

Data Mining dalam Menganalisis Faktor Alasan Pemilihan Perumahan

Erlin Elisa^{a,*}, Alfannisa Annurullah Fajrin^b

^{ab} Universitas Putera Batam, Batam

*Erlin.Elisa@puterabatam.ac.id

Abstract

The house is a basic human need, where the house is a means of resting and gathering with the family, the choice of housing type greatly affects the comfort of someone who will live in the house, such as choosing where the location, facilities, accessibility and even more important is the price that fits the pocket of the candidate. housing owners, this study discusses analysis in the selection of housing in batam city, the variables used are the accessibility factors related to distance and ease of reaching a place., typology is related to the size of the settlement, price, payment for houses, data analysis is carried out with the concept of extracting information on the C4.5 algorithm by testing the results and obtaining the decision tree using the opensource rapidminer tools. From the data processing, the result shows that the factors that really influence people to choose and buy a house in the research object are price, accessibility and typology.

Keywords: *Housing; Data Mining; C4.5 Algorithm; Rapidminer.*

Abstrak

Rumah merupakan kebutuhan pokok manusia, dimana rumah adalah sarana tempat istirahat dan berkumpul dengan keluarga, pemilihan tipe perumahan sangat mempengaruhi kenyamanan bagi seseorang yang akan tinggal dirumah tersebut seperti memilih dimana lokasi, fasilitas, aksesibilitas bahkan yang tidak kalah penting adalah harga yang sesuai kantong bagi calon pemilik perumahan, penelitian ini membahas tentang analisa dalam pemilihan perumahan pada suatu daerah dikota batam, variable yang digunakan yaitu aksesibilitas faktor yang berhubungan dengan jarak dan kemudahan untuk mencapai suatu tempat., tipologi berhubungan dengan faktor luasnya pemukiman, harga, pembayaran terhadap rumah, analisa data dilakukan dengan konsep penggalian informasi algoritma C4.5 dengan pengujian hasil dan perolehan pohon keputusan menggunakan tools *opensource rapidminer*. Dari pengolahan data didapat hasil bahwasanya faktor yang sangat mempengaruhi orang untuk memilih dan membeli rumah pada objek penelitian adalah harga, aksesibilitas dan tipologi.

Kata Kunci: *Perumahan; Data Mining; Algoritma C4.5; Rapidminer.*

1. Pendahuluan

Rumah merupakan kebutuhan primer yang harus dipenuhi oleh masyarakat sebagai tempat berteduh, melepas lelah sejenak dari hiruk pikuknya kehidupan, semua orang menjadikan rumah sebagai tempat berkumpul dengan sanak keluarga dan orang-orang terdekat. Kota Batam adalah salah satu kota di Provinsi Kepulauan Riau yang banyak didatangi para perantau untuk mencari pekerjaan dan menetap dikota ini. Kota dengan slogan Bandar Madani ini hanya berjarak 13 km dari wilayah Negara Singapura. Badan Pusat Statistik (BPS) Kota Batam mencatat proyeksi pertumbuhan penduduk di Batam pada 2019 ini sebanyak 1.376.009 jiwa. Jumlah ini meningkat dibanding jumlah penduduk tahun 2018 sebesar 1.329.773 jiwa, atau bertambah

234.193 jiwa dibandingkan pertumbuhan jumlah penduduk lima tahun lalu. Dengan jumlah penduduk yang semakin bertambah apalagi pada tahun 2020 ini maka pemukiman akan semakin bertambah pula sejalan dengan kebutuhan masyarakat akan tempat tinggal, salah satunya perumahan yang bernaung adalah cipta asri grub yang berlokasi pada kelurahan tembesi kecamatan sagulung kota batam, dimana cipta asri menyajikan perumahan dengan fasilitas yang bersaing dengan perumahan lainnya, hal ini sangat mempengaruhi terhadap calon pemilih rumah untuk memilih tinggal disini. Perumahan cipta asri tersedia dengan beberapa macam tipe yaitu Tipe 22, Tipe 36 dan Tipe 36+ dan harga rumah berkisar 250 juta hingga 300 jutaan tergantung tipe dan kelebihan tanah yang

dimiliki seperti lokasi di Huk atau persimpangan.

