



## Computer Based Information System Journal

ISSN (Print): 2337-8794 | E- ISSN : 2621-5292  
 web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



# ALGORITMA CLASSIFIER UNTUK MENGANALISIS FAKTOR PEMILIHAN HUNIAN TEMPAT TINGGAL

Erlin Elisa<sup>1)</sup>, Alfannisa Anurullah Fajrin<sup>2)</sup>

Universitas Putera Batam, Indonesia.

### INFORMASI ARTIKEL

Diterima Redaksi: Juli 2020  
 Diterbitkan Online: September 2020

### KATA KUNCI

Hunian Tempat Tinggal; *Datamining*:  
 Algoritma C4.5, *Dessicion Tree*.

### KORESPONDENSI

E-mail:  
 Erlin.Elisa@puterabatam.ac.id

### A B S T R A C T

*Batam City is the largest city in the RIAU archipelago, the population is also included in the island with the largest population when compared to other islands. According to data from the Central Statistics Agency (BPS), Batam City recorded the projected population growth in Batam in 2019 as many as 1,376,009 people. The development of the increasing population will gradually start to cause new problems, namely the problem of housing or housing. When viewed from the incident so far, the factors or reasons for choosing a residential place to live are not understood by the community so that when they buy, problems often occur, such as not matching the price or problems with developers such as difficult mortgage management and an unsuitable environment and many problems. other. This study will analyze the choice of residential housing with the C4.5 datamining technique and rapid mining as data testing tools, the results of this study show that the main factors for housing selection are clean water sanitation, accessibility, accessibility and price after testing the accuracy of C4.5 data. 25 sample data obtained 88.00% accuracy of the performance of this algorithm.*

## I. Latar Belakang

Kota Batam merupakan kota yang terbesar di wilayah kepulauan RIAU, jumlah penduduk pun termasuk kepada pulau dengan jumlah penduduk yang paling besar jika dibandingkan dengan pulau-pulau lain, hal ini tentu bukan hanya besar dikertas (secara data) saja karena secara nyata bisa dilihat dari kehidupan dan keseharian yang ada. Menurut data 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 Untuk menanggulangi hal tersebut pemerintah kota Batam terus berbenah hal ini dibuktikan dengan banyaknya perbaikan dari segi jalan-jalan yang ada, jalan-jalan yang lama diperbesar dan diperbaiki menjadi lebih baik, perbaikan dan pembangunan jalan-jalan ini pada saat sekarang bisa dilihat diseluruh penjuru Kota Batam terutama yang menjadi objek-objek vital pada Kota Batam itu sendiri.

Perkembangan penduduk yang semakin banyak tersebut secara peralihan akan mulai menimbulkan permasalahan baru yaitu masalah hunian atau rumah tempat tinggal, dengan terbatasnya lahan yang disediakan Badan Pengusahaan Batam dan Pemerintah Kota Batam, maka warga Kota Batam harus benar-benar pintar mencari hunian dalam hal ini disesuaikan dengan kemampuan keuangan mereka, walaupun hunian ini merupakan kebutuhan primer dan siapa yang tidak membutuhkan rumah demi kenyamanan hidup baik mereka yang masih *single* ataupun berkeluarga, jika dilihat fenomena yang terjadi dikota tentang masalah hunian memang sangat menarik hal ini dapat dilihat pada beberapa persen dari penduduk kota Batam yang ditinggal di RULI, istilah ini memang tidak asing bagi warga Kota Batam yang berarti adalah rumah liar, artinya penduduk tinggal dan menempati tanah pemerintah secara ilegal dan suatu saat harus siap digusur oleh pemerintah jika tanah tersebut harus dipakai untuk keperluan tertentu ataupun untuk keperluan tata kota, tapi jika dihitung dalam persentase jumlah warga yang menempati RULI sangat sedikit, karena lebih banyak dari penduduk menempati rumah hunian dalam bentuk perumahan resmi yang dibangun oleh developer karena banyak dari warga kota Batam mencari aman. Apabila dilihat dari kejadian selama ini faktor atau alasan seseorang tentang memilih hunian tempat tinggal kurang dipahami masyarakat sehingga pada saat mereka membeli sering terjadi masalah, seperti kurang cocok dengan harga atau permasalahan dengan developer seperti pengurusan KPR yang sulit serta lingkungan yang tidak cocok dan masih banyak masalah lainnya.

