

## Penerapan *Data Mining* Pengajuan Pembiayaan Perumahan (*Consumen Financing*) Individual Menggunakan *Algoritma C4.5*

Tri wahyudi<sup>[1]</sup>; Mira kusmira<sup>[2]</sup>; Agung Baitul Hikmah<sup>[3]</sup>

Ilmu komputer, STMIK Nusa Mandiri Jakarta<sup>[1][2]</sup>

Jln.Keramat Raya No 18, Senen, Jakarta Pusat

Sistem Informasi, Universitas BinaSarana Informatika<sup>[3]</sup>

Jln. Tanuwijaya No.4 Tawang Tasikmalaya<sup>[3]</sup>

Email : [triwahyudi100390@gmail.com](mailto:triwahyudi100390@gmail.com)<sup>[1]</sup>, [mira.mik@bsi.ac.id](mailto:mira.mik@bsi.ac.id)<sup>[2]</sup>, [agung.abl@bsi.ac.id](mailto:agung.abl@bsi.ac.id)<sup>[3]</sup>

### ABSTRAKSI

Setiap manusia membutuhkan rumah untuk tempat tinggal terutama bagi mereka yang sudah berumah tangga, tetapi dengan harga rumah yang cenderung tiap tahun mengalami kenaikan harga membuat tiap orang mengurungkan niatnya untuk membeli secara tunai. Banyak pihak terutama perbankan atau lembaga keuangan yang menawarkan pembiayaan perumahan secara angsuran, tetapi dalam memberikan pembiayaan ada resiko yang dihadapi oleh pihak perbankan atau lembaga keuangan yaitu resiko kredit macet dan gagal bayar. Oleh karena itu untuk mengatasi permasalahan itu pihak perbankan atau lembaga keuangan perlu memprediksi kelayakan pengajuan pembiayaan terlebih dahulu. Teknik data mining digunakan untuk menentukan prediksi kelayakan pengajuan pembiayaan dalam membuat keputusan kepada nasabah yang berhak diterima atau ditolak pengajuannya. Dan memperoleh tingkat akurasi sebesar 77,27%

**Kata kunci:** Pembiayaan, Data mining, Decision tree, Algoritma C4.5

### ABSTRACT

*Every human being needs a house to live in, especially for those who are already married, but with house prices that tend to increase in price each year makes everyone discourage them from buying in cash. Many parties, especially banks or financial institutions, offer housing payments in installments, but in providing financing there are risks faced by banks or financial institutions, namely the risk of bad credit and default. Therefore, to overcome this problem the banks or financial institutions need to predict the feasibility of applying for financing first. Data mining techniques are used to determine predictions of the feasibility of filing financing in making decisions to customers who are entitled to be accepted or rejected submissions. And obtain an accuracy rate of 77.27%*

**Keywords:** Financing, Data mining, Decision tree, C4.5 Algorithm

### 1. PENDAHULUAN

Pembiayaan konsumen adalah kegiatan pembiayaan untuk pengadaan barang berdasarkan kebutuhan konsumen dengan pembayaran secara angsuran. Pembiayaan konsumen termasuk kedalam jasa keuangan dan dapat dilakukan baik pihak bank atau lembaga keuangan non-bank dalam bentuk perusahaan pembiayaan. Perbankan dan instansi keuangan memiliki peranan yang strategis dalam pembangunan nasional. Badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada

masyarakat dalam bentuk pembiayaan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup masyarakat banyak di bahas pada Undang-Undang Perbankan No.10 tahun 1998. Pada hakikatnya manusia mempunyai kebutuhan akan hidupnya yang harus dipenuhi terutama kebutuhan primer seperti perumahan, kendaraan, pendidikan, modal usaha dan lain-lainnya. Setiap manusia membutuhkan pembiayaan akan kebutuhan hidupnya jika penghasilan yang dimiliki tidak mencukupi untuk memenuhi salah satu dari kebutuhan hidupnya, Dengan semakin

meningkatnya kebutuhan manusia maka keperluan akan modal semakin meningkat pula. Di dalam memberikan pembiayaan tentunya keberadaan jaminan menjadi pertimbangan yang khusus guna merealisasikan suatu pembiayaan kepada masyarakat. Adapun tujuan penjaminan untuk menjamin pembiayaan dari kemungkinan terjadinya pembiayaan macet oleh masyarakat selaku sebagai debitur.