Perumahan Cipta Asri group yang memiliki banyak kelebihan dan dijual dengan harga yang kompetitif, disini lain memiliki kondisi lokasi perumahan yang terletak tidak jauh dari industri batamindo dan wisata barelang dan sangat aman dan nyaman sebagai tempat tinggal serta fasilitas yang dimiliki seperti sekolah, berada pada lingkungan perumahan dan merupakan daerah bebas banjir. Hal ini yang membuat peneliti tertarik untuk melakukan penelitian terkait dengan menganalisa faktor pemilihan perumahan pada cipta asri group sehingga dapat membantu masyarakat dalam memilih perumahan yang ideal untuk ditinggali. Metode yang digunakan untuk menganalisa adalah salah satu metode data mining dengan teknik pohon keputusan atau algoritma C4.5 dengan memanfaatkan *rapidminer* sebagai alat analisis data faktor pemilihan. Algoritma C4.5 pernah dimanfaatkan oleh peneliti sebelumnya dalam mengklasifikasi potensi *droup out* mahasiswa yang merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas (Nasrullah, 2018). Dan kali ini peneliti akan mengujinya dengan kasus faktor pemilihan perumahan.

2. Kajian Literatur

2.1 Perumahan

Kebutuhan manusia akan rumah sangatlah penting dalam kehidupan, seperti yang dikatakan menurut Maslow menyebutkan bahwa setelah terpenuhinya kebutuhan manusia akan sandang, pangan, dan kesehatan, maka kebutuhan akan rumah atau tempat tinggal merupakan salah satu motivasi untuk pengembangan kehidupan yang lebih baik. Tempat tinggal pada dasarnya merupakan wadah bagi manusia atau keluarga dalam melangsungkan kehidupannya (Kalesaran et al., 2013).

Berdasarkan Undang-undang Nomor 1 Tahun 2011 tentang Perumahan dan Pemukiman. Perumahan adalah kelompok rumah yang berfungsi sebagai lingkungan tempat tinggal atau lingkungan hunian yang dilengkapi dengan sarana dan prasarana lingkungan.

2.2 Data mining

Data mining sebagai proses menemukan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi dari sekumpulan data besar untuk membantu dalam pengambilan keputusan (Harahap & Sulindawaty, 2020).

Data mining, sering juga disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian atau, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. *knowledge discovery in database* (KDD) pada intinya adalah proses menemukan pengetahuan yang bermanfaat dari kumpulan data (Tukino, 2019).

Menurut Tan, *data mining* merupakan pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Pola yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru. Data mining adalah bagian dari *Knowledge Discovery in Database*. *Knowledge Discovery in Database* adalah sebuah proses otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola (Halim & Prasetyo, 2018).

2.3 Klasifikasi

Klasifikasi dapat digambarkan sebagai berikut. Data input, disebut juga *training set*, terdiri atas banyak contoh (*record*), yang masing-masing memiliki beberapa atribut. Selanjutnya, tiap contoh diberikan sebuah label *class* khusus (Erlin Elisa, 2019).

Lebih lanjut, input didefinisikan sebagai sekumpulan *record* (*training set*), dan setiap *record* terdiri atas sekumpulan atribut, salah satu atribut adalah klas. Adapun model klasifikasi digunakan untuk antara lain:

- Pemodelan Deskriptif sebagai perangkat penggambaran untuk membedakan objek-objek dari klas berbeda.
- Pemodelan Prediktif digunakan untuk memprediksi label klas untuk *record* yang tidak diketahui atau tidak dikenal (Erlin Elisa, 2019).

