

# ANALISIS DATA MINING UNTUK MENENTUKAN VARIABEL – VARIABEL YANG MEMPENGARUHI KELAYAKAN KREDIT KEPEMILIKAN RUMAH MENGUNAKAN TEKNIK KLASIFIKASI

**Aulia Fitrul Hadi**

Universitas Putra Indonesia “YPTK” Padang

E-mail: fitrulhadi@upiypk.ac.id

---

## Abstrak

**Abstrak** - Algoritma C4.5 adalah algoritma klasifikasi data bertipe pohon keputusan. Pohon keputusan algoritma C4.5 dibangun dengan beberapa tahap yang meliputi pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai dan membagi kasus dalam cabang. Tahap-tahap ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Dari penyelesaian pohon keputusan maka akan didapatkan beberapa rule suatu kasus. Dalam hal ini penulis mengklasifikasikan pengajuan dari calon debitur baru dalam pengajuan kredit kepemilikan rumah. Analisa dilakukan untuk menemukan variabel yang paling berpengaruh dalam pengajuan kredit kepemilikan rumah.

**Kata Kunci** : Algoritma, C4.5, Kredit, Analisa, Perancangan

---

## 1. Pendahuluan

*Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006). Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* (Basuki dan Syarif, 2003).

Analisa terhadap variabel variabel yang mempengaruhi kredit kepemilikan rumah belum menjadi perhatian dari analis kredit maupun pihak yang berkaitan dalam proses kredit tersebut, sedangkan diperlukan analisa yang kuat untuk membantu suatu proses pengajuan kelayakan kredit dapat diterima. Untuk itu analisa dilakukan untuk mendapatkan variabel variabel yang mempengaruhi pengajuan tersebut.

## 2. Landasan Teori

### 2.1 *Knowledge Discovery in Databases (KDD)*

Data Mining sesungguhnya merupakan salah satu rangkaian dari proses pencarian pengetahuan pada database (*Knowledge Discovery in Database/KDD*). KDD adalah pencabutan yang trivial informasi implisit, yang sebelumnya tidak diketahui, dan berpotensi berguna dari data (. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. KDD adalah analisis eksplorasi secara otomatis dan pemodelan *repository* data yang besar. KDD adalah proses terorganisir untuk mengidentifikasi pola dalam data yang besar dan kompleks dimana pola data tersebut ditemukan yang bersifat sah, baru, dan dapat bermanfaat (Anik Andriani, 2012).

Sebuah model proses set tugas yang harus dilakukan untuk mengembangkan elemen tertentu, serta unsur-unsur yang diproduksi di setiap tugas (*output*) dan unsur-unsur yang diperlukan untuk melakukan tugas (*input*) (Óscar Marbán, Gonzalo Mariscal dan Javier Segovia, 2009). Fase dijelaskan seperti berikut (Óscar Marbán, Gonzalo Mariscal dan Javier Segovia, 2009) :

- *Business understanding*  
Fase ini berfokus pada pemahaman tujuan dan kebutuhan proyek dari perspektif bisnis, kemudian mengubah pengetahuan ini menjadi sebuah definisi masalah DM dan rencana awal yang dirancang untuk mencapai tujuan.
- *Data Understanding*  
Data fase pemahaman dimulai dengan pengumpulan data awal dan hasil kegiatan dalam rangka untuk mendapatkan akrab dengan data, untuk mengidentifikasi masalah kualitas data, untuk menemukan wawasan pertama ke dalam data atau untuk mendeteksi subset yang menarik untuk membentuk hipotesis untuk informasi yang tersembunyi.
- *Data Preparation*  
Tahap persiapan data mencakup semua kegiatan yang diperlukan untuk membangun dataset akhir dari data mentah awal. Tugas persiapan data kemungkinan akan dilakukan berulang-ulang dan tidak dalam urutan yang ditentukan.
- *Modeling*  
Pada tahap ini, berbagai teknik pemodelan yang dipilih dan diterapkan dan parameter mereka dikalibrasi dengan nilai-nilai yang optimal. Biasanya, ada beberapa teknik yang sama jenis masalah DM. Beberapa teknik memiliki persyaratan tertentu pada bentuk data. Oleh karena itu, itu sering perlu untuk melangkah kembali ke tahap persiapan data.
- *Evaluation*  
Perspektif analisis data, model berkualitas tinggi tampaknya akan dibangun pada tahap ini proyek. Sebelum melanjutkan penyebaran model akhir, penting untuk mengevaluasi model lebih teliti dan meninjau langkah-langkah yang diambil untuk membangun untuk memastikan bahwa itu benar mencapai tujuan bisnis. Pada akhir tahap ini, keputusan harus dicapai tentang cara menggunakan hasil DM.
- *Deployment*  
Tujuan dari model ini adalah untuk meningkatkan pengetahuan dari data, pengetahuan yang didapat akan perlu diorganisir dan disajikan dengan cara yang pelanggan dapat menggunakannya