Berdasarkan kepada permasalahan yang telah dikemukakan diatas maka peneliti akan mencoba mencari solusinya dalam sebuah penelitian dengan judul "Algoritma *Classifier* Untuk Menganalisis Faktor Pemilihan Hunian Tempat Tinggal". Sebagaimana yang diketahui

<http://ejournal.upbatam.ac.id/index.php/cbis>

bahwa algoritma C4.5 merupakan salah satu dari metode yang dikemukakan dalam bidang ilmu *datamining* yang masuk kedalam kelompok klasifikasi dimana dapat menggambarkan hasil dalam bentuk pohon keputusan yang mudah dibaca dan diinterpretasikan oleh pengguna atau *user* yang membutuhkan [1].

## II. Kajian Literatur

### 2.1 Perumahan Atau Hunian

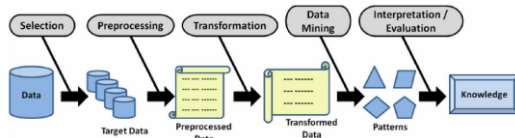
Tempat tinggal adalah kebutuhan pokok setiap orang, oleh karena itu demikian setiap orang akan berlomba untuk memenuhi kebutuhan tersebut baik itu dibeli atau disewa terlebih dahulu, pada sebuah penelitian yang mengkaji tentang bagaimana strategi pembangunan kawasan perumahan dikawasan perkotaan mengatakan dalam teori penelitiannya bahwa perumahan yang dikutip dalam *The Dictionary of Real Estate Appraisal* (2002:313) merupakan lahan atau lahan luas dan kosong yang nanti akan disediakan untuk tempat kediaman, yang berupa rumah tapak, apartemen, rumah susun [2].

Sedangkan pada teori yang telah digunakan juga pada penelitian mengenai strategi pemukiman yang sehat mengatakan Perumahan adalah hunian yang memiliki manfaat sebagai lingkungan tempat tinggal atau lingkungan hunian dengan fasilitas umum yang tersedia didalamnya, yang bertujuan agar hunian tempat tinggal berubah menjadi lingkungan dengan suasana yang sehat, aman, serasi, dan teratur kemudian akan berguna sesuai *planning* pembangunannya [3].

### 2.2 Knowledge Discovery Database

Penggalan data atau yang sering disebut juga dengan KDD berkali-kali dimanfaatkan dengan cara bergantian dengan tujuan untuk menerangkan proses dari penambahan data dalam suatu tumpukan data dalam jumlah yang sangat besar [4]. Kedua maksud istilah tersebut mempunyai teori yang atau konsep yang berbeda, akan tetapi memiliki keterkaitan satu

dengan yang lainnya, dari semua proses KDD tersebut terdapatlah *datamining*, untuk lebih rincinya bisa kita lihat pada gambar proses dibawah ini.



Gambar 1. Tahapan *Knowledge Discovery In Database*

### 2.3 *Datamining*

Sebuah penelitian yang mengkaji tentang perbandingan tentang algoritma C4.5 dan KNN dalam mengidentifikasi mahasiswa yang memiliki potensi *droup out*, dalam penelitian tersebut peneliti memakai teori *datamining* yang merupakan proses otomatisasi yang bersifat semi dengan memakai perhitungan statistika, aritmatika, *Artificial Intelegency*, dan *machine learning* yang berguna dalam mengurai dan menemukan pengetahuan tersembunyi yang bernilai serta berguna yang didapat dalam *database* berkapasitas besar [5].

Pada karya ilmiah oleh penulis tentang analisa faktor penyebab kecelakaan kerja, pada penelitian ini penulis mengambil teori *datamining* merupakan penambangan pengetahuan baru yang akan dijadikan gagasan berita baru yang diambil dari gunungan data-data berkapasitas banyak atau dikenal dengan *database* [6].