2.4 Pohon Keputusan

Pohon keputusan merupakan salah satu metode klasifikasi yang populer karena dapat dengan mudah diinterpretasi oleh manusia. Pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Pohon keputusan memiliki *node* pohon yang merepresentasikan atribut yang telah diuji dan setiap cabangnya merupakan suatu pembagian hasil uji serta *node* daun (*leaf*) merepresentasikan kelompok kelas tertentu (Bimo et al., 2020).

Data dalam pohon keputusan dinyatakan dalam bentuk tabel dengan atribut dan *record*.

2.5 Algoritma C4.5

Konsep dasar dari algoritma C4.5 yaitu mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*). Algoritma C4.5 memetakan atribut menjadi kelas yang dapat diterapkan untuk klasifikasi baru. Kelebihan dari algoritma C4.5 yaitu mudah dimengerti, fleksibel dan menarik karena mampu divisualisasikan dalam gambar. Secara umum langkah-langkah algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Ente et al., 2020).

- Memilih atribut sebagai akar dengan berdasarkan nilai gain tertinggi pada masing-masing atribut yaitu dengan cara:

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

dimana S adalah himpunan kasus, A adalah atribut, n merupakan jumlah partisi atribut A, $|S_i|$ adalah jumlah kasus pada partisi ke-i dan $|S|$ adalah jumlah kasus dalam S. Sementara itu, untuk menghitung nilai entropy dapat

$$\text{Entropy}(S) = -\sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

dimana S adalah himpunan kasus, A adalah fitur, n adalah jumlah partisi S, p_i adalah proporsi dari S_i terhadap S.

- Membuat cabang pada masing-masing nilai.
- Membagi kasus dalam cabang
- Mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2.6 Rapidminer

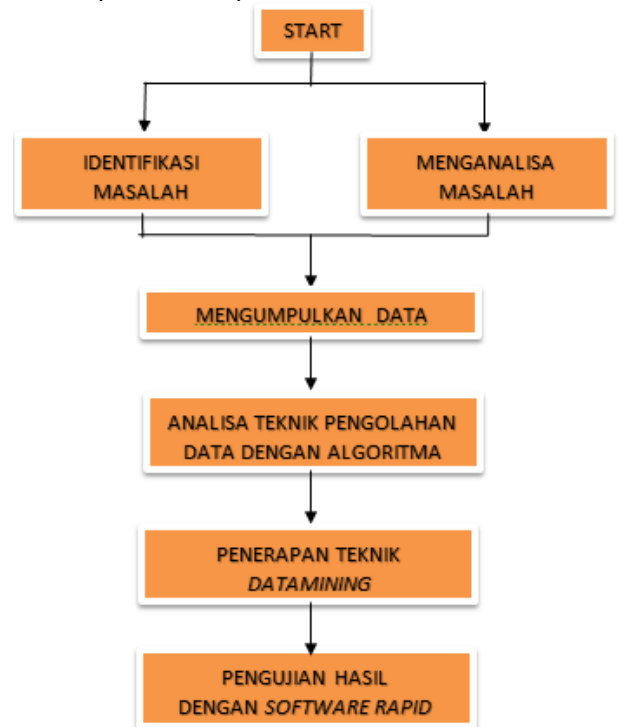
Rapidminer merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi (Luluk Elvitaria, 2017).

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan

menggunakan program Java di bawah lisensi GNU Public Licence dan Rapid Miner dapat dijalankan di sistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-lain (Haryati et al., 2015).