Serangkaian proses dan tahap tersebut dapat dilihat pada gambar 2.1 dibawah, yang mana tahap itu adalah sebagai berikut :

1. Pembersihan data dan integrasi data (*cleaning and integration*)  
Proses ini digunakan untuk membuang data yang tidak konsisten dan bersifat *noise* dari data yang terdapat di berbagai basisdata yang mungkin berbeda format maupun *platform* yang kemudian diintegrasikan dalam satu database *datawarehouse*.
2. Seleksi dan transformasi data (*selection and transformation*)  
Data yang terdapat dalam database *datawarehouse* kemudian direduksi dengan berbagai teknik. Proses reduksi diperlukan untuk mendapatkan hasil yang lebih akurat dan mengurangi waktu komputasi terutama untuk masalah dengan skala besar (*large scale problem*). Beberapa cara seleksi, antara lain:
  - *Sampling*, adalah seleksi subset representatif dari populasi data yang besar.
  - *Denoising*, adalah proses menghilangkan *noise* dari data yang akan ditransformasikan
  - *Feature extraction*, adalah proses membuka spesifikasi data yang signifikan dalam konteks tertentu.
  - *Centering*, mengurangi setiap data dengan rata-rata dari setiap atribut yang ada.

- *Normalisation*, membagi setiap data yang *dicentering* dengan standar deviasi dari atribut bersangkutan.
  - *Scaling*, mengubah data sehingga berada dalam skala tertentu.
3. Penambahan data (*Data Mining*)
- Metode Data Mining pendidikan sering berbeda dari metode yang lebih luas dari Data Mining sastra, di eksplisit memanfaatkan beberapa tingkat hirarki yang berarti dalam data pendidikan. Metode dari literatur psikometri sering terintegrasi dengan metode dari mesin belajar dan Data Mining literatur untuk mencapai tujuan ini (Baker, 2008).
- Data-data yang telah diseleksi dan ditransformasi ditambang dengan berbagai teknik. Proses Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan fungsi-fungsi tertentu. Fungsi atau algoritma dalam Data Mining sangat bervariasi. Pemilihan fungsi atau algoritma yang tepat sangat bergantung pada tujuan dan proses pencarian pengetahuan secara keseluruhan.
4. Evaluasi pola dan presentasi pengetahuan
- Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Langkah terakhir KDD adalah mempresentasikan pengetahuan dalam bentuk yang mudah dipahami oleh pengguna.

## 2.2 Definisi data Mining

*Data Mining* didefinisikan sebagai proses menemukan pola dalam data. Proses ini harus otomatis atau biasanya secara semi-otomatis. Pola yang dihasilkan harus berarti bahwa pola tersebut memberikan beberapa keuntungan. Pola tersebut diidentifikasi, divalidasi, dan digunakan untuk membuat sebuah prediksi (Anik Andriani, 2012). Berdasarkan sudut pandang operasional, Data Mining adalah proses terpadu dari analisis data yang terdiri dari serangkaian kegiatan yang berjalan berdasarkan definisi tujuan yang akan dianalisis, dengan analisis datanya sampai interpretasi dan evaluasi hasil (Anik Andriani, 2012).

Data Mining berguna setiap kali sistem adalah berurusan dengan set data yang besar. Dalam sistem pendidikan, catatan mahasiswa yaitu rincian pendaftaran, kriteria kelayakan saja, bunga saja dan prestasi akademik dapat menjadi pertimbangan penting untuk menganalisis berbagai kecenderungan karena semua sistem sekarang berbasis computer (Agarwal, Pandey, dan Tiwari, 2012).

Secara sederhana *Data Mining* bisa dikatakan sebagai proses menyaring atau “menambang” pengetahuan dari sejumlah data yang besar. Istilah lain untuk Data Mining adalah *Knowledge Discovery in Database* atau KDD. Sebuah lembaga dengan efisien Pergudangan Data dan pendekatan Data Mining dapat mengetahui cara baru untuk meningkatkan perilaku system (Agarwal, Pandey, dan Tiwari, 2012).