### 2.4 Pohon Keputusan

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara

<http://ejournal.upbatam.ac.id/index.php/cbis>

sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain [7].

### 2.5 Klasifikasi

Menurut Tan et al (2009), klasifikasi adalah proses untuk menyatakan suatu objek ke salah satu kategori yang sudah didefinisikan sebelumnya. Klasifikasi juga bisa didefinisikan sebagai proses pembelajaran fungsi target (model klasifikasi) yg memetakan setiap sekumpulan atribut  $x$  (input) ke salah satu kelas  $y$  yang didefinisikan sebelumnya [8]

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.

### 2.6 Algoritma C4.5

Dalam teknik *classifier* menggunakan C4.5 yang akan membuat sebuah pohon keputusan yang didasari atas *training* data yang disediakan tinjauan pustaka ini dimanfaatkan dalam penelitian terdahulu dalam klasifikasi tingkat keganasan hama pada tanaman padi [9].

Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai berikut: dapat mengatasi *missing value*, dapat mengatasi *continue data*, dan *pruning* [10].

Tahapan dalam membuat sebuah *Decision Tree* dalam algoritma C4.5, yaitu

- Mempersiapkan data training. Data training didapat dari data history yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
- Menghitung akar dari pohon. Akar akan diambil dengan cara menghitung nilai gain

dari masing-masing atribut. Nilai gain yang paling tinggi akan menjadi akar yang pertama. Sebelum menghitung nilai gain dari atribut, terlebih dahulu hitung nilai entropy [10].

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

Di mana :

*S* : himpunan kasus

*A* : fitur

*N* : jumlah partisi *S*

*pi* : proporsi dari *Si* terhadap *S*

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [11].

- Memilih Akar dari Atribut yang ada.
- Membuat cabang pada tiap-tiap nilai
- Membagi kasus dalam bentuk cabang
- Mengulangi Proses sampai nilai sama pada setiap cabang
- Tentukan *gain* paling tinggi.

c) Kemudian rumus mencari *gain* :

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

Di mana :

*S* : Kasus Kelompok

*A* : atribut

*N* : jumlah partisi atribut *A*

*|Si|* : jumlah kasus pada partisi ke-*i*

*|S|* : jumlah kasus dalam *S*

d) Ulangi langkah tahap kedua dan ketiga sampai *record* terpartisi.

e) Proses decision tree akan berhenti ketika:

- Semua record dalam simpul n mendapat kelas yang sama
- Tidak ada atribut di dalam record yang dipartisi lagi
- Tidak ada record di dalam cabang yang kosong

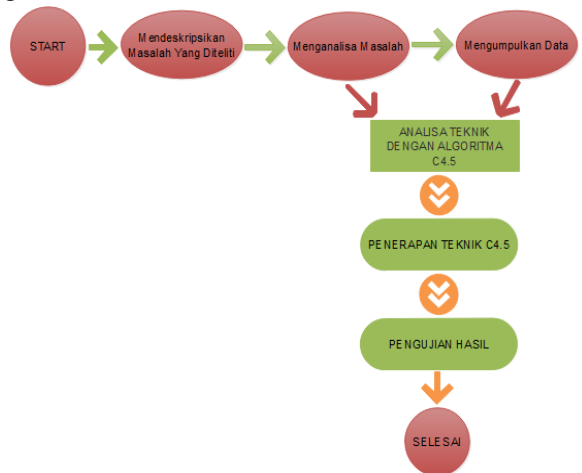
### 2.7 Rapid Miner

*Rapid Miner* merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology 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 [12].

### III. Metodologi

Pedoman atau panduan dan untuk mencapai tujuan akan dituangkan dalam desain yang akan digunakan oleh peneliti dapat dilihat pada gambar dibawah ini.



