

OPTIMALISASI ALGORITMA C4.5 MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI KELULUSAN SISWA SMKN 2 CIMAH

Castaka Agus Sugianto*¹, Nadya Selvi Pujiyanita²

^{1,2}Program Studi Teknik Informatika, Politeknik TEDC Bandung, Indonesia
e-mail: *¹castaka@poltektedc.ac.id, ²pujiyanita16@gmail.com

Abstrak

Kata Kunci:

Klasifikasi;
Kelulusan Siswa;
Data Mining;
Prediksi;
Algoritma C4.5;
Algoritma Genetika;
T-test.

SMKN 2 Cimahi merupakan lembaga pendidikan sekolah menengah kejuruan di cimahi. Sekolah menengah kejuruan sangat berperan untuk menciptakan lulusan terbaik bagi kebutuhan dunia kerja. Proses kelulusan adalah kegiatan paling akhir dari manajemen peserta didik. kriteria kelulusan siswa untuk ujian nasional harus memenuhi nilai minimal, menyelesaikan semua mata pelajaran, mengikuti ujian sekolah, ujian nasional, uji kompetensi. Pada angkatan tahun 2015 yang berjumlah 300 siswa, siswa yang tidak lulus pada ujian nasional berjumlah 1 siswa, pada angkatan tahun 2016 yang berjumlah 299 siswa, siswa yang tidak lulus pada ujian nasional berjumlah 1 siswa, pada angkatan tahun 2017 yang berjumlah 302 siswa 100% siswa lulus tepat waktu, namun pada tahun yang akan mendatang masih belum diketahui apakah siswa akan lulus 100% atau tidak. Maka dari itu penelitian ini dilakukan untuk memprediksi kelulusan siswa dengan menggunakan algoritma C4.5+ algoritma genetika. Algoritma C4.5 mempunyai tingkat *accuracy* sebesar 99,78%, *precision* 99,78%, *recall* 100% dan *execution time* 1 *second*. Sedangkan algoritma C4.5 + algoritma genetika mempunyai tingkat *accuracy* sebesar 99,78%, *precision* 99,78%, *recall* 100% AUC 0,500, dan *execution time* 36 *second* dan setelah dilakukan pengujian *T-test* antara algoritma C4.5 dan algoritma C4.5 + algoritma genetika dikatakan signifikan jika $\alpha=0.050$. Dalam hasil penelitian ini didapat $\alpha=1.000$.

Abstract

Keywords:

Classification;
Student Graduation;
Data Mining;
Prediction;
C4.5 algorithm;
Genetic Algorithm;
T-test.

SMK 2 Cimahi is an educational institution Vocational high school in Cimahi. Vocational high schools play an important role in creating the best graduates for the needs of the workforce. The graduation process is the latest activity of student management. Student graduation criteria for national exams must meet minimum scores, complete all subjects, take school exams, national exams, and competency tests. In the 2015 class of 300 students, students who did not pass the national exam were 1 student, in the 2016 class of 299 students, students who did not pass the national exam were 1 student, whereas in the 2017 class totaling 302 students 100% students graduate on time. However in the coming year, it is still unknown whether students will be 100% graduated or not. Therefore this study was conducted to predict student graduation using the C4.5 algorithm + Genetic Algorithm. C4.5 algorithm has an accuracy rate of 99.78%, precision of 99.78%, 100% recall and 1 second execution time. While the C4.5 + genetic algorithm has an accuracy rate of 99.78%, precision of 99.78%, 100% recall AUC 0.500, and 36 second execution time and after T-test testing between the C4.5 Algorithm and C4 Algorithm .5 +, Genetic Algorithm is said to be significant if $\alpha = 0.050$. The results of this study obtained $\alpha = 1,000$.

✉ Alamat korespondensi :

E-mail : castaka@poltektedc.ac.id

p-ISSN: 2087-1627, e-ISSN: 2685-9858

1. Pendahuluan

Sekolah Menengah Kejuruan (SMK) merupakan suatu lembaga pendidikan formal yang lulusannya dipersiapkan untuk memasuki dunia kerja. Standar proses pendidikan adalah standar nasional pendidikan yang berkaitan dengan pelaksanaan pembelajaran pada satu-satuan pendidikan untuk mencapai standar kompetensi lulusan (peraturan pemerintah No. 19 tahun 2005 Bab 1 pasal 1 Ayat 6).