3. Metode Penelitian

Adapun desain penelitian ini adalah:



Gambar 1. Desain Penelitian

Langkah awal yang dilakukan peneliti adalah mengidentifikasi permasalahan sesuai dengan topik penelitian yaitu pemilihan perumahan bagi masyarakat yang akan membeli atau memilih rumah di cipta asri grub, dengan variabel ukuran yaitu Aksesibilitas, Tipologi, Harga dan Pembayaran dengan keputusan Beli dan Tidak Beli kemudian dari permasalahan yang maka tahapan yaitu menganalisa masalah kemudian mengumpulkan data dilapangan untuk kemudian diolah dengan teknik datamining algoritma C4.5, data yang di dapat akan dilakukan pra-proses sebelum dilakukan penetapan teknik C4.5, kemudian akan di uji kedalam software rapidminer untuk melihat pohon keputusan yang di hasilkan mengenai faktor pemilihan perumahan.

4. Hasil dan Pembahasan

4.1 Hasil

Tipe perumahan dengan memanfaatkan teknik data mining algoritma C4.5, pada tahap pra-proses dibutuhkan data keputusan terhadap pemilihan perumahan oleh calon penghuni pada objek penelitian. Berikut adalah data sampel yang berupa tabel yang akan dilakukan penambangan sesuai dengan langkah Decision Tree .

Tabel 1. Data Sampel

Aksesibilitas	Tipologi	Harga	Pembayaran	Keputusan
Dekat	Luas	Rendah	Kontan	Beli
Dekat	Luas	Tinggi	Kontan	Tidak Beli
Jauh	Sempit	Rendah	Kontan	Tidak Beli
Jauh	Luas	Rendah	Kredit	Tidak Beli
Sedang	Sempit	Rendah	Kontan	Beli
Jauh	Sempit	Tinggi	Kontan	Tidak Beli
Sedang	Luas	Tinggi	Kontan	Tidak Beli
Dekat	Luas	Rendah	Kredit	Beli
Dekat	Sempit	Rendah	Kontan	Tidak Beli
Sedang	Luas	Rendah	Kontan	Beli
Dekat	Sempit	Rendah	Kredit	Tidak Beli
Sedang	Luas	Rendah	Kontan	Beli
Dekat	Luas	Rendah	Kontan	Beli
Dekat	Luas	Tinggi	Kontan	Tidak Beli

Kemudian akan dilakukan perhitungan nilai *entropy* dan *gain* masing-masing atribut, dapat dilihat pada tabel dibawah ini.

Tabel 2.Perhitungan *Entropy* dan *Gain Node 1*

Klasifikasi Penilaian	Jumlah Kasus	Beli S1	Tidak Beli S2	Entropy	Gain
Total	14	6	8	0,985234	
Aksesibilitas					
Jauh	J	3	0	3	0,2535
Sedang	S	4	3	1	0,811278
Dekat	D	7	3	4	0,999783
Tipologi					
Luas	L	9	5	4	1,005576 0,0809
Sempit	S	5	1	4	0,721928
Harga					
Tinggi	T	4	0	4	0,2916

Rendah	R	10	6	4	0,970951
Pembayaran					
Kontan	KN	11	5	6	1,00853 -0,0070
Kredit	K	3	1	2	0,932795

Dari tabel diatas dilihat nilai gain yang tertinggi ada pada attribute Harga dengan nilai 0,291698 maka harga akan menjadi sebuah akar pertama, kemudian akan dilanjutkan perhitungan *node* ke 2 dengan perhitungan pada tabel dibawah ini.

Tabel 3.Perhitungan *Entropy* dan *Gain Node 2*

Klasifikasi Penilaian	Jumlah Kasus	Beli S1	Tidak Beli S2	Entropy	Gain
Total	10	6	4	0,9709	
Aksesibilitas					
Jauh	J	2	0	2	0,48547
Sedang	S	3	3	0	0
Dekat	D	5	3	2	0,9709
Tipologi					
Luas	L	6	5	1	0,6646 0,24764
Sempit	S	4	1	3	0,8112
Pembayaran					
Kontan	KN	7	5	2	0,8776 0,081092
Kredit	K	3	1	2	0,9183

Tabel diatas menunjukkan perhitungan *node 2* dengan nilai gain tertinggi adalah pada atribut aksesibilitas yaitu 0,4854753, maka aksesibilitas akan menjadi akar selanjutnya, karena aksesibilitas dekat belum ditentukan maka perhitungan akan dilanjut ke *node 3* dibawah ini.