Gambar 2. Desain Penelitian

### IV. Pembahasan

Pada penelitian ini akan mengkaji faktor pemilihan hunian berdasarkan penyebab seseorang atau masyarakat memutuskan untuk memilih tempat tinggal, variable yang digunakan adalah Harga hunian, Sanitasi Air Bersih, Keamanan lingkungan hunian, Fasilitas yang tersedia di sekitar hunian tempat tinggal dan aksesibilitas ketempat hunian dan tempat umum lainnya seperti, pasar, sekolah dan rumah sakit.

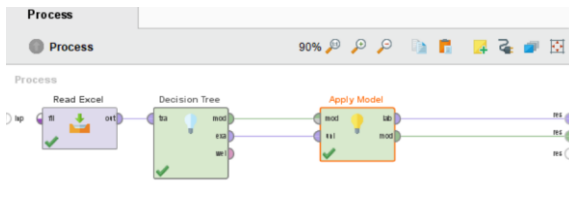
Berikut ini adalah sampel data yang berupa tabel yang kemudian akan ekstraksi melalui aplikasi Rapidminer studio untuk mendapatkan *New Knowledge* berupa *rule* dalam bentuk kondisi *If Then*.

Tabel 1. Sampel Data Tes

Harga	Sanitasi Air Bersih	Keamanan	Fasilitas	Aksesibilitas	Tingkat Pilihan
Tinggi	Layak	Cukup	Layak	Jauh	Tidak
Normal	Tidak Layak	Layak	Layak	Dekat	Tidak
Normal	Layak	Cukup	Layak	Sedang	Ya
Tinggi	Tidak Layak	Layak	Tidak Layak	Jauh	Tidak
Normal	Layak	Layak	Layak	Sedang	Ya
.....	.....	.....	.....	.....	.....
Normal	Tidak Layak	Cukup	Layak	Sedang	Tidak
Tinggi	Layak	Tidak Layak	Tidak Layak	Dekat	Tidak
Normal	Tidak Layak	Tidak Layak	Layak	Dekat	Tidak
Normal	Layak	Layak	Layak	Sedang	Ya

Pada tabel diatas yang meihatkan attribute keputusan adalah tingkat Pilihan, sedangkan attribute kondisi nya adalah Harga, Sanitasi Air Bersih, Keamanan, Fasilitas dan Aksesibilitas.

Selanjutnya akan dilanjutkan dengan proses pengujian dengan *rapidminer studio*, dapat dilihat pada gambar dibawah ini.



Gambar 3. Koneksi Proses *Rapidminer*

Dari proses koneksi *rapidminer* diatas dihasilkan pohon keputusan sebagai berikut.



Gambar 4. Pohon Keputusan

Pohon keputusan diatas menghasilkan empat node akar dengan urutan faktor yang mempengaruhi seseorang untuk memilih hunian tempat tinggalnya adalah Sanitasi Air Bersih, Fasilitas, Aksesibilitas dan harga, untuk lebih jelasnya dapat dilihat dari *description tree* dibawah ini.

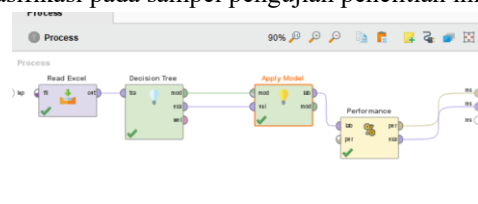
**Tree**

```

Sanitasi Air Bersih = Layak
| Fasilitas = Layak
| | Aksesibilitas = Dekat
| | | Harga = Normal: Ya (Tidak=0, Ya=2, Tidak =0)
| | | Harga = Tinggi: Tidak (Tidak=1, Ya=1, Tidak =0)
| | | Aksesibilitas = Jauh: Tidak (Tidak=2, Ya=1, Tidak =0)
| | | Aksesibilitas = Sedang: Ya (Tidak=0, Ya=5, Tidak =0)
| | Fasilitas = Tidak Layak: Tidak (Tidak=2, Ya=0, Tidak =0)
| Sanitasi Air Bersih = Tidak Layak: Tidak (Tidak=10, Ya=0, Tidak =1)
    
```

Gambar 5. *Description Tree*

Selanjutnya akan melihat tingkat akurasi dari klasifikasi pada sampel pengujian penelitian ini.