Ujian nasional adalah penilaian hasil belajar oleh pemerintah yang bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok ilmu pengetahuan dan teknologi. Ujian Nasional merupakan agenda Nasional yang dilakukan setiap Tahun di SMKN 2 Cimahi.

Proses kelulusan adalah kegiatan paling akhir dari manajemen peserta didik. Kelulusan adalah pernyataan dari lembaga pendidikan (sekolah) tentang telah diselesaikannya program pendidikan yang harus diikuti oleh peserta didik. Kriteria kelulusan siswa untuk ujian nasional harus memenuhi nilai minimal, menyelesaikan semua mata pelajaran, mengikuti ujian sekolah, ujian nasional, uji kompetensi. Pengolahan data dengan teknik *data mining*, di harapkan dapat memanfaatkan data yang sebelumnya untuk dapat menentukan faktor mana yang menjadi penunjang paling dominan di dalam tingkat kelulusan seorang siswa.

Di SMKN 2 Cimahi setiap mendekati waktu ujian nasional sekolah kesulitan dalam hal perencanaan program dan kebijakan yang akan diterapkan khusus untuk meluluskan semua siswa yang akan melakukan ujian nasional. Pada angkatan tahun 2015 yang berjumlah 300 siswa, siswa yang tidak lulus pada ujian nasional berjumlah 1 siswa, pada angkatan tahun 2016 yang berjumlah 299 siswa, siswa yang tidak lulus pada ujian nasional berjumlah 1 siswa, pada angkatan tahun 2017 yang berjumlah 302 siswa 100% siswa lulus tepat waktu, namun pada tahun yang akan mendatang masih belum diketahui apakah siswa akan lulus 100% atau tidak. Sehingga pihak sekolah setiap tahun selalu mengawatirkan tentang tingkat kelulusan siswanya.

Proses pengolahan melalui *data mining* dapat di gunakan untuk memprediksi tingkat kelulusan siswa SMKN 2 Cimahi di masa yang akan datang Banyak penelitian yang melakukan hal yang sama seperti halnya yang dilakukan oleh Alimancon Sijabat yang berjudul “Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree” hasilnya bahwa perilaku baik lah yang menghantarkan siswa

ini lulus, hal ini membuktikan algoritma C4.5 cukup efektif [1][2]. Algoritma C4.5 ini Iayak digunakan untuk prediksi nilai kelulusan siswa sekolah menengah [3]. Sedangkan penelitian abdul rohman yang berjudul “Model Algoritma *K-Nearest Neighbor* (K-Nn) Untuk Prediksi Kelulusan Mahasiswa”. Dalam penelitian ini dalam memprediksi kelulusan mahasiswa dengan menggunakan algoritma klasifikasi *data mining K-Nearest Neighbor* dengan mengklaster *data* $k=1$, $k=2$, $k=3$, $k=4$, dan $k=5$. Hasil yang diperoleh dengan *cluster* data $k=5$ *accuracy* adalah 85,15% dan nilai *AUC* adalah 0.888 adalah akurasi paling tinggi [4].

Beberapa algoritma dalam klasifikasi diantaranya adalah C4.5, ID3, *k-nearstneighbor classifer*, *Neive Bayes*, SVM, dan ANN[5]. Algoritma C4.5 Merupakan algoritma yang cukup efektif untuk memprediksi data kelulusan[6]. Algoritma genetika adalah suatu algoritma yang memodelkan proses evolusi dari spesies biologi melalui seleksi alam, sedangkan keuntungan dari algoritma genetika adalah sifat metoda pencariannya yang lebih optimal, tanpa terlalu memperbesar ruang pencarian[7]. Beberapa penelitian yang menggunakan algoritma genetika seperti yang dilakukan irfan ali dkk Penelitian ini mencoba menerapkan metode *artificial neural network* menggunakan algoritma genetika untuk memprediksi kelulusan mahasiswa, mendapatkan nilai akurasi untuk model *artificial neural network* sebesar 71.48% dan akurasi untuk model *artificial neural network* berbasis algoritma genetika sebesar 99.33%[8].

Sedangkan penelitian Meilani & Susanti yang berjudul “Aplikasi *Data Mining* Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode *Algoritma Genetika*”. Pada analisa data yang dilakukan diproses *testing*, di dapat tingkat keakuratan sistem sekitar 99,82% dan memiliki nilai *error* 0.18%[9].

