

## KLASIFIKASI STATUS KREDIT NASABAH BMT MENGGUNAKAN ALGORITMA C5.0

<sup>1</sup>Firstiana Istiqomah, <sup>2</sup>Yuliana Susanti, <sup>3</sup>Etik Zukhronah

<sup>1,2,3</sup> Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret  
e-mail: [firstianaistiqomah@gmail.com](mailto:firstianaistiqomah@gmail.com)

### Abstrak

Nasabah kredit *Baitul Maal wa Tamwil* (BMT) terkadang mengalami tunggakan kredit atau berstatus kredit macet. Kredit macet menimbulkan masalah bagi BMT. Penelitian terhadap nasabah berstatus kredit macet perlu dilakukan melalui perilaku mereka. Klasifikasi berdasarkan pendidikan terakhir, jenis pekerjaan, penghasilan, serta jangka peminjaman dapat dijadikan pihak lembaga untuk mengantisipasi terjadinya kredit macet. Teknik klasifikasi menggunakan pohon keputusan dapat digunakan untuk melihat aturan dan mengelompokkan data ke dalam suatu kelas tertentu. Pohon keputusan yang digunakan pada penelitian ini adalah algoritma *Commercial Version 5.0* (C5.0). Algoritma C5.0 menghasilkan pohon keputusan dengan 7 klasifikasi. Ketepatan klasifikasi menghasilkan akurasi sebesar 83.8% dan eror sebesar 16.2%.

**Kata Kunci:** status kredit, klasifikasi, pohon keputusan, C5.0

### Abstract

*Baitul Maal wa Tamwil* (BMT) credit customers sometimes experience credit arrears or bad credit status. Bad loans cause problems for BMT. Research on bad credit customers needs to be done through their behavior. Classification based on recent education, type of work, income, and the duration of the loan can be used as an institution to anticipate the occurrence of bad credit. Classification techniques using decision trees can be used to view rules and group data into a particular class. The decision tree used in this study is the *Commercial Version 5.0* (C5.0) algorithm. The C5.0 algorithm produces a decision tree with 7 classifications. Classification accuracy produces an accuracy of 83.8% and an error of 16.2%.

**Keywords:** credit status, classification, decision tree, C5.0

## PENDAHULUAN

Masyarakat melakukan pinjaman berupa kredit pada lembaga finansial. Lembaga finansial merupakan suatu lembaga yang bergerak dalam bidang keuangan dimana nasabahnya adalah masyarakat dari berbagai lapisan dengan berbagai perilaku (Sucipto, 2015). Salah satu lembaga finansial yang menyediakan layanan pengajuan kredit adalah lembaga keuangan mikro syariah *Baitul Maal wa Tamwil* (BMT) Wira Darma Sukoharjo. BMT Wira Darma Sukoharjo memberikan pinjaman kredit kepada masyarakat sebagai modal usaha. Masyarakat yang telah menjadi nasabah BMT masih mengalami tunggakan kredit atau berstatus kredit macet. Kredit macet membuat BMT mengalami keresahan. Pada penelitian ini perilaku nasabah menjadi fokus untuk melakukan pendekatan dalam mengurangi masalah ini. Peninjauan terhadap nasabah berstatus kredit macet perlu dilakukan melalui perilaku mereka. Klasifikasi berdasarkan pendidikan terakhir, jenis pekerjaan, penghasilan, serta jangka peminjaman dapat dijadikan pihak lembaga untuk mengantisipasi terjadinya kredit macet.

Sucipto (2015) menerapkan algoritma C4.5 pada data calon nasabah yang mengajukan kredit dapat diketahui atau dideteksi bahwa calon nasabah tersebut akan mengalami kemungkinan macet atau lancar. Data yang digunakan untuk mengantisipasi terjadinya kredit macet dapat dikelompokkan ke dalam kelas-kelas tertentu, sehingga teknik klasifikasi menggunakan *decision tree* (pohon keputusan) dapat digunakan pada penelitian ini. Pohon keputusan dapat mempresentasikan *rule* (aturan) dan mengelompokkan data ke dalam suatu kelas tertentu. Menurut Patil *et. al.* (2012), algoritma C5.0 lebih baik daripada algoritma C4.5 dalam hal memori dan akurasi. Oleh karena itu, pohon keputusan yang digunakan pada penelitian ini menggunakan algoritma *Commercial Version 5.0* (C5.0). Formula yang digunakan berupa *information gain* tertinggi untuk mengklasifikasikan data, membuat pohon keputusan, serta membuat aturan klasifikasi dari data nasabah BMT Wira Darma Sukoharjo.