Gambar 6. Koneksi Pengujian Performance Akurasi

Hasil pengujian diatas menunjukkan tingkat akurasi dengan teknik algoritma C4.5 adalah 88.00% dari data 25 data uji, lebih jelasnya dapat dilihat pada *result* akurasi *rapidminer* dibawah ini.

accuracy: 88.00%

	true Tidak	true Ya	true Tidak	class precision
pred. Tidak	15	2	1	83.33%
pred. Ya	0	7	0	100.00%
pred. Tidak	0	0	0	0.00%
class recall	100.00%	77.78%	0.00%	

Gambar 6. Akurasi Kinerja Algoritma C4.5

## V. Kesimpulan

Dari proses kinerja datamining teknik klasifikasi algoritma C4.5 dengan *tool rapidminer* dalam menentukan faktor pemilihan hunian tempat tinggal didapatkan hasil:

1. Faktor yang mendasari seseorang memilih hunian tempat tinggal adalah sanitasi air bersih yang disediakan, fasilitas yang terdapat pada lingkungan perumahan termasuk fasilitas dalam mendapatkan kredit KPR, Aksesibilitas dari dan ketempat hunian termasuk pula akses tempat-tempat umum seperti sekolah, pasar, dan rumah sakit dan yang terakhir adalah harga.
2. Tingkat akurasi dari pengujian data sampel dari hasil pengumpulan data dilapangan menunjukkan kinerja Algoritma C4.5 senilai 88.00 % .

## Ucapan Terima Kasih

Terimakasih kami ucapkan kepada tim yang telah membantu penelitian ini dan Ristekdikti sebagai pemberi hibah serta tak lupa pula kepada Universitas Putera Batam dan LPPM yang telah mendukung kami dalam penyelesaian penelitian ini.

## Daftar Pustaka

- [1] I. P. Astuti, "Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5," *Fountain Informatics J.*, vol. 2, no. 2, p. 5, 2017, doi: 10.21111/fij.v2i2.1067.
- [2] Totok Harjanto, "Strategi pembangunan perumahan di kawasan perkotaan," vol. 5, no. 2, pp. 44–70, 2016.
- [3] A. Hariyanto, "Strategi Penanganan Kawasan Kumuh Sebagai Upaya Menciptakan Lingkungan Perumahan dan Pemukiman Yang Sehat," no. 4, pp. 11–37, 2013.
- [4] Y. S. Luvia, A. P. Windarto, S. Solikhun, <http://ejournal.upbatam.ac.id/index.php/cbis>
- and D. Hartama, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 1, no. 1, p. 75, 2018, doi: 10.30645/jurasik.v1i1.12.
- [5] A. S. Yeyen Dwi Atma, "Perbandingan algoritma c4.5 dan k-nn dalam identifikasi mahasiswa berpotensi drop out," *Metik J.*, vol. 2, no. 2, pp. 31–37, 2018.
- [6] 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.
- [7] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," vol. VI, no. 1, pp. 15–20, 2014.
- [8] T. Tukino, "Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia," *J. Sist. Inf. Bisnis*, vol. 9, no. 1, p. 39, 2019, doi: 10.21456/vol9iss1pp39-46.
- [9] P. A. Sularno, "Penerapan Algoritma C4.5 Untuk Klasifikasi Tingkat Keganasan Hama Pada Tanaman Padi," *J. Sains Dan Inform.*, vol. 2, pp. 54–60, 2016.
- [10] A. Ulfa, D. Winarso, and E. Arribe, "Sistem Rekomendasi Jurusan Kuliah Bagi Calon Mahasiswa Baru Menggunakan Algoritma C4.5," *J. Fasilkom*, vol. 10, no. 1, pp. 61–65, 2020, doi: 10.37859/jf.v10i1.1511.
- [11] S. Lestari and A. Suryadi, "Model Klasifikasi Kinerja Dan Seleksi dosen Berprestasi Dengan," *Proseding Semin. Bisnis Teknol.*, pp. 15–16, 2014.
- [12] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama Vol.*, vol. 11, no. 2, pp. 130–138, 2015.