Tujuan yang ingin di capai yaitu untuk mengetahui pola dan indikator yang mempengaruhi kelulusan siswa smkn 2 cimahi dengan menggunakan algoritma C4.5 + algoritma genetika. Diharapkan dengan dilakukan penelitian ini dapat membantu SMKN 2 Cimahi untuk mengetahui lebih awal faktor apa saja yang menjadi indikator kelulusan SMKN 2 Cimahi.

2. Metode Penelitian

2.1 Penentuan Masalah

Penentuan masalah yang di lakukan peneliti merupakan proses awal dalam penelitian ini, yaitu mengenai masalah tentang prediksi kelulusan siswa SMKN 2 cimahi.

Tabel 1. Tabel Dataset Data Kelulusan SMKN 2 Cimahi

NO	NO PESERTA	NAMA PESEERTA	ASAL SEKOLAH	TAHUN	X1	X2	X3	X4	X5	X6	X7	X8	X9	X 10	X 11	X 12	JUMLAH	KETERANGAN
1.	4-15-02-05-102-001-8	AGUNG DANA PRAWIRA	SMPN 1 CISARUA	2015	75	52	15	82	76	81	75	73	8	8	72	72	285	LULUS
2.	4-15-02-05-102-002-7	AGUS RIAN	SMPN 1 BATUJAJAR	2015	79	64	42	78	82	83	75	73	86	54	45	47	318	LULUS
3.	4-15-02-05-102-046-3	IRSYAD ARDIANSYAH	SMPN 1 PANGALENGAN	2015	0	0	0	0	0	0	0	0	74	78	6	65	27.7	TIDAK LULUS
4	4-16-02-05-102-154-7	MEGI SETIA	SMP PGRI 1 CIMAH	2017	0	0	0	0	0	0	0	0	70	52	6	74	25.6	TIDAK LULUS
...
...
901	41702051023023	YUSTISIA ALDI DARMAWAN	SMP NEGERI 2 CIMAH	2017	76	60	32	85	82	76	78	73	62	52	42	45	305	LULUS

2.2 Mencari Literature

Mencari *literature* merupakan mencari informasi yang berkaitan dengan penelitian seperti *literature data mining, algoritma C4.5, decision tree*, dan prediksi kelulusan. Adapun yang dilakukan dalam penelitian ini adalah dengan melakukan studi pustaka, yang merupakan sebuah metode yang dilakukan dengan cara mengumpulkan dan mencari sumber dari buku, internet, jurnal yang berkaitan dengan penelitian yang dilakukan.

2.3 Pengumpulan Data

a. Data Penelitian

Proses pengumpulan data dilakukan dengan cara mengambil data dari daftar kolektif hasil ujian nasional di SMKN 2 Cimahi. Data yang digunakan dalam penelitian ini diambil dari hasil daftar kolektif hasil ujian nasional tahun pelajaran 2015 s/d tahun pelajaran 2017 dengan sebanyak 901 siswa. Data tersebut diambil dari awal bulan desember 2017 sampai bulan Mei 2018. Proses preparasi data dilakukan, yaitu dengan mengubah dari data *hardcopy* kedalam format *excel*, seperti pada Tabel 1.

b. Proses Reparasi Data set

Sebelum dilakukan proses *data mining* pada proses ini *dataset* akan di reparasi yaitu menghilangkan *field* yang tidak perlu dipakai ketika proses data mining di *RapidMiner*. *Dataset* yang digunakan terdapat 19 *field* bisa dilihat pada tabel 2 pada proses ini *field* yang di hilangkan adalah *field no*. Sehingga *field* yang digunakan pada *dataset* yang digunakan untuk proses *data mining* menjadi 18 *field*.

Tabel 2. Tabel Keterangan Field

No	Atribut	Jenis Data	Type Data
1	NO	<i>integer</i>	<i>Id</i>
2	NO PESERTA	<i>integer</i>	<i>atribut</i>
3	NAMA PESERTA	<i>varchar</i>	<i>atribut</i>
4	ASAL SEKOLAH	<i>integer</i>	<i>atribut</i>
5	TAHUN	<i>integer</i>	<i>atribut</i>
6	X1	<i>integer</i>	<i>Atribut</i>
7	X2	<i>integer</i>	<i>atribut</i>
8	X3	<i>integer</i>	<i>atribut</i>
9	X4	<i>integer</i>	<i>atribut</i>
10	X5	<i>integer</i>	<i>atribut</i>
11	X6	<i>integer</i>	<i>atribut</i>
12	X7	<i>integer</i>	<i>atribut</i>
13	X8	<i>integer</i>	<i>atribut</i>
14	X9	<i>integer</i>	<i>atribut</i>
15	X10	<i>integer</i>	<i>atribut</i>
16	X11	<i>integer</i>	<i>atribut</i>
17	X12	<i>integer</i>	<i>atribut</i>
18	JUMLAH	<i>integer</i>	<i>atribut</i>
19	KETERANGAN	<i>varchar</i>	<i>atribut</i>