Berdasarkan latar belakang, tujuan dilakukan penelitian ini adalah 1) Menentukan klasifikasi status kredit nasabah menggunakan algoritma C5.0, dan 2) Menentukan tingkat akurasi dan eror dari aturan klasifikasi.

### **Baitul Maal wa Tamwil (BMT)**

Tujuan didirikannya BMT adalah meningkatkan kualitas usaha ekonomi untuk kesejahteraan anggota pada khususnya dan masyarakat pada umumnya. Dalam rangka mencapai tujuannya, BMT difungsikan sebagai *financial intermediary* (perantara keuangan) antara pemilik dana baik sebagai pemodal maupun penyimpan dengan penggunaan dana untuk pengembangan usaha produktif (Ridwan, 2014). Pada konteks ini, penggunaan dana dapat diartikan sebagai kredit. Kredit adalah semua jenis pinjaman yang harus dibayar kembali oleh peminjam sesuai dengan perjanjian yang telah disepakati (Hasibuan, 2015). Namun, pinjaman yang telah dipakai oleh nasabah tidak selalu dikembalikan pada waktu yang telah disepakati. Nasabah yang melakukan tunggakan pengembalian pinjaman dapat dikatakan berstatus macet.

### **Decision tree (pohon keputusan)**

*Decision tree* merupakan pohon keputusan pada teknik klasifikasi untuk mempresentasikan aturan. Teknik pohon keputusan bermanfaat untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara sejumlah variabel prediktor dengan sejumlah variabel respon. *Decision tree* dibentuk melalui pengklasifikasian variabel prediktor dan variabel respon. Kedua variabel tersebut dikenal dengan nama atribut. Atribut digunakan sebagai acuan untuk mendapatkan keputusan berupa *if...then*. Himpunan atribut yang digunakan akan membentuk *record*. Pada proses ini, *record* akan diubah menjadi bentuk pohon keputusan. Berikut adalah tiga *node* (simpul) yang terdapat pada pohon keputusan.

1. *Root node* atau *node* akar, merupakan simpul paling atas yang memiliki cabang lebih dari satu.
2. *Internal node* atau *node* cabang, merupakan simpul yang terdapat satu cabang masuk dan cabang keluar lebih dari satu.
3. *Leaf node* atau *node* terminal, merupakan simpul terakhir yang tidak memiliki cabang keluar.

### **Algoritma C5.0**

Algoritma C5.0 merupakan algoritma klasifikasi yang menghasilkan pohon keputusan ditemukan oleh Ross Quinlan pada tahun 1987. Algoritma ini merupakan penyempurnaan dari algoritma ID3 dan C4.5. Algoritma C5.0 lebih baik daripada algoritma C4.5 dalam hal memori dan akurasi (Patil *et. al.*, 2012). Kecepatan dalam membuat pohon keputusan melalui *software* dinilai sangat cepat dibandingkan algoritma lainnya. Algoritma C5.0 menerapkan *rule based model* (model berbasis aturan) sehingga memudahkan untuk melihat aturan pada pohon keputusan. Selain itu, algoritma C5.0 dapat mengatasi *missing value* atau nilai yang hilang. Hal ini menjadi kelebihan C5.0 yang dinilai lebih unggul dibanding algoritma lainnya. Menurut Elsayad dan Elsalamony (2013), model pada algoritma C5.0 bekerja dengan *split* (pemisahan) sampel berdasarkan atribut yang memiliki *information gain* tertinggi. Menurut Patil *et.al.* (2012) untuk menghitung informasi himpunan kasus pada kelas  $i$  digunakan rumus

$$I_{(S_1, S_2, \dots, S_m)} = - \sum_{i=1}^m p_i \log_2(p_i)$$

dengan  $I_{(S_1, S_2, \dots, S_m)}$  merupakan informasi dari himpunan kasus pada kelas  $i$ ,  $m$  merupakan banyaknya kelas,  $p_i$  merupakan proporsi kelas  $i$  yang dirumuskan sebagai  $p_i = S_i/S$ ,  $S_i$  merupakan jumlah sampel pada kelas  $i$ , dan  $S$  merupakan himpunan kasus. Langkah selanjutnya menghitung informasi himpunan kasus pada kelas  $i$  dan *subset*  $j$  sebagai berikut.