### 2.2.1 Teknik-teknik Data Mining

*Data Mining* digunakan untuk menemukan ilmu yang tersembunyi dari pusat perekaman data (Mohamed Hussain Tawarish dan Farooqui Wassmuddin, 2013).

*Data Mining* dijelaskan sebagai proses penemuan pola dalam kuantitas data yang besar dengan berbagai teknik, metode dan *tool* yang beraneka ragam (Kabakchieva, 2012).

*Data Mining* dibagi menjadi beberapa teknik berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain.

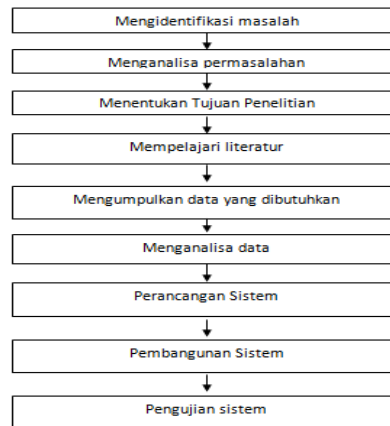
6. Asosiasi

Tugas asosiasi dalam *Data Mining* adalah menemukan *atribut* yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Algoritma-algoritma *Data Mining* kebanyakan berasal atau merupakan bentuk pengembangan dari algoritma-algoritma dari bidang ilmu *machine learning*, statistika, inteligensia buatan dan jaringan saraf tiruan. Karena tidak dirancang untuk menangani data dalam ukuran yang sangat besar, sedangkan *Data Mining* dimaksudkan untuk menangani data dalam ukuran yang demikian, maka salah satu arah penelitian di bidang *Data Mining* adalah mengembangkan algoritma-algoritma tersebut agar dapat menangani data yang berukuran sangat besar.

### 3. Metodologi Penelitian

Metode yang dipakai dalam penelitian yang tercakup dalam kerangka kerja penelitian mulai dari mengidentifikasi masalah, menganalisa masalah, menentukan tujuan, mempelajari *literature*, teknik pengumpulan data, cara mendesain sistem, teknik pengujian sistem sampai pada teknik mengevaluasi sistem. Tahap-tahap kerangka kerja dibuat agar penelitian menjadi terarah dan mencapai tujuan yang ditentukan dalam penelitian ini. Adapun kerangka kerja penelitian ini dapat dilihat pada gambar berikut :



Gambar 3.1 : Kerangka Kerja Penelitian

## 4. Analisa dan Pengolahan Data

### 4.1 Analisa Data Mining dengan Algoritma C4.5

Sistem yang akan dirancang dan digunakan dalam menentukan variabel yang mempengaruhi kredit kepemilikan rumah dirancang dengan menggunakan metode analisis terstruktur, pada metode transformasi input, proses dan *output* dinyatakan dengan Diagram Arus Data (*Data Flow Diagram*) sebagai pernyataan algoritma pelatihan dan pengujian. Pada tahapan ini penggunaan notasi-notasi digunakan untuk menggambarkan arus data sistem, di mana akan sangat membantu dalam proses komunikasi dengan pemakai.

Pada penulisan penelitian ini proses klasifikasi variabel kredit diproses dengan membuat pohon keputusan untuk menghasilkan suatu *output*. Di mana data tersebut mempunyai beberapa kriteria yang merupakan syarat dalam pengklasifikasian variabel kredit. Data tersebut merupakan data *input* klasifikasi variabel kredit kepemilikan rumah. Data yang telah didapat dianalisa sehingga dikelompokkan menjadi beberapa kelompok data yang dapat diproses dengan merancang pohon keputusan.

Proses pengambilan keputusan dalam klasifikasi variabel kredit kepemilikan rumah dikelompokkan menjadi beberapa kriteria yaitu :

1. BI Checking
2. Jumlah Penghasilan
3. Harga Rumah
4. DP Rumah

### 4.2 Perancangan Sistem

Salah satu cara untuk melakukan analisis kredit adalah dengan melakukan klasifikasi dari kumpulan data calon debitur. Salah satu model klasifikasi adalah dengan membuat pohon keputusan.