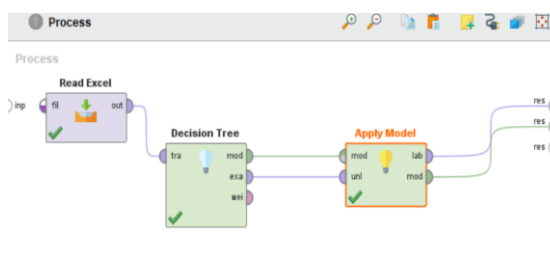
Tabel 4.Perhitungan *Entropy* dan *Gain Node 3*

Klasifikasi Penilaian	Jumlah Kasus	Beli S1	Tidak Beli S2	Entropy	Gain
Total	5	3	2	0,97095	
Tipologi					
Luas	L	3	3	0	0,97095
Sempit	S	2	0	2	0
Pembayaran					
Kontan	KN	3	2	1	0,57095
Kredit	K	2	1	1	

Tabel *node* 1.3 merupakan perhitungan terakhir dari teknik datamining algoritma C4.5 karena telah ditemukan klasifikasi akar terakhir dari pohon keputusan yaitu attribute Tipologi dengan nilai *gain* 0,97095.

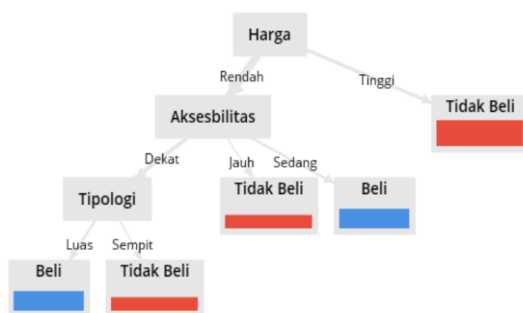
4.2 Pembahasan

Langkah berikutnya yang akan dilakukan yaitu melihat hasil pengujian dengan menggunakan tools data mining rapidminer, dimana nantinya proses akan dimulai dari koneksi basisdata , operator rapidminer dan pohon keputusan yang dihasilkan, disini penulis menggunakan rapidminer studio, untuk lebih jelasnya dapat dilihat pada gambar berikut.



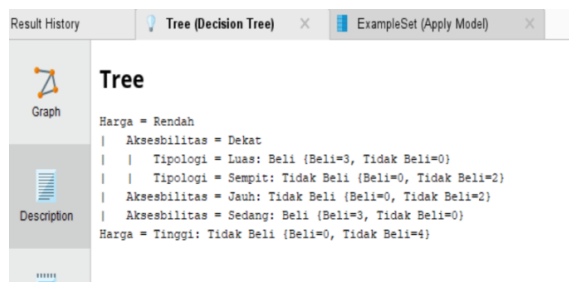
Gambar 2. Koneksi Proses Rapidminer

Proses ini dilanjutkan dengan menjalankan (run) dan dapat kita lihat hasil pohon keputusan yang dibentuk dan sama dengan hasil perhitungan manual yang telah dilakukan diatas.