$$I(S_{1j}, S_{2j}, \dots, S_{mj}) = - \sum_{ij=1j}^{mj} p_{ij} \log_2(p_{ij})$$

$I(S_{1j}, S_{2j}, \dots, S_{mj})$  merupakan informasi dari himpunan kasus kelas  $i$  dan *subset*  $j$  dan  $p_{ij}$  merupakan proporsi kelas  $i$  dan *subset*  $j$ . Informasi dari himpunan kasus pada kelas  $i$  dan *subset*  $j$  dapat digunakan untuk menghitung nilai *entropy*, sebagai berikut.

$$E(A) = \sum_{j=1}^y \{(S_{1j} + S_{2j} + \dots + S_{mj})/S\} * I(S_{1j}, S_{2j}, \dots, S_{mj})$$

dengan  $E(A)$  merupakan *entropy* atribut  $A$  dan  $S_{ij}$  merupakan sampel dari kelas  $i$  dan *subset*  $j$  dari atribut  $A$ . Langkah terakhir menghitung *information gain* sebagai pemilihan atribut yang digunakan sebagai *node*.

$$Gain(A) = I_{(S_1, S_2, \dots, S_m)} - E(A) \quad (1)$$

Proses dilakukan sampai subset sampel tidak dapat dilakukan *split*.

### Confusion Matrix

*Confusion matrix* terdiri dari beberapa sel. Setiap sel berisi angka yang menunjukkan banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi (Han dan Kamber, 2006). Adapun tabel *confusion matrix* terdapat pada Tabel 1.

**Tabel 1. Confusion Matrix**

Classification	Classified Matrix	
	Class +	Class -
Class +	TP ( <i>true positive</i> )	FN ( <i>false negative</i> )
Class -	FP ( <i>false positive</i> )	TN ( <i>true negative</i> )

$$Akurasi (\%) = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (2)$$

$$Error (\%) = \frac{FN+FP}{TP+FN+FP+TN} \times 100\% \quad (3)$$

### METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah seluruh data nasabah kredit BMT Wira Darma Sukoharjo tahun 2019 berjumlah 228. Data diolah menggunakan *software* R. Atribut yang digunakan terdiri dari variabel respon dan variabel prediktor seperti yang terdapat pada Tabel 2.

**Tabel 2. Atribut Data Nasabah**

Variabel	Kategori
Y	Status kredit Lancar Macet
X <sub>1</sub>	Pendidikan SMA Diploma Sarjana
X <sub>2</sub>	Pekerjaan PNS Wiraswasta Swasta
X <sub>3</sub>	Pendapatan Rendah (< Rp 1.500.000,00 per bulan) Sedang (Rp 1.500.000,00 – Rp 2.500.000,00 per bulan) Tinggi (Rp 2.500.000,00 – Rp 3.500.000,00 per bulan) Sangat Tinggi (> Rp 3.500.000,00 per bulan)
X <sub>4</sub>	Jangka waktu Pendek (< 1 tahun) Menengah (1-3 tahun) Panjang (> 3 tahun)

Adapun langkah-langkah dalam penelitian ini yaitu, mendeskripsikan data nasabah kredit BMT tahun 2019. Selanjutnya dilakukan perhitungan *information gain* tertinggi untuk memperoleh *node* akar menggunakan persamaan 1. Perhitungan *information gain* dilakukan untuk memperoleh *node* cabang dan *node* terminal. Berikutnya dilakukan perhitungan *confusion matrix* untuk mengetahui prediksi benar dan salah dari klasifikasi menggunakan persamaan 2 dan 3.

### HASIL DAN PEMBAHASAN

Ringkasan jumlah kategori masing-masing atribut yang terdapat pada Tabel 3 digunakan untuk menentukan *node*.

**Tabel 3. Jumlah kategori masing-masing atribut data nasabah**

		Lancar	Macet	Total
<b>Jumlah</b>		183	45	228
<b>Pendidikan</b>	SMA	154	30	184
	Diploma	8	10	18
	Sarjana	21	5	26
<b>Pekerjaan</b>	PNS	4	0	4
	Swasta	85	24	109
	Wiraswasta	94	21	115
<b>Pendapatan</b>	Sedang	136	39	175
	Tinggi	33	5	38
	Sangat Tinggi	14	1	15
<b>Jangka</b>	Pendek	9	1	10
	Menengah	174	44	218

Selanjutnya menghitung nilai *information gain* tertinggi sebagai *node akar*. Tabel 4 menunjukkan hasil perhitungan nilai *information gain* untuk setiap atribut.