**Tabel 1: Tabel Rincian Posisi Status Kredit Nasabah**

NO	NAMA NASABAH	BI CHECKING	PENGHASILAN	HARGA RUMAH	DP RUMAH	KEPUTUSAN
1	SHERLY SUALDANI	Ada	Sedang	Medium	Low	Hold
2	AGUSRIO HENDRA	Ada	Sedang	Kecil	Low	Checking
3	ROCKY LADIMO	Ada	Sedang	Kecil	Low	Checking
4	HENDRIK MANER	Tidak	Sedang	Kecil	Low	Approve
5	CATERIN	Tidak	Sedang	Kecil	Low	Approve
6	FANNY MAKMUR	Tidak	Sedang	Medium	Low	Approve
7	RICKY SINATRA	Ada	Sedang	Medium	Low	Hold
8	MUHAMMAD IMAN PRIADI	Tidak	Sedang	Medium	Low	Approve
9	MUSA	Tidak	Tinggi	Besar	High	Approve
10	NOFRIANTI	Ada	Sedang	Medium	Low	Hold
11	ROMINSEN ST	Tidak	Sedang	Medium	Low	Approve
12	BUDHI HARYADI BASRY	Tidak	Sedang	Medium	Low	Approve
13	RIVALDI	Ada	Sedang	Kecil	Low	Checking
14	RAHMAT HIDAYAT	Tidak	Sedang	Medium	Low	Approve
15	WITANTO EFFENDI	Ada	Tinggi	Besar	High	Checking
16	SILVIA ANGGRAINI	Ada	Sedang	Medium	Low	Hold
17	ALEXANDER CANDRA	Ada	Sedang	Besar	Low	Hold
18	LINDA WAHYUNI	Ada	Sedang	Medium	Low	Hold
19	YANCE SYAHRIAL	Ada	Tinggi	Besar	High	Checking
20	JAMALUS	Ada	Sedang	Kecil	Low	Checking
21	HARRY HUJNIR	Ada	Tinggi	Besar	Low	Hold
22	NOVENDRIATI	Ada	Sedang	Medium	Low	Hold
23	SUSI YANTI DELI	Ada	Sedang	Medium	Low	Hold
24	MUHAMMAD DOVI	Ada	Sedang	Besar	Low	Hold

**Gambar 2 : Tabel Kredit**

Dari data-data tersebut, yang diambil sebagai variabel keputusannya adalah nilai kelengkapan data lengkap, dan tidak lengkap. Sedangkan yang diambil sebagai variabel penentu dalam pembentukan pohon keputusan adalah Bi Cheking, Jumlah Penghasilan, DP Rumah dan Harga Rumah.

Menghitung Nilai Entropy total :

*Entropy* (Total)

$$Entropy(total) = \left( -\frac{8}{24} * \log_2 \left( \frac{8}{24} \right) \right) + \left( -\frac{6}{24} * \log_2 \left( \frac{6}{24} \right) \right) + \left( -\frac{10}{24} * \log_2 \left( \frac{10}{24} \right) \right)$$

$$= 1.55458517$$

Gain tertinggi menjadi akar dan dilanjutkan hingga tidak ada lagi kemungkinan yang menjadi akar

Node		Jumlah Kasus (S)	Approve (S1)	Checking (S2)	Hold (S3)	Entropy	Gain
1	TOTAL	24	8	6	10	1.55458517	
	BI Checking						0.9182958
	Ada	16	0	6	10	0.95443400	
	Tidak	8	8	0	0	0	
	Penghasilan						0.0438455
	Sedang	20	7	4	9	1.51288762	
	Tinggi	4	1	2	1	1.5	
	Harga Rumah						0.8350768
	Rendah	6	2	4	0	0.91829583	
	Sedang	12	5	0	7	0.97986876	
	Tinggi	6	1	2	3	1.45914792	
	DP Rumah						0.1328023
	Low	21	7	4	10	1.49370955	
	High	3	1	2	0	0.91829583	

**Gambar 3 : Tabel Node 1**

**Tabel 4.4 Perhitungan Node 1.2**

Node			Jumlah Kasus (S)	Approve (S1)	Checking (S2)	Hold (S3)	Entropy	Gain
2	BI Checking – Ada		16	0	6	10	0.95443400	
	Penghasilan							0.95744855
		Sedang	13	0	4	9	0.8904916	
		Tinggi	3	0	2	1	0.9182958	
	Harga Rumah							1.55458516
		Rendah	4	0	4	0	0	
		Sedang	7	0	0	7	0	
		Tinggi	5	0	2	3	0.97095059	
	DP Rumah							1.051098171
		Low	14	0	4	10	0.86312057	
		High	2	0	2	0	0	

