

Analisis Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Nilai Gagal Menggunakan Metode Algoritma C4.5

Ela Roza Batubara¹, Zakarias Situmorang²

^{1,2}Jurusan Ilmu Komputer, Universitas Potensi Utama

e-mail: ¹ elabatubara22@gmail.com , ² zakarias65@yahoo.com

Intisari

Tujuan dari penelitian ini adalah untuk mengetahui faktor yang paling dominan terhadap kesulitan mahasiswa dalam menyelesaikan nilai gagal. Pada penelitian ini peneliti menggunakan teknik Data Mining dengan menggunakan algoritma C4.5. Sumber data penelitian yang digunakan diperoleh dari wawancara dan pemberian angket kepada mahasiswa secara acak di lingkungan AMIK Tunas Bangsa Pematangsiantar. Adapun variabel penelitian yang digunakan adalah biaya yang dikeluarkan (C1), waktu yang disempatkan (C2), motivasi untuk berusaha mengerjakan (C3), kesadaran mahasiswa/mahasiswi (C4), dan informasi yang didapat (C5). Pada penelitian ini alternatif yang digunakan sebagai sampel adalah Mahasiswa AMIK Tunas Bangsa, semester enam. Hasil dari klasifikasi menggunakan algoritma C4.5 dan pengujian dengan software rapidminer didapatkan faktor yang paling dominan kesulitan mahasiswa menyelesaikan nilai gagal adalah biaya yang dikeluarkan (C1) dengan nilai gain sebesar 0,15577244.

Kata kunci— Algoritma C4.5, Data Mining, klasifikasi

Abstract

The purpose of this study was to determine the most dominant factor in students' difficulties in completing failed grades. In this study, researchers used Data Mining techniques using the C4.5 algorithm. Sources of research data used were obtained from interviews and giving questionnaires to students at random in

Analisis Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Nilai Gagal Menggunakan Metode Algoritma C4.5

(Ela Roza Batubara, Zakarias Situmorang)

the AMIK Tunas Bangsa Pematangsiantar environment. The research variables used are the costs incurred (C1), time taken (C2), motivation to try to work (C3), student awareness (C4), and information obtained (C5). In this study, the alternative used as a sample was AMIK Tunas Bangsa students, sixth semester. The results of the classification using the C4.5 algorithm and testing with rapidminer software found that the most dominant factor in the difficulty of students completing the failed score was the cost incurred (C1) with a gain value of 0.15577244.

Keywords— C4.5 Algorithm, Data Mining, classification

PENDAHULUAN

Dalam mengetahui faktor yang paling dominan terhadap kesulitan mahasiswa menyelesaikan skripsi dapat menggunakan hasil klasifikasi data mahasiswa. Klasifikasi merupakan kegiatan dalam mengekstrak dan memprediksi kategori untuk masing-masing data. “Klasifikasi adalah suatu teknik yang dilakukan untuk memprediksi class atau properti dari setiap instace data. Teknik ini dilakukan pada data baru dengan memanipulasi data yang ada dan telah diklasifikasikan menggunakan hasilnya untuk memberikan sejumlah aturan [1]. Tujuan utama dari penambangan data ini untuk menemukan pengetahuan baru yang tersembunyi dari database tersebut [2]. Decision tree adalah struktur aturan yang divisualisasikan kedalam bentuk pohon (tree), dalam setiap simpul internal bertanda tes atribut, setiap cabang menggambarkan hasil tes, dan simpul daun menggambarkan kelas atau distribusi kelas. Decision tree memiliki alur diawali dari simpul akar, batang sampai kepada simpul daun yang menjelaskan hasil prediksi kelas [3]. manfaat utama penggunaan pohon keputusan adalah kemampuannya untuk membuat proses pengambilan keputusan yang kompleks menjadi lebih sederhana sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan [4]

“Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma ini memiliki kelebihan yaitu mudah dipahami, fleksibel dan menarik karena dapat diterapkan dalam bentuk gambar (pohon keputusan)”[5].

METODE PENELITIAN

Data Mining

Data mining merupakan proses perulangan dan memerlukan interaksi manusia dalam prosesnya untuk menemukan pola atau model baru yang dapat digeneralisasi untuk masa yang akan datang [6] Secara garis besar data mining juga di sebut dengan (KDD) yang mana merupakan singkatan dari knowledge Discovery in Database yang berfungsi memanfaatkan database.

Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma *decision tree*. Algoritma ini mempunyai *input* berupa training samples dan samples [6] 1 Secara umum langkah untuk membangun pohon keputusan Algoritma C4.5 adalah sebagai berikut

- a) Pilih atribut sebagai pakar
- b) Buat cabang untuk setiap nilai
- c) Bagi kasus dalam cabang
- d) Ulangi proses untuk setiap cabang sampai semua cabang mempunyai kelas yang sama Untuk menghitung *gain* menggunakan rumus seperti sebagai berikut

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

Keterangan

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai *gain* maka dicari terlebih dahulu nilai *entropy*. Untuk mencari nilai *entropy* menggunakan rumus sebagai berikut :

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

S : Himpunan Kasus

n : Jumlah partisi S

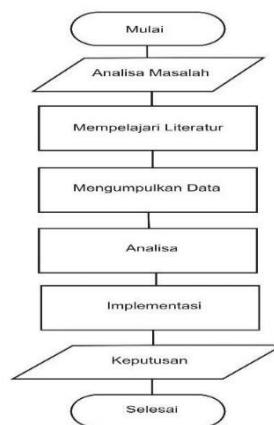
p_i: Proporsi dari S_i terhadap S

Metodologi penelitian memberikan gambaran rancangan penelitian yang meliputi antara lain : rincian semua pelaksanaan penelitian dan menganalisis serta perancangna sistem yang digunakan. Tujuan dari penelitian ini adalah untuk mengumpulkan informasi dan mengelola data untuk menyelesaikan permasalahan dalam penelitian untuk mencari jawaban terhadap suatu masalah yang akan diteliti.[7].

HASIL DAN PEMBAHASAN

Pada bab ini akan mendapatkan hasil dari penelitian untuk mencari klasifikasi factor mana yang menyebabkan mahasiswa sulit atau tidak sulitnya menyelesaikan Nilai Gagal. Proses perhitungan untuk menentukan klasifikasi yang sering terjadi menggunakan Algoritma C4.5 adalah dengan jumlah kasus untuk Setelah itu akan dilakukan perhitungan manual dan penyesuaian hasil melalui pengujian menggunakan software RapidMiner 8.1.

Berikut hasil flowcart rancangan penelitian yang penulis lakukan, dapat di lihat pada gambar berikut:



Gambar 1. Rancangan Penelitian

Gambar 1 menjelaskan rancangan penelitian yang dilakukan untuk mencari faktor dominan mahasiswa kesulitan dalam menyelesaikan skripsi dengan menggunakan algoritma C4.5 yang terdiri dari sebagai berikut :

1. Analisa Masalah

Masalah yang terkait dengan kesulitan mahasiswa dalam menyelesaikan skripsi dilakukan dengan menganalisa faktor-faktor yang membuat sulit dalam menyelesaikan skripsi.

2. Mempelajari Literatur

Penelitian ini didasari rujukan yang digunakan untuk mendapatkan informasi dalam penelitian.

Analisis Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Nilai Gagal Menggunakan Metode Algoritma C4.5

(Ela Roza Batubara, Zakarias Situmorang)

3. Mengumpulkan Data

Data dikumpul dengan menggunakan kuisisioner kepada mahasiswa AMIK Tunas Bangsa semester 6 angkatan 2016 selama 1 minggu dari tanggal 3 Februari 2019 s/d 8 Februari 2019.

4. Analisa

Proses yang dilakukan untuk mencari faktor dominan mahasiswa kesulitan dalam menyelesaikan skripsi dengan menggunakan faktor- faktor yang digunakan dalam penelitian.

5. Implementasi

Implementasi dilakukan dengan menggunakan aplikasi *RapidMiner* versi 5.0 sebagai sistem untuk mencari keputusan dalam analisa mahasiswa sulit menyelesaikan skripsi.

6. Keputusan

Hasil yang diberikan oleh sistem yang akan ditindaklanjuti oleh pihak akademik dan dapat mencari solusi untuk menimalisir perpanjangan SK dan adanya pemebelian skripsi.

Analisis Data

Dalam penelitian ini data yang digunakan akan diolah dari hasil kuesioner, yang diberikan kepada mahasiswa semester 6 dengan jumlah mahasiswa sebanyak 140. Penulis memberikan kuesioner kepada mahasiswa dengan jumlah 100 sampel data. Berikut tabel pengambilan sampel data mahasiswa yang digunakan :