**Tabel 4. Nilai *Information Gain* masing-masing atribut**

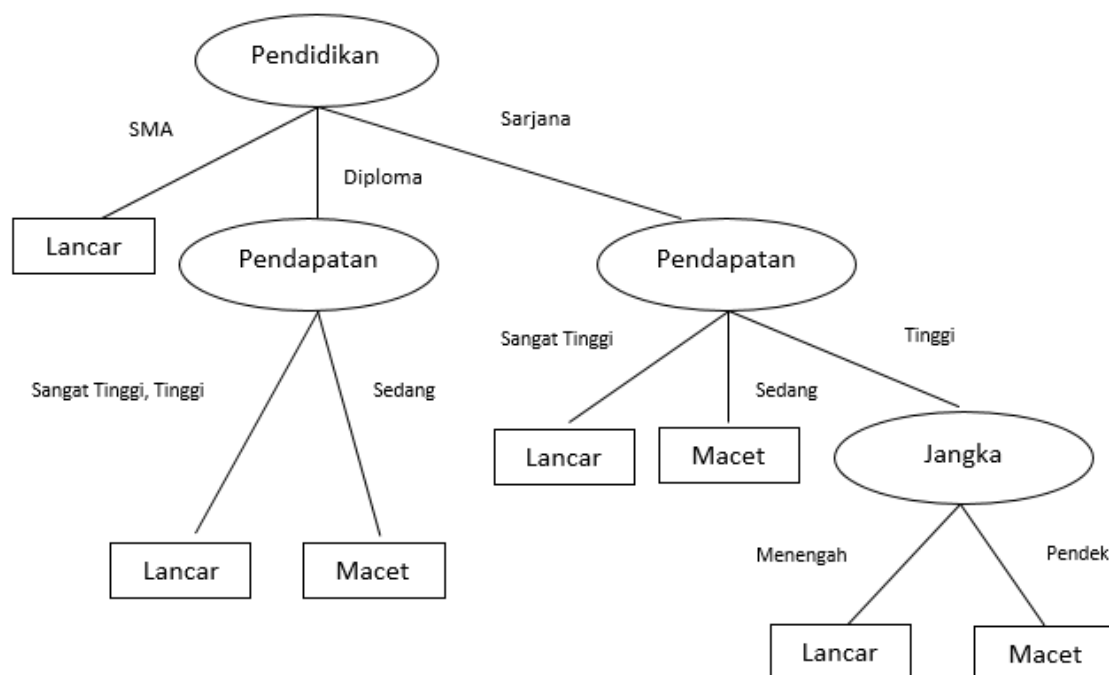
	Variabel	<i>Information Gain</i>
<b>X<sub>1</sub></b>	Pendidikan	<b>0.040116</b>
<b>X<sub>2</sub></b>	Pekerjaan	0.007177
<b>X<sub>3</sub></b>	Pendapatan	0.012321
<b>X<sub>4</sub></b>	Jangka waktu	0.002305

Berdasarkan Tabel 4 *information gain* tertinggi adalah variabel pendidikan. Oleh karena itu, pendidikan dijadikan sebagai *node akar*. *Node akar* dipilih berdasarkan nilai *information gain* tertinggi setelah dilakukan penghapusan terhadap atribut yang menjadi *node akar*. Langkah selanjutnya adalah menghitung *node cabang*. Kategori atribut pendidikan dibagi menjadi tiga kategori, sehingga dilakukan iterasi masing-masing kategori seperti pada Tabel 5.

**Tabel 5. Iterasi kategori**

Iterasi	Kategori	Atribut	<i>Information Gain</i>
1	Pendidikan SMA	Pekerjaan	0.007244
		Pendapatan	0.015354
		Jangka	0.011439
2	Pendidikan Diploma	Pekerjaan	0.011328
		Pendapatan	0.249688
		Jangka	0
3	Pendidikan Diploma   Pendapatan Sedang	Pekerjaan	0.236562
		Jangka	0
4	Pendidikan Diploma   Pendapatan Tinggi	Pekerjaan	0.170951
		Jangka	0
5	Pendidikan Sarjana	Pekerjaan	0.031828
		Pendapatan	0.099446
		Jangka	0.02933
6	Pendidikan Sarjana   Pendapatan Tinggi	Pekerjaan	0.024086
		Jangka	0.160705
7	Pendidikan Sarjana   Pendapatan Sangat Tinggi	Pekerjaan	0.199204
		Jangka	0.025851

Nilai *information gain* pada iterasi mengklasifikasikan semua kasus ke dalam kelas tertentu. Pohon keputusan yang terbentuk berdasarkan *information gain* tertinggi ditunjukkan pada Gambar 1.