## 2. TINJAUAN PUSTAKA

Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Menurut Ian H. Witten, Frank Eibe, and Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Asma Stephan and Burlington, Eds. United States of America: Morgan Kaufmann, 2011. data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data, data mining memecahkan masalah dengan menganalisis data yang telah ada dalam database. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan meliputi pengumpulan pemakaian data, historis untuk menemukan keteraturan pola atau hubungan dalam set data berukuran besar. Data mining memiliki beberapa proses yang dilakukan untuk membentuk data sebagai berikut

### 2.1. Data Mining

Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Menurut Ian H. Witten, Frank Eibe, and Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Asma Stephan and Burlington, Eds. United States of America:

Morgan Kaufmann, 2011. data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data, data mining memecahkan masalah dengan menganalisis data yang telah ada dalam database. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan meliputi pengumpulan pemakaian data, historis untuk menemukan keteraturan pola atau hubungan dalam set data berukuran besar. Data mining memiliki beberapa proses yang dilakukan untuk membentuk data sebagai berikut:

#### a. Metode Pelatihan

Secara garis besar metode pelatihan yang digunakan dalam teknik-teknik data mining dibedakan menjadi dua pendekatan, yaitu

1. *Unsupervised learning*, metode ini yang diterapkan tanpa adanya latihan (training) dan tanpa ada guru (*teacher*). Yang dimaksud guru disini adalah label dari data lain.
2. *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses *training*.

#### b. Tahap – tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses. Tahap-tahap proses tersebut bersifat interaktif pemakai terlibat langsung atau dengan perantara *knowledge base*.

Tahap-tahap data mining sebagai berikut :

1. Pembersihan data (*data cleaning*)  
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
2. Intergrasi data (*data integration*)  
Intergrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
3. Seleksi data (*data selection*)  
Data yang ada pada database sering kali semuanya tidak semuanya dipakai, oleh

karena itu hanya data yang sesuai di-analisis yang akan diambil dari database.

4. Transformasi data (*data transformation*)  
Data diubah atau digabung ke dalam format yang sesuai untuk di-proses dalam data mining.
5. Proses data (*data process*)  
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *data mining*.
6. Evaluasi pola (*pattern evaluation*)  
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
7. Presentasi pengetahuan (*Knowledge presentation*)  
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

## 2.2. RapidMiner

RapidMiner adalah sebuah *software* untuk pengolahan data mining. 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 *input*, *output*, *data preprocessing* dan *visualisasi*. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang diintegrasikan pada produknya sendiri. RapidMiner ditulis menggunakan bahasa *java* sehingga dapat bekerja disemua sistem operasi.

Beberapa fitur dari RapidMiner sebagai berikut :

- a. Banyaknya *algoritma* data mining, seperti *decision tree*.
- b. Bentuk grafis yang canggih, seperti tumpang tindih *diagram histogram*, *tree chart* dan *3D scatter plots*.
- c. Banyaknya variasi *plugin*, seperti *plugin* untuk melakukan analisis teks.
- d. Menyediakan prosedur data mining dan *machine learning* termasuk *ETL (Extraction,*

*Transformasi, Loading)*, *data preprocessing*, *visualisasi*, *modeling* dan *evaluasi*.

- e. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML dan dibuat dengan GUI.
- f. Mengintegrasikan proyek data mining weka dan statistika R.

## 2.3. Decision Tree (Pohon Keputusan)

Pohon keputusan atau yang biasa lebih dikenal dengan istilah *decision tree* merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah-masalah tersebut dengan memperhitungkan berbagai macam faktor yang berkaitan di dalam lingkup masalah tersebut. Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah sehingga dengan memperhitungkan faktor-faktor tersebut dapat menghasilkan penyelesaian yang terbaik untuk masalah tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternative pemecahan masalah.

Pohon keputusan dalam analisis pemecahan masalah pengambilan keputusan merupakan pemetaan alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon keputusan juga memperlihatkan faktor-faktor kemungkinan yang dapat mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita megambil alternatif keputusan tersebut.