Tabel 1. Hasil Data Kusioner Mahasiswa

N	Nama Mahasiswa	Biaya	Waktu	Motivas	Kesadaran	Informasi	Keterangan
1	Agus Triyono	Baik	Baik	Cukup	Cukup	Baik	Sulit
2	Apri Yanti Simanullang	Baik	Cukup	Baik	Sangat Baik	Sangat Baik	Sulit
3	Andirwan Sembiring	Baik	Cukup	Baik	Baik	Baik	Tidak Sulit
4	Agus Alfiansyah	Sangat Baik	Baik	Sangat Baik	Kurang Baik	Baik	Sulit
5	Ari Aldiansyah	Sangat Baik	Cukup	Cukup	Cukup	Baik	Sulit
6	Andre Sanseko	Sangat Baik	Cukup	Baik	Sangat Baik	Baik	Tidak Sulit
7	Cahyo Pramana	Sangat Baik	Cukup	Baik	Sangat Baik	Sangat Baik	Sulit
8	Edgar Yosua	Sangat Baik	Baik	Baik	Kurang Baik	Sangat Baik	Sulit
9	Febri Fernanda	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Sangat Baik	Tidak Sulit
10	Gabriel Elang	Baik	Baik	Baik	Kurang Baik	Baik	Sulit
11	Gita Cahyani	Sangat Baik	Cukup	Baik	Sangat Baik	Baik	Sulit
12	Hendra Pasrah	Baik	Cukup	Baik	Sangat Baik	Baik	Tidak Sulit
13	Hnais Fahira Purba	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Tidak Sulit
14	Hans Martua	Sangat Baik	Baik	Sangat Baik	Baik	Sangat Baik	Tidak Sulit

Alat Analisis Data

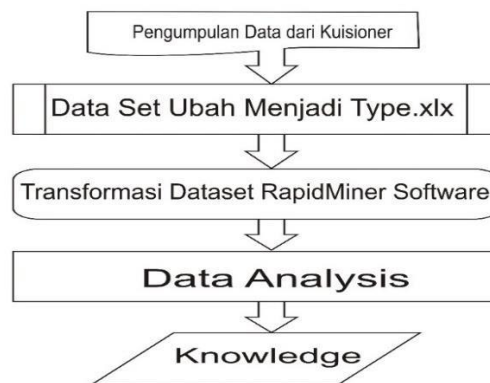
RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source) yang dibuat dengan menggunakan bahasa pemrograman java sehingga dapat diakses oleh semua sistem operasi. RapidMiner dapat menjadi solusi dalam melakukan analisis terhadap datamining, Rapidminer 5.0 digunakan pada studi ini untuk melakukan analisis faktor kesulitan mahasiswa dalam menyelesaikan nilai gagal.

Instrumen Penelitian

Dalam penelitian ini, penulis melakukan beberapa dataset file dalam bentuk *spreadsheet* file excel 2013 (xls). Transformasi data ini diperlukan sebagai masukan untuk perangkat lunak yang digunakan dalam penelitian ini. Dalam pengolahan data transformasi, penulis menggunakan perangkat lunak *RapidMiner* 5.0 Instrumen penelitian seperti Gambar 3 dapat dijelaskan bahwa data penelitian berasal dari data kusioner mahasiswa semester 6. Setelah data dipilih yang berbentuk klasifikasi maka data kemudian dianalisis menggunakan *software* rapidminer seperti disajikan pada Gambar 2.

Analisis Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Nilai Gagal Menggunakan Metode Algoritma C4.5

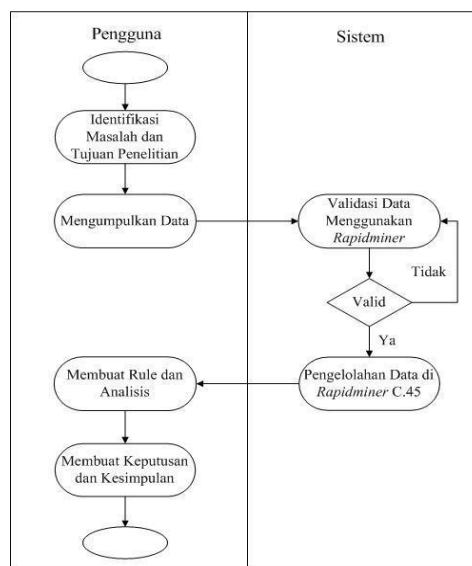
(Ela Roza Batubara, Zakarias Situmorang)



Gambar 2. Proses Percobaan

Diagram Kerja Aktifitas Penelitian

Berikut ini alur kerja yang akan dilakukan pada penelitian ini yang digambarkan dalam diagram aktivitas pada Gambar 3.3.



Gambar 3. Diagram Aktivitas Kerja Penelitian

Gambar 3 menjelaskan penulis atau pengguna mengidentifikasi masalah dan tujuan penelitian yang dilakukan, mengumpulkan data berupa kuisisioner yang diberikan, pada aplikasi *RapidMiner*.