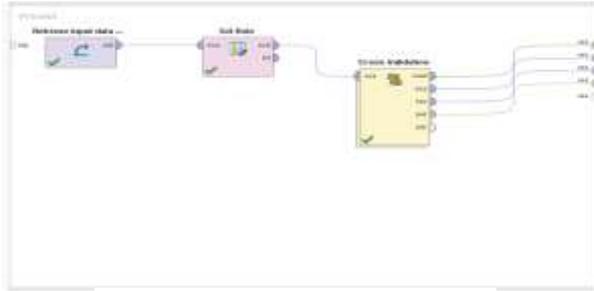
c. Proses dan Hasil

Proses data mining yang dilakukan menggunakan algoritma *C4.5* yang menghasilkan nilai akurasi, dan pohon keputusan dengan menggunakan algoritma genetika.

3. Hasil dan Pembahasan

3.1 Pengujian Menggunakan Algoritma C4.5

Proses Pengujian ini dilakukan untuk mengetahui nilai *accuracy, precision, recall, Area Under Curve* yang keluar serta mengetahui seperti apa hasil *tree* (pohon keputusan) beserta *rule* yang akan dihasilkan pada proses data mining. Proses tersebut dapat dilihat pada Gambar 1.



Gambar 1. Pengujian Algoritma C4.5

Berikut adalah Hasil dari pengujian algoritma C4.5 dengan menggunakan validasi *cross validation*. dari tabel tabel 3 menunjukkan bahwa algoritma C4.5 mendapatkan hasil *accuracy* sebesar 99,78%, *Precision* 99,78%, *Recall* 100,00%, dan *Execution Time* 1 second.

Tabel 3. Hasil Pengujian Algoritma Naïve Bayes

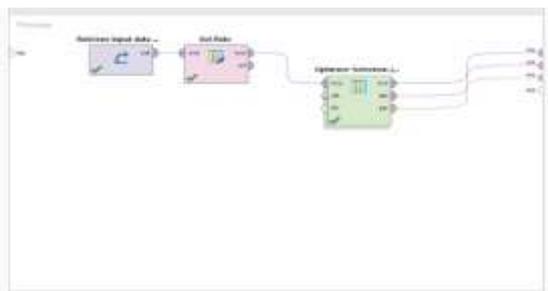
	<i>true</i> LULUS	<i>true</i> TIDAK LULUS	<i>Class</i> <i>precision</i>
pred.LULUS	899	2	99.78%
pred. TIDAK LULUS	0	0	0.00%
<i>class recall</i>	100.00%	0.00%	



Gambar 2. Hasil *tree* pohon keputusan

3.2 Pengujian Model

Dalam pengujian ini menggunakan algoritma C4.5 + algoritma genetika prosesnya bisa dilihat pada gambar 3.



Gambar 3. Pengujian algoritma C4.5 dan Algoritma Genetika

Berikut adalah Hasil dari pengujian menggunakan algoritma C4.5 + algoritma genetika dengan menggunakan validasi *cross validation*. dari tabel 4 menunjukkan bahwa algoritma C4.5 + algoritma genetika mendapatkan hasil *accuracy* sebesar 99,78%, *Precision* 99,78%, *Recall* 100,00%, *AUC* 0,500, dan *Execution Time* 36 second.

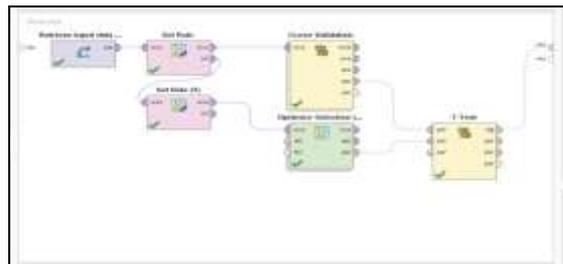
Tabel 4. Hasil pengujian akgoritma C4.5 + algoritma genetika

	<i>true</i> LULUS	<i>True</i> TIDAK LULUS	<i>Class</i> <i>precision</i>
pred.LULUS	899	2	99.78%
pred. TIDAK LULUS	0	0	0.00%
<i>class recall</i>	100.00%	0.00%	

3.3 Hasil Pengujian Menggunakan T-Test

Pengujian *t-test* dalam penelitian ini menggunakan *dataset* hasil kelulusan SMKN 2 Cimahi dan di validasi dengan menggunakan dua Algoritma C4.5 + Algoritma Genetika. Proses pengujian bisa di lihat pada gambar 4.