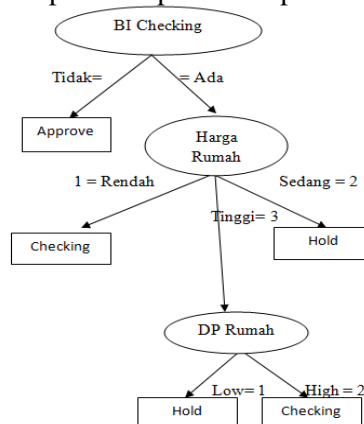
**Gambar 4 : Tabel Node 2**

**Ta**

Node			Jumlah Kasus (S)	Approve (S1)	Checking (S2)	Hold (S3)	Entropy	Gain
3	Harga Rumah - Tinggi		5	0	2	3	0.97095059	
	Penghasilan							0.41997311
		Sedang	2		0	2	0	
		Tinggi	3		2	1	0.9182958	
	DP Rumah							0.97095059
		Low	3		0	3	0	
		High	2		2	0	0	

**Gambar 5 : Tabel Node 3**

Dari ketiga *node* tersebut di dapatkan pohon keputusan seperti berikut :



**Gambar 4.4 Pohon Keputusan Hasil Perhitungan Node 3**

**Gambar 6 : Diagram Pohon**

Berdasarkan pohon keputusan terakhir yang terbentuk pada gambar 4.4 di atas, maka aturan atau *rule* yang terbentuk adalah sebagai berikut :

1. Jika *BI Checking* = T (Tidak), maka Keputusan = *Approve*
2. Jika *BI Checking* = A(Ada) dan Harga Rumah = 1 (Rendah), Maka Keputusan = *Checking*
3. Jika *BI Checking* = A(Ada) dan Harga Rumah = 2 (Sedang), Maka Keputusan = *Hold*
4. Jika *BI Checking* = A(Ada) dan Harga Rumah = 3 (Tinggi) dan Penghasilan = 1 (Sedang) Maka Keputusan = *Hold*
5. Jika *BI Checking* = A(Ada) dan Harga Rumah = 3 (Tinggi) dan Penghasilan = 2 (Tinggi) dan DP Rumah = 1 (*Low*), Maka Keputusan = *Hold*

Jika *BI Checking* = A(Ada) dan Harga Rumah = 3 (Tinggi) dan Penghasilan = 2 (Tinggi) dan DP Rumah = 2 (*High*), Maka Keputusan = *Checking*

## 5. Kesimpulan

Berdasarkan hasil yang didapat dari pembahasan maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Metode pohon keputusan yang diproses dengan Weka lebih efektif dan fleksibel jika digunakan pada proses pengklasifikasian kredit kepemilikan rumah.
2. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang akan digunakan dalam memprediksi juga sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan.
3. Sistem yang dibangun dapat membantu dalam pengklasifikasian kriteria debitur yang umumnya masih dilakukan secara manual.
4. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi eksekutif dan sistem digunakan untuk menggambarkan proses yang terkait dengan pengklasifikasian kriteria debitur.
5. Sistem pengklasifikasian debitur menggunakan algoritma C4.5 dapat digunakan dalam menentukan variabel apa saja yang berpengaruh dalam kredit kepemilikan rumah.

## Daftar Pustaka

- [1]. Anik Andriani. (2012). "*Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Drop Out*"
- [2]. Dorina Kabakchieva. (2012). "*Student Performance Prediction by Using Data Mining Classification Algorithms*"
- [3]. Oscar Marban1, Gonzalo Mariscal2, Javier Segovia1. (2010). "*A Data Mining and Knowledge Discovery Process Model*"
- [4]. Hai-Fen Lin1, Chih-Hung Hsu2, Mao-Jiun J Wang3, Yu-Cheng Lin4. (2008). "*An Application of Data Mining Technique in Developing Sizing System For Army Soldiers in Taiwan*"
- [5]. Mohamed Hussain Tawarish1, Farooqui Waseemuddin2. (2013). "*Analysis of Data Mining Concepts in Higher Education With Needs to Najran University*"
- [6]. Surjeet Kumar Yadav, Saurabh Pal. (2012). "*A Prediction For Performance Improvement of Engineering Students using Classification*"
- [7]. Kabra, Bichkar. (2011). "*Performance Prediction of Engineering Students using Decision Tree*"
- [8]. Padmapriya. (2012). "*Prediction of Higher Education Admissibility using Classification Algorithm*"
- [9]. Nesreen K Ahmed, Jennifer Neville, Ramana Kompella. (2012). "*Network Sampling Designs for Relational Classification*"
- [10]. Sonali Agarwal, Pandey, Tiwari. (2012). "*Data Classification and Decision Tree Approach*"



[Type text]