Pengelolaan Data Menggunakan Algoritma C4.5

Dalam melakukan implemetasi hasil akhir dari algoritma C4.5 dilakukan dua tahap yaitu perhitungan manual dan penyusaian hasil akhir dari perhitungan manual dengan *software rapidminer 5.3*.

Perhitungan Entropy dan Gain

Perhitungan algoritma C4.5 dimulai dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan sulit dan jumlah kasus keputusan tidak sulit. variabel penelitian yang digunakan adalah biaya yang dikeluarkan (C1), waktu yang disempatkan (C2), motivasi untuk berusaha mengerjakan (C3), kesadaran mahasiswa/mahasiswi (C4), dan informasi yang didapat (C5). Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Hasil perhitungan menggunakan excel dapat di lihat pada table 4.1 dibawah ini.

Tabel 1. Perhitungan *Node 1*

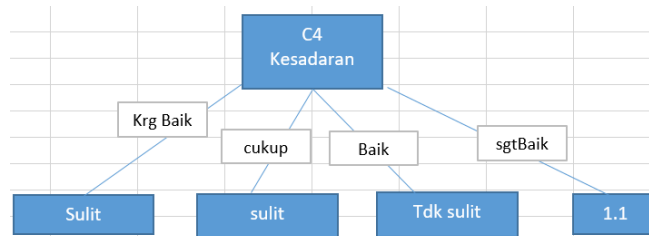
nilai	jih kasus	sulit	tidak sulit	entropy	gain
baik	5	4	1	0,16820242	1,226002343
sangat baik	9	6	3	0,143212766	
cukup	7	4	3	0,07103869	
baik	6	4	2	0,143212766	1,281243741
sangat baik	1	1	0	0	
cukup	2	2	0	0	
baik	8	7	1	0,192645078	1,268057084
sangat baik	4	0	4	0	
kurang baik	3	3	0	0	
cukup	2	2	0	0	1,342620641
baik	2	0	2	0	
sangat baik	7	4	3	0,07103869	
baik	8	6	2	0,171073935	1,219006552
sangat baik	6	4	2	0,143212766	

Langkah 3 : Pada tabel diatas dapat kita lihat bahwa atribut Membangun Suasana Kerja memiliki *gain* paling tinggi yaitu **1,342620641**. Maka atribut Membangun Suasana Kerja menjadi *node* akar. Dimana terdiri dari 2 sub atribut yaitu Puas dan Tidak Puas. Berdasarkan nilai *entropy* dari kedua sub atribut diatas, Maka perlu dilakukan perhitungan lebih lanjut untuk menemukan node akar Dari hasil diatas dapat dilihat bahwa faktor dengan nilai *Gain* tertinggi

Analisis Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Nilai Gagal Menggunakan Metode Algoritma C4.5

(Ela Roza Batubara, Zakarias Situmorang)

Berikut adalah hasil dari pohon keputusan untuk node 1:

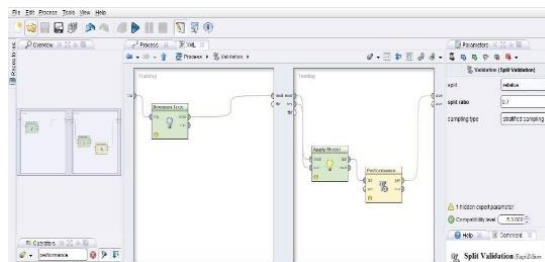


Gambar 4. Pohon Keputusan Node 1

Berdasarkan dari node 1 diatas, dapat dilihat bahwa kriteria kasus masih harus di perhitungkan lagi serta dianalisis lebih lanjut sehingga mendapatkan hasil yang diinginkan. Untuk mencari nilai Gain dan entropy digunakan persamaan (1) dan (2) sama seperti cara sebelumnya. Dan berikut adalah hasil dari perhitungan node 1.1 “Jenis Pidana-Penjara atau Denda”. Begitu sterusnya.

