tabel 2. Hasil perhitungan iterasi pertama untuk menentukan akar dari pohon keputusan

	JML KASUS (S)	TIDAK (S ₁)	IYA (S ₂)	ENTROPY	GAIN
Root	183	21	162	0,514091	

	JML KASUS (S)	TIDAK (S ₁)	IYA (S ₂)	ENTROPY	GAIN
Nilai lulus					0,005796
Rendah	14	3	11	0,749595	
Sedang	144	16	128	0,503258	
Tinggi	25	2	23	0,402179	
Umur					0,006233
FG	141	14	127	0,466743	
MG	33	5	28	0,613619	
OG	9	2	7	0,764205	
Jenis Sekolah					0,045173
IPA	43	1	42	0,159350	
IPS	48	10	38	0,738285	
Teknik	48	4	44	0,413817	
Non Teknik	31	5	26	0,637387	
Transfer	6	1	5	0,650022	
Lainnya	7	0	7	0	
Letak Sekolah					0,046846
Cluster 1	122	17	105	0,582519	
Cluster 2	35	0	35	0	
Cluster 3	6	0	6	0	
Cluster 4	20	4	16	0,721928	
Jenis Kelamin					0,003689
P	61	9	52	0,603652	
L	122	12	110	0,463777	

Berdasarkan hasil analisis pada Tabel 2, atribut “letak sekolah” mendapatkan nilai gain tertinggi sebesar 0,046846. Dengan demikian atribut ini dipilih sebagai simpul akar (*root*) dalam struktur pohon keputusan. Atribut “letak sekolah” terdiri dari empat kriteria, dimana dua diantaranya berada pada kelas yang sama dan memiliki nilai entropy 0. Oleh sebab itu, kedua kriteria tersebut langsung diklasifikasikan sebagai daun (*leaf*). Sementara dua kriteria lainnya berfungsi sebagai simpul cabang (*node*) yang akan dianalisis lebih lanjut pada tahap iterasi berikutnya.

Pada tahap iterasi kedua, berdasarkan hasil analisis yang ditampilkan dalam Tabel 3, atribut “jenis sekolah” memperoleh nilai gain tertinggi sebesar 0,054041. Oleh sebab itu, atribut tersebut dipilih sebagai simpul cabang pertama (node 1.1) dalam struktur pohon keputusan, dimana atribut ini memiliki satu simpul daun dan empat cabang.

Tabel 3. Hasil perhitungan pada iterasi 2 untuk menentukan node 1.1

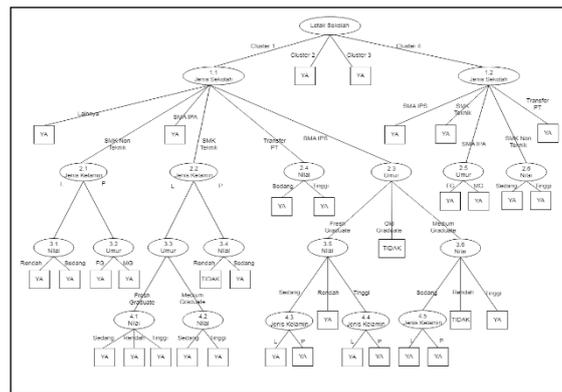
	JML KASUS (S)	TIDAK (S ₁)	IYA (S ₂)	ENTROPY	GAIN
Letak Sekolah					
Node 1.1 Cluster 1					
Total	122	17	105	0,582519	
Nilai lulus					0,014237
Rendah	9	3	6	0,918296	
Sedang	96	12	84	0,543564	
Tinggi	17	2	15	0,522559	
Kriteria Umur					0,010855
FG	94	11	83	0,520738	
MG	21	4	17	0,702467	
OG	7	2	5	0,863121	
Jenis Sekolah					0,054041
IPA	20	0	20	0	
IPS	36	8	28	0,764205	
Teknik	36	4	32	0,503258	
Non Teknik	24	4	20	0,650022	
Transfer	4	1	3	0,811278	
Lainnya	2	0	2	0	
Jenis Kelamin					0,018577
P	41	9	32	0,759276	
L	81	8	73	0,465070	

Selanjutnya, pada iterasi ketiga yang ditampilkan pada Tabel 4, atribut “jenis sekolah” kembali mendapatkan nilai gain tertinggi sehingga ditetapkan sebagai simpul lanjutan (node 1.2) dengan struktur yang terdiri atas tiga simpul daun dan dua cabang tambahan.

Tabel 4. Hasil perhitungan iterasi 3 untuk menentukan node 1.2

	JML KASUS (S)	TIDAK (S1)	IYA (S2)	ENTROPY	GAIN
Letak Sekolah					
Node 1,2 Cluster 4					
Total	20	4	16	0,721928	
Nilai lulus					
Rendah	1	0	1	0	
Sedang	18	4	14	0,764205	
Tinggi	1	0	1	0	
Kriteria Umur					
FG	16	3	13	0,696212	
MG	3	1	2	0,918296	
OG	1	0	1	0	
Jenis Sekolah					
IPA	7	1	6	0,591673	
IPS	5	2	3	0,970951	
Teknik	3	0	3	0	
Non Teknik					
Transfer	2	0	2	0	
Lainnya	0	0	0	0	
Jenis Kelamin					
P	2	0	2	0	
L	18	4	14	0,764205	

Secara keseluruhan, pohon keputusan yang terbentuk setelah seluruh proses iterasi selesai terdiri dari 19 cabang dan 33 simpul daun dengan atribut “letak sekolah” menjadi akar utama pada struktur pohon keputusan, seperti yang ditunjukkan pada Gambar 3.