Secara umum, pohon keputusan adalah suatu gambaran permodelan dari suatu persoalan yang terdiri dari serangkaian keputusan yang mengarah kepada solusi yang dihasilka. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori *graf*. Seiring dengan perkembangannya, pohon keputusan kini telah banyak dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

2.4. Algoritma C4.5

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan dibutuhkan algoritma C4.5. Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam data mining :

- a. Pilih atribut sebagai simpul akar.
- b. Buat cabang untuk tiap-tiap nilai.
- c. Bagi kasus tiap cabang.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan atribut sebagai simpul, baik akar (root) atau simpul internal didasarkan pada nilai Gain tertinggi dari atribut yang ada. Perhitungan nilai Gain digunakan rumus seperti dalam persamaan 1.

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

di mana:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S<sub>i</sub>| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Untuk menghitung nilai entropy dapat dilihat pada persamaan 2.

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

di mana:

S : Himpunan kasus

n : Jumlah partisi S

p<sub>i</sub> : Proporsi dari S<sub>i</sub> terhadap S

3. METODOLOGI

Berikut sistematika tahapan dalam mengelolah data pembiayaan sesuai dengan :

- a. Wawancara
 

Pada tahap ini penulis mewawancarai pihak terkait mengenai data nasabah.
- b. Pengumpulan data
 

Dalam tahap ini penulis mencari bahan-bahan dasar yaitu data nasabah yang berkaitan dengan data pembiayaan.
- c. Jurnal dan studi pustaka
 

Tahap ini merupakan landasan awal dalam penelitian ini. Dari studi pustaka ini didapatkan banyak referensi yang akan

mendukung penelitian ini, sekaligus untuk memperkuat pengetahuan dasar dan teori yang digunakan pada penelitian ini.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 25 data pengajuan pembiayaan baik yang diterima, atau ditolak pengajuannya. Semua atribut pada data pengajuan bernilai kategori, seperti terlihat pada tabel 1. Data pembiayaan terdiri dari 19 atribut.

Tabel 1. Atribut dan field data pembiayaan

Variabel	Field	Skor
Cara pendaftaran	Referal	16
	Solicit	14
	Walk in client	10
Usia Nasabah	35 - 45 tahun	12
	45 - 55 tahun	9
	< 35 tahun	6
	> 55 tahun	3
Status Perkawinan dan jumlah anak	Kawin & jlh anak 1-3	10,5
	Kawin & jlh anak 0	7,5
	Kawin & jlh anak > 3	6
	Tidak kawin/janda/duda	4
Pendidikan	S2/S3	16
	S1	12
	D3	8
	SMU	4
Lama menetap	> 4 tahun	15
	3 - 4 tahun	9
	1 - 2 tahun	6
	< 1 tahun	0
Kepemilikan Rumah	Rumah sendiri	15
	Rumah keluarga	7,5
	Rumah dinas	4,5
	Rumah sewa	3
Nilai aset yang dimiliki	> Rp 350 juta	12
	Rp 200 - 350 juta	9
	Rp 100 - 200 juta	6
	S/d Rp 100 juta	3
Rekening bank	BMI/bank syariah lain	15
	Bank Konvensional	12
	Tidak punya rekening	3
Tempat	First tier	28

bekerja	Second tier	24,5
	Lainnya	17,5
Posisi	Snr.Mgr/Owner	20
	Middle Manager	15
	Junior Manager	10
	Staff Biasa	5
Pengalaman kerja	> 10 tahun	20
	5 - 10 tahun	17,5
	2 - 5 tahun	12,5
Mutasi kerja	Belum pernah pindah	20
	1 kali pindah kerja	15
	2 kali pindah kerja	10
	3 - 4 kali pindah kerja	5
Sumber Pendapatan	Fixed income	36
	Fixed & non-fixed income	28
	Non-fixed income	16
Total income per bulan	> Rp 5 juta	40
	Rp 3 - 5 juta	35
	Rp 1 - 3 juta	20
	Rp < 1 juta	5
Rasio angsuran	< 30%	50
terhadap income	30 - 50%	30
	50 - 70%	20
	> 70%	0
Jenis Jaminan	Deposito	20
	Lahan dan bangunan	15
	Kendaraan bermotor	10
	Jaminan perusahaan	5
Rasio terhadap Nilai Jaminan	< 50%	40
	51 - 70%	30
	71 - 80%	20
	> 81%	10
Kepemilikan Jaminan	Milik sendiri	16
	Milik suami atau istri	14
	Akan balik nama	10
Marketability Jaminan	Sangat Marketable	25
	Cukup Marketable	20
	Kurang Marketable	10