Gambar 3. Pohon Keputusan

Dari pohon keputusan diatas maka pada *rapidminer* di dapat diperoleh *rule* aturan Sebagai berikut:



Gambar 4. Rules Yang Didapat

Aturan-aturan yang didapat dari hasil pohon keputusan dan deskripsi *tree* diatas untuk menganalisa faktor pemilihan perumahan berdasarkan atribut pada penelitian ini adalah:

1. Jika Harga = Tinggi maka konsumen = Tidak Beli.
2. Jika Harga = Rendah, Aksesibilitas =Jauh, maka konsumen = Tidak Beli.
3. Jika Harga = Rendah, Aksesibilitas = Sedang, Maka Konsumen = Beli
4. Jika Harga = Rendah, Aksesibilitas = Dekat, Tipologi=Luas, Maka Konsumen = Beli
5. Jika Harga = Rendah, Aksesibilitas = Dekat, Tipologi = Sempit, maka Konsumen = Tidak Beli.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Hasil penelitian yang telah dilakukan untuk menganalisa faktor pemilihan perumahan dengan teknik klasifikasi C4.5 maka dapat disimpulkan faktor yang menjadi penyebab utama masyarakat memilih untuk bertempat tinggal didaerah perumahan cipta asri group adalah Harga Perumahan disini, aksesibilitas yang dekat dari jalan utama,sekolah, rumah sakit dan pusat perbelanjaan dan terakhir adalah tipologi atau keluasan kawasan perumahan baik dari jalan,kelebihan tanah dan tersedia pula kawasan untuk fasilitas umum.

5.2 Saran

Adapun saran untuk kedepanya perlu di analisis dengan jumlah yang lebih besar dan variabel yang lebih banyak agar dapat menyempurnakan hasil klasifikasi,dan untuk hasil yang lebih maksimal bisa di kerjakan dengan jenis algoritma lain pada penelitian berikutnya supaya dapat melihat perbandingan hasil yang diperoleh.

Ucapan Terima Kasih

Pertama-tama penulis mengucapkan terimakasih kepada ristekdikti yang telah memberikan dana hibah agar terlaksanaknyA penelitian ini, kemudian kepada Universitas Batam dan LPPM yang telah bersedia menjadi wadah bagi peneliti untuk mengembangkan ilmu kami sebagai peneliti terimakasih atas segala bantuannya selama ini semoga kebaikan ini dibalas oleh Tuhan YME.

Daftar Pustaka

Bimo, P., Setio, N., Retno, D., Saputro, S., & Winarno, B. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.

Ente, D. R., Thamrin, S. A., Arifin, S., Kuswanto, H.,

- & Andreza, A. (2020). Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus Di Rumah Sakit Unhas Menggunakan Algoritma C4.5. *Indonesian Journal of Statistics and Its Applications*, 4(1), 80–88. <https://doi.org/10.29244/ijsa.v4i1.330>
- Erlin Elisa, R. H. (2019). ALGORITMA C4.5 UNTUK KLASIFIKASI PEMILIHAN CALON KETUA RW Erlin. *STMIK Kota Bumi*, 7(2), 87–96.
- Halim, C., & Prasetyo, H. (2018). P Enerapan a Rtficial I Ntelligence Dalam C Omputer a Ided I Nstructure (Cai). *Penerapan AI Dalam Computer Aided Instructure*, 01(01), 45–51.
- Harahap, P. N., & Sulindawaty, S. (2020). Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah). *Matics*, 11(2), 46. <https://doi.org/10.18860/mat.v11i2.7821>
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama Vol.*, 11(2), 130–138.
- Kalesaran, R., Mandagi, R., & Waney, E. (2013). Analisa Faktor-Faktor Yang Mempengaruhi Keputusan Konsumen Dalam Pemilihan Lokasi Perumahan Di Kota Manado. *Jurnal Ilmiah Media Engineering*, 3(3), 99064.
- Luluk Elvitaria, M. H. (2017). Smk Analisis Kesehatan Abdurrah Menggunakan Algoritma. (*Jurnal Teknologi Dan Sistem Informasi Univrab*, 2(2), 220–233.
- Nasrullah, A. H. (2018). Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, 10(2), 244–250. <https://doi.org/10.33096/ilkom.v10i2.300.244-250>
- Tukino, T. (2019). Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia. *Jurnal Sistem Informasi Bisnis*, 9(1), 39. <https://doi.org/10.21456/vol9iss1pp39-46>