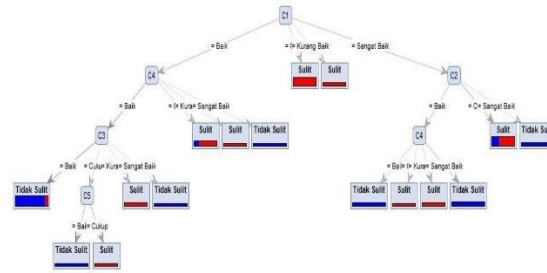
Pengujian Menggunakan Rapidminer

Tampilan menghubungkan port decision tree, apply model dan performance



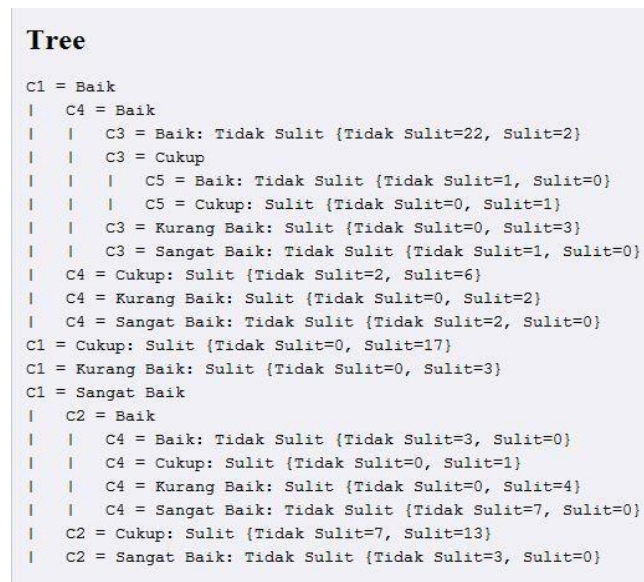
Gambar 5. Tampilan menghubungkan port decision tree, apply model dan performance

Untuk mendapatkan hasil berupa pohon keputusan maka pada tahap selanjutnya dapat dilakukan dengan cara mengklik tanda panah biru yang terdapat diposisi tengah atas pada bagian *toolbar*. Hasil akhir yang akan ditampilkan adalah berupa *decision tree* dimana hasil dari pengujian data akan tampak seperti pohon yang bercabang yang menampilkan isi dan hasil dari penelitian. Tampak pada gambar berikut :



Gambar 6. Pohon Keputusan

Setelah dilakukan perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir.:



Gambar 7. Rule Model Decision Tree

HASIL DAN PEMBAHASAN

Berdasarkan dari penjelasan diatas mengenai tahap-tahap penggunaan serta hasil yang telah ditampilkan maka selanjutnya membahas mengenai keterkaitan dari hasil yang didapat antara perhitungan manual Algoritma dengan hasil yang ditampilkan oleh *tools RapidMiner*.

Analisis Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Nilai Gagal Menggunakan Metode Algoritma C4.5

(Ela Roza Batubara, Zakarias Situmorang)

Dalam melakukan validasi perhitungan manual Algoritma harus telah menampilkan hasil akhir berupa pohon keputusan, serta data yang digunakan haruslah data yang *valid* dan sama dengan yang dipakai pada *tools*. Dari gambar diatas dapat disimpulkan bahwa data yang digunakan adalah data *valid* dan dapat dibuktikan dengan *tools RapidMiner* serta dapat menampilkan hasil yang sama dari perhitungan manual maupun dengan *tools RapidMiner*.

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, penulis menarik beberapa kesimpulan diantaranya:

1. Permasalahan menentukan faktor kesulitan mahasiswa menyelesaikan nilai gagal dapat diselesaikan menggunakan teknik data mining, yaitu dengan Algoritma C4.5. Menghasilkan 14 (empat belas) rules dan Tingkat akurasi yang dihasilkan oleh metode tersebut adalah 95,00%.
2. Dari perhitungan dengan algoritma C4.5 maka didapatkan faktor yang paling dominan adalah motivasi lulus tepat waktu (C1) dengan nilai gain sebesar 0,204593531

UCAPAN TERIMA KASIH

Terimakasih pada program studi Ilmu Komputer Universitas Potensi Utama atas dukungan dalam terlaksananya penelitian ini serta bapak zakarias situmorang.

DAFTAR PUSTAKA

- [1] A. Asroni, B. Masajeng Respati, and S. Riyadi, "Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta," *Semesta Tek.*, vol. 21, no. 2, pp. 158–165, 2018, doi: 10.18196/st.212222.
- [2] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi: 10.15575/join.v2i1.71.
- [3] A. Pratama, A. A. Wicaksana, and A. Razi, "ANALISA KESESUAIAN LAHAN TANAH UNTUK TANAMAN PADI (ORYZA SATIVA L .) DENGAN METODE DECISION TREE BERBASIS WEB (STUDI KASUS : KABUPATEN ACEH UTARA)," vol. 6, no. 1, 2022.
- [4] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [5] A. H. Nasrullah, "Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 244–250, 2018, doi: 10.33096/ilkom.v10i2.300.244-250.
- [6] E. S. Pribadi, P. Poningsih, and H. S. Tambunan, "Analisa Tingkat Kepuasan Masyarakat Terhadap Pelayanan Pengadilan Agama Pematangsiantar Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 1, pp. 33–40, 2020, doi: 10.30645/brahmana.v2i1.46.
- [7] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.