Setelah masing-masing memiliki score penilaian maka terdapat satu tabel dimana tabel tersebut merupakan nilai *max-min score* untuk

menentukan layak atau tidak dalam pengajuan pembiayaan.

Tabel 2. Grade

Grade/ Rating	Total Score		Keterangan
	Min	Max	
R-1	425		Diterima
R-2	400	425	
R-3	375	400	
R-4	350	375	
R-5	325	350	
R-6	300	325	Ditolak
R-7	275	300	
R-8	250	275	
R-9	225	250	
R-10	0	225	

Pada tahapan pengujian, data yang sudah ditransformasikan ke dalam bentuk yang lebih sederhana. Data yang digunakan adalah data nasabah yang mengajukan pembiayaan konsumen secara acak yang diambil sebanyak 25 data sampel. Data akan diolah dengan menerapkan *data mining algoritma* C4.5 pada proses pengajuan pembiayaan konsumen. Beberapa *atribut* yang digunakan yaitu terlihat pada tabel 3 (lampiran). Kemudian hitung *entropy* dengan rumus sebagai berikut :

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

Jadi  $Entropy(S) = ((-19 / 25) * \log_2 (19/25)) + ((-6 / 25) * \log_2 (6/25)) = 0,795040279$

Tabel 3. Hasil Perhitungan pada Dataset

Jumlah Kasus	Diterima	Ditolak	entropy
25	19	6	0,795040279

Setelah mendapatkan *entropy* dari keseluruhan kasus seperti pada tabel 4, kemudian hitung *entropy* dan *Gain* pada setiap *atribut*. Untuk menghitung *Gain* setiap *atribut* menggunakan rumus :

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

Dan akan didapatkan hasil seperti terlihat pada tabel 5 (lampiran) dan dalam melakukan pengujian akan menghasilkan nilai akurasi. Hasil akurasi *confusion matrix* pada pengujian 1 sebagai berikut:

	True tolak	True terima
Pred. Ditolak	0	0
Pred. diterima	5	17

Dari pengujian menggunakan rapidminer menghasilkan akurasi sebesar 77,27%  
Berikut adalah perhitungan akurasi :

$$\begin{aligned}
 \text{Akurasi} &= \frac{tp+tn}{tp+tn+fp+fn} \times 100\% \\
 &= \frac{0 + 17}{0 + 17 + 0 + 5} \times 100\% \\
 &= 17 / 32 \times 100\% \\
 &= 0,7727 \times 100 \% \\
 &= 77,27\%
 \end{aligned}$$

## 5. KESIMPULAN

Dari hasil penelitian yang dilakukan, dapat disimpulkan untuk pengajuan pembiayaan perumahan dapat diprediksi dan dievaluasi dengan memanfaatkan *data mining* menggunakan algoritma C4.5 untuk memprediksi ditolak atau diterima dalam pengajuan. Dari pengujian memiliki nilai akurasi sebesar 77,27%

## REFERENSI

- <https://id.wikipedia.org/wiki/Bank>, diakses pada juni 2016
- [https://id.wikipedia.org/Pembiayaan\\_konsumen](https://id.wikipedia.org/Pembiayaan_konsumen), diakses pada juni 2016
- [http://www.academia.edu/7712860/Belajar\\_Data\\_Mining\\_dengan](http://www.academia.edu/7712860/Belajar_Data_Mining_dengan_RapidMiner) RapidMiner, diakses pada juli 2016
- <http://romisatriawahono.net/dm/>, diakses pada agustus 2016
- Ian H. Witten, frank Eibe, and Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed., Asma Stephan and Burlington, Eds. United States of America: Morgan Kaufmann, 2011.