Gambar 3. Pohon keputusan prediksi pendaftaran ulang mahasiswa baru

3.3. Evaluasi

Tahap evaluasi dilakukan dengan mengukur tingkat akurasi dari pohon keputusan yang telah dibentuk melalui penggunaan metode *confusion matrix*. Teknik yang digunakan adalah *percentace split*, yang memisahkan data menjadi dua bagian yaitu data pelatihan dan data pengujian. Kombinasi nilai prediksi dan nilai aktual hasil klasifikasi dalam Tabel 5, guna memberikan gambaran performa model dalam mengidentifikasi kelas secara tepat.

Tabel 5. Kombinasi nilai prediksi dan nilai aktual hasil klasifikasi

Kelas	Prediksi		
	Positif	Negatif	
Aktual	Positif	347	8
	Negatif	72	0

Berdasarkan Tabel 5, tingkat akurasi dari pohon keputusan mencapai 81,26% yang menunjukkan efektivitas model dalam memprediksi pendaftaran ulang mahasiswa baru. Ketepatan informasi yang dihasilkan oleh sistem terhadap kebutuhan pengguna, yang direpresentasikan melalui nilai *precision*, mencapai 82,82%. Sementara itu, kemampuan sistem dalam mengenali dan mengembalikan informasi yang relevan ditunjukkan oleh nilai *recall* sebesar 97,75%.

Hasil evaluasi tersebut mencerminkan bahwa model klasifikasi yang digunakan memiliki performa yang cukup baik dalam konteks akurasi, ketepatan dan keberhasilan pengambilan informasi. Dengan demikian, pohon keputusan

yang dibentuk dapat dijadikan sebagai alat bantu prediktif yang dapat diandalkan dalam mendukung pengambilan keputusan institusional.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma C4.5 dalam proses data mining untuk memprediksi calon mahasiswa baru yang akan melakukan pendaftaran ulang di STT Payakumbuh. Melalui tahapan pengumpulan, pembersihan, reduksi dan pengolahan data selama tujuh tahun terakhir, terbentuk sebuah model pohon keputusan yang mampu mengidentifikasi pola dan kecenderungan dari data historis. Evaluasi terhadap model menggunakan confusion matrix menunjukkan tingkat akurasi sebesar 81,26% dengan nilai *precision* 82,82% dan *recall* sebesar 97,75% yang mencerminkan kinerja klasifikasi yang cukup optimal. Dengan demikian, algoritma C4.5 terbukti efektif sebagai alat prediksi dan analisis yang dapat membantu perguruan tinggi dalam merancang strategi penerimaan dan pengelolaan mahasiswa baru secara lebih akurat dan terukur.

Daftar Rujukan

- [1] Herdianto. 2013. Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Saraf Tiruan Backpropagation. Tesis. Universitas Sumatera Utara, Medan
- [2] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey and Sons, Inc.
- [3] Permata Sari, D., Buana, W. ., & Febri Mayang Sari, M. . 2025. Implementasi Data Mining pada Penjualan Barang dengan Teknik K Means. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, 5(1), 106–112.
- [4] Heizer Jay dan Render, Barry. 2017. *Manajemen Operasi Edisi 11*. Salemba Empat. Jakarta.
- [5] Linda M. Fitriani dan A. Setyono. 2018, Penerapan Algoritma C4.5 untuk Seleksi Penerimaan Siswa Baru pada SD Islam Terpadu Permata Bunda Demak. *CCIT (Creative Communication and Innovative Technology) Journal*. Vol. 11 No. 2: 158-170.
- [6] Ratna Puspita Sari Putri dan Indra Waspada. 2018. Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika, *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*. Vol. 4 No. 1: 1-7.
- [7] Siska Narulita, Andreas Tigor Oktaga, Ika Susanti. 2021. Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4.5 untuk Penentuan Peminatan Peserta Didik. *Jurnal Media Aplikom*. Vol 13 No.2:68-82. Purwokerto: Stikom Yos Sudarso Publisher.
- [8] Sugiyono. 2018. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*, Alfabeta.
- [9] Normah, I. Yulianti, D. Novianti, M. N. Winnarto, A. Zumarniansyah, and S. Linawati. 2020. Comparison of Classification C4.5 Algorithms and Naïve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program. *Journal of Physics:Conference Series*. Vol. 1641 No. 1.
- [10] Devinda Yulia, Abdi Pandu Kusuma, Dimas Fanny H. Permadi. 2022. Penerapan Algoritma C4.5 Untuk Prediksi Minat Penjurusan Siswa Di SMKN 1 Kademangan. *JATI (Jurnal Mahasiswa Teknik Informatika)*. Vol. 6 No.2:893-900
- [11] Robi Nurhidayat, Kania Evita Dewi. 2023. Penerapan Algoritma K-Nearest Neighbor dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek. *KOMPUTA:Jurnal Ilmiah Komputer dan Informatika*. Vol. 12 No. 1:91-100
- [12] Siswanto, D., Zamzami, Nijal, L., & Rajab, S. 2022. Analisa Sentimen Publik Mengenai Perekonomian Indonesia Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Metode Klasifikasi K-NN Dan Svm. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*. Vol 2 No.1:1–9