Pengujian menggunakan *t-test* untuk mengetahui apakah data yang telah di olah menggunakan Algoritma C4.5 dan Algoritma C4.5 + Algoritma Genetika hasilnya signifikan atau tidak, dikatakan signifikan jika $\alpha=0.050$. Dalam hasil penelitian ini didapat $\alpha=1.000$, jadi dalam penelitian ini hasilnya tidak signifikan karna di hasil *t-test* menunjukkan hasil diatas 0.050. Dikatakan tidak signifikan bisa di lihat pada tabel 5 pada hasil *t-test* dalam pengujian antara Algoritma C4.5 dan Algoritma C4.5 + Algoritma Genetika.



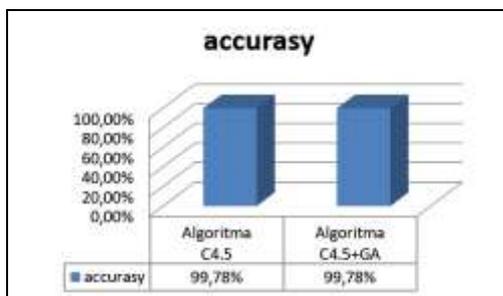
Gambar 4. Pengujian menggunakan *t-test*

Tabel 5. Hasil Pengujian T-Test

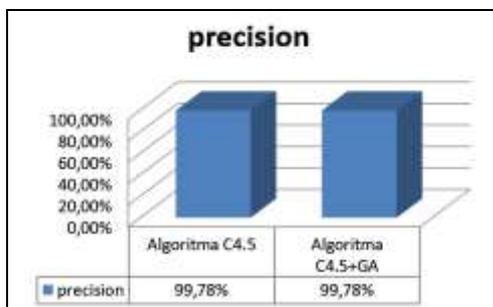
A	B	C
	C4.5	C4.5 + Algoritma Genetika
C4.5		1.000
C4.5 + Algoritma Genetika		

3.4 Evaluasi Sistem

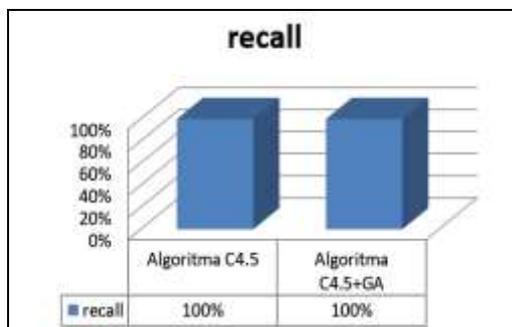
Hasil perbandingan dari Algoritma C4.5 dan Algoritma C4.5 + algoritma genetika mendapatkan hasil yang sama namun dalam segi waktu eksekusi algoritma sangat berbeda. Algoritma C4.5 mempunyai tingkat *accuracy* sebesar 99,78%, *precision* 99,78%, *recall* 100% dan *execution time* 1 *second*. Sedangkan algoritma genetika mempunyai tingkat *accuracy* sebesar sebesar 99,78%, *precision* 99,78%, *recall* 100% dan *execution time* 36 *second*. Pada gambar 5 bisa dilihat grafik *accuracy* hasil perbandingan. Pada gambar 6 bisa dilihat grafik *precision* hasil perbandingan dan pada gambar 7 bisa dilihat grafik *Recall* hasil perbandingan.



Gambar 5. Grafik Accuracy



Gambar 6 Grafik Precision



Gambar 7 Grafik Recall

Berdasarkan keseluruhan pengujian dan validasi algoritma C4.5 dan algoritma C4.5 + algoritma genetika mendapatkan hasil yang sama namun dalam segi waktu eksekusi algoritma sangat

berbeda. Lebih cepat algoritma C4.5 dibandingkan dengan algoritma C4.5 + algoritma genetika. Dan hasil pengujian *t-test* mendapatkan hasil $\alpha=1.000$, bahwa hasilnya tidak signifikan karena hasil *t-test* menunjukkan hasil di atas 0.050.