**Gambar 1. Pohon keputusan algoritma C5.0**

Berdasarkan Gambar 1 diperoleh klasifikasi pohon keputusan. Berikut klasifikasi pohon keputusan.

1. Pendidikan SMA mengklasifikasikan ke dalam status kredit lancar
2. Pendidikan Diploma, pendapatan sangat tinggi dan tinggi mengklasifikasikan ke dalam status kredit lancar
3. Pendidikan Diploma, pendapatan sedang mengklasifikasikan ke dalam status kredit macet
4. Pendidikan Sarjana, pendapatan sangat tinggi mengklasifikasikan ke dalam status kredit lancar
5. Pendidikan Sarjana, pendapatan sedang mengklasifikasikan ke dalam status kredit macet
6. Pendidikan Sarjana, pendapatan tinggi, jangka waktu menengah mengklasifikasikan ke dalam status kredit lancar
7. Pendidikan Sarjana, pendapatan tinggi, jangka waktu pendek mengklasifikasikan ke dalam status kredit macet

Hasil klasifikasi pohon keputusan dapat menghasilkan prediksi benar dan salah. Oleh karena itu, perlu dilakukan ketepatan klasifikasi menggunakan *confusion matrix* pada Tabel 6.

**Tabel 6. Confusion Matrix**

Klasifikasi	Prediksi Klasifikasi	
	Lancar	Macet
Lancar	180	3
Macet	34	11

Tabel 6 menunjukkan bahwa 191 *record* diklasifikasikan dengan benar yang berarti prediksi sesuai dengan kelas prediksi, sedangkan 37 *record* diklasifikasikan salah. Sejumlah 34 *record* seharusnya diklasifikasikan sebagai status macet namun diprediksi sebagai status lancar dan sejumlah 3 *record* seharusnya diklasifikasikan sebagai status lancar namun diprediksi sebagai status macet. Kesalahan klasifikasi tersebut menyebabkan eror sebesar 16.2%. Namun, prediksi menghasilkan tingkat akurasi yang baik sebesar 83.8%.

## SIMPULAN DAN SARAN

Berdasarkan pembahasan dapat dilihat bahwa atribut yang masuk pada pohon keputusan adalah atribut pendidikan, pendapatan, dan jangka waktu peminjaman. Pohon keputusan menghasilkan 7 klasifikasi. Klasifikasi tersebut menghasilkan keakuratan sebesar 83.8% serta eror sebesar 16.2%. Didapatkan kesimpulan bahwa klasifikasi menggunakan algoritma C5.0 menghasilkan akurasi yang cukup tinggi sehingga dapat digunakan sebagai alternatif pihak pemangku kebijakan BMT dalam melihat klasifikasi kredit nasabah bersatatus lancar atau macet. Penelitian selanjutnya dapat ditambah beberapa variabel mengenai perilaku nasabah karena jika semakin banyak variabel yang digunakan untuk pengujian, maka hasil klasifikasi semakin akurat.

## DAFTAR PUSTAKA

- Elsayad, A.M., & Elsalamony, H.A. (2013). Diagnosis Breast Cancer using Decision Tree Models and SVM. *International Journal of Computer Applications*, Vol. 83, No. 5.
- Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehnique*. San Fransisco: Morgan Kauffman Publishers.
- Hasibuan, H.M.S.P. (2015). *Dasar-Dasar Perbankan*. Jakarta: PT Bumi Aksara.
- Patil, N., Lathi, R., & Chitre, V. (2012). Customer Card Classification Based on C5.0 and CART Algorithm. *International Journal of Engineering Reasearch and Applications*, 12(4), 164-167.
- Ridwan, M. (2014). *Manajemen Baitul Maal Wa Tamwil (BMT)*. Yogyakarta: UII Press.
- Sucipto, A. (2015). Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam dengan Menggunakan Metode Algoritm Klasifikasi C4.5. *Jurnal DISPROTEK*, Vol. 6, No. 1.