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, menemukan pola untuk memprediksi kelulusan SMKN 2 Cimahi yaitu hasil *tree* pohon keputusan yang menunjukkan Jumlah yaitu jika Jumlah >203.750 dinyatakan lulus dan jika ≤ 203.750 dinyatakan tidak lulus. Berdasarkan keseluruhan pengujian dan validasi algoritma C4.5 dan algoritma C4.5 + algoritma genetika mendapatkan hasil yang sama namun dalam segi waktu eksekusi algoritma sangat berbeda. Algoritma C4.5 mempunyai tingkat *accuracy* sebesar 99,78%, *precision* 99,78%, *recall* 100% dan *execution time* 1 *second*. Sedangkan algoritma genetika mempunyai tingkat *accuracy* sebesar sebesar 99,78%, *precision* 99,78%, *recall* 100% dan *execution time* 36 *second*. Pengujian *t-test* mendapatkan hasil pengujian yang didapat $\alpha=1.000$, bahwa hasilnya tidak signifikan karena di hasil *t-test* menunjukkan hasil di atas 0.050.

Dari penelitian ini terdapat 11 indikator yang mempengaruhi prediksi kelulusan SMKN 2 cimahi yaitu: Asal sekolah, Tahun Pelajaran, Nilai Ujian Nasional Bahasa Indonesia, Nilai Ujian Nasional Bahasa Inggris, Nilai Ujian Sekolah Bahasa Inggris, Nilai Ujian kompetensi, Nilai Ebtanas Murni SMP Bahasa Indonesia, Nilai Ebtanas Murni SMP Bahasa Inggris, Nilai Ebtanas Murni SMP Matematika, Nilai Ebtanas Murni SMP IPA, Jumlah.

Beberapa saran dari penulis untuk penelitian selanjutnya adalah disarankan bagi penelitian selanjutnya sebaiknya membuat aplikasi. Untuk mendapatkan hasil yang lebih baik lagi kedepannya perlu dilakukan penambahan indikator seperti perilaku siswa, pergaulan dilingkungan, pengaruh permainan elektronik dari komputer, serta telepon seluler dan dilakukan juga mencoba metode algoritma yang berbeda untuk mendapatkan hasil prediksi lebih baik dan akurat seperti *K-Nearest Neighbor*, *Neural Network*, *Support Vector Machine*, *ID3*, *CART*, *Linear Discriminant Analysis*.

Daftar Pustaka

- [1] A. Sijabat, "Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree (Studi Kasus : Yayasan Perguruan Kristen Andreas)," *Maj. Ilm. Inf. dan Teknol. Ilm.*, vol. 5, no. 3, pp. 7–12, 2015.
- [2] C. A. Sugianto, "PENERAPAN TEKNIK DATA MINING UNTUK MENENTUKAN

- HASIL SELEKSI MASUK SMAN 1 GIBEBER UNTUK SISWA BARU MENGGUNAKAN DECISION TREE,” *J. TEDC*, vol. 9, no. 1, pp. 39–43, 2015.
- [3] R. H. Pambudi, “Penerapan Algoritma C4.5 untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal,” Universitas Brawijaya, 2017.
- [4] A. Rohman, “Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa,” *Neo Tek.*, vol. 1, no. 1, pp. 1–9, 2015.
- [5] S. S. Nikam, “A Comparative Study of Classification Techniques in Data Mining Algorithms,” *Int. J. Mod. Trends Eng. Res.*, vol. 4, no. 7, pp. 58–63, 2017.
- [6] S. Maslahah, “Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5,” Universitas Islam Negeri Maulana Malik Ibrahim, 2013.
- [7] M. A. Kartika, “PENCOCOKAN KATA SECARA ACAK DENGAN METODE ALGORITMA GENETIKA MENGGUNAKAN PROGRAM PASCAL,” *J. Matametika*, vol. 2, no. 2, pp. 1–9, 1975.
- [8] I. Ali, L. Sularto, D. Sudrajat, D. A. Kurnia, and Y. A. Wijaya, “Optimasi Parameter Artificial Neural Network Menggunakan Algoritma Genetika Untuk Prediksi Kelulusan Mahasiswa,” *J. ICTInformation Commun. Technol.*, vol. 18, no. 1, pp. 54–59, 2019.
- [9] B. D. Meilani and N. Susanti, “Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naïve Bayes,” *J. Ilm. NERO*, vol. 1, no. 3, pp. 182–189, 2015.