

Implementasi Algoritma *Decision Tree* (J.48) untuk Memprediksi Resiko Kredit pada BMT

Atik Febriani^{*1}, Violita Anggraini²

^{1,2}Program Studi Teknik Industri, Institut Teknologi Telkom Purwokerto
Jalan D.I Panjaitan No 128, Purwokerto Kidul, Karangreja, Purwokerto, Jawa Tengah 53147
e-mail: ^{*1}Atik@ittelkom-pwt.ac.id, ²18106060@ittelkom-pwt.ac.id

(artikel diterima: 23-07-2020, artikel disetujui: 12-04-2021)

Abstrak

Kredit merupakan hal utama pada lembaga keuangan yang berpengaruh pada pertumbuhan dan perkembangan lembaga tersebut. Lemahnya pengawasan dan manajemen dalam proses pemberian kredit kepada nasabah dapat menyebabkan tingginya kredit macet. Masalah ini terjadi pada salah satu lembaga keuangan yang memberikan kredit pada nasabah yaitu BMT X. Data tahun 2019 menunjukkan terdapat 600 ajuan kredit multiguna. Dari jumlah tersebut, hanya sekitar 76% menunjukkan kolektabilitas yang baik. Kondisi kolektabilitas kredit yang tidak maksimal menyebabkan BMT X harus mengeluarkan biaya lebih untuk mengumpulkan angsuran yang harus dibayarkan oleh debitur secara langsung. Kredit macet ini menimbulkan kerugian pada lembaga keuangan yang bersangkutan. Untuk itu, dalam memberikan kredit, BMT X harus cerdas menilai kelayakan nasabah. Tujuan penelitian ini adalah menyusun rancangan kebijakan BMT X guna meminimisir kesalahan prediksi nasabah dengan kategori kredit macet. Teknik yang digunakan pada penelitian ini yaitu *data mining* klasifikasi dengan algoritma J.48. Untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data, harus dipilih atribut yang memiliki *information gain* terbesar yang akan diletakkan pada *root node*. Penelitian ini menghasilkan enam *rule* dengan tingkat akurasi sebesar 80,2% sehingga dapat digunakan pihak BMT X untuk menggali informasi kelayakan nasabah untuk mendapatkan kredit.

Kata kunci: Algoritma J.48, data mining, pohon keputusan, resiko kredit

Abstract

Credit is crucial in financial institutions that affects the growth and development of these institutions. Weak supervision and management in the process of extending credit to customers can lead to high non-performing loans. This problem occurred in one of the financial institutions that provides credit to customers, namely BMT X. Data for 2019 showed that there were 600 applications for multipurpose loans. Of these, only about 76% showed good collectability. The condition of credit collectability that is not optimal causes BMT X to spend more to collect installments that must be paid by the debtor directly. This bad credit causes losses to the financial institution. Thus, in providing credit, BMT X must be smart in assessing customer's feasibility. The purpose of this research is to design credit policies in order to minimize the prediction errors of customers with bad credit category. The technique used in this research is classification data mining with the J.48 algorithm. To measure the effectiveness of an attribute in classifying a data sample set, it is necessary to select the attribute that has the greatest information gain which will be placed at the root node. This research produces six rules with an accuracy level of 80,2% so as it can be used by BMT X to search customer's feasibility to gain credit.

Keywords: Algorithm J.48, data mining, decision tree, credit risk

1. PENDAHULUAN

Beberapa indikator krisis keuangan dan perbankan antara lain adalah tingginya tingkat kredit bermasalah, kesulitan likuiditas serta kepailitan dari lembaga keuangan maupun perbankan. Krisis perbankan ini ditandai dengan penilaian terhadap *negative net present value* dari *cashflow* dan penilaian terhadap *net worth* dari lembaga keuangan atau perbankan. Bank merupakan lembaga keuangan yang kegiatannya menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kembali ke masyarakat serta memberikan jasa-jasa bank lainnya (Kasmir, 2014). Dalam beberapa tahun terakhir, industri perbankan telah mengalami perubahan besar. Industri perbankan menjadi lebih kompetitif karena deregulasi peraturan.

Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga (UU RI NO.10 Tahun 1998 Tentang Perbankan). Pemberian kredit merupakan salah satu kegiatan utama industry perbankan yang mengandung resiko, dan sangat berpengaruh terhadap perkembangan dan pertumbuhan bank. Besarnya jumlah kredit yang disalurkan bank dalam aktiva bank adalah $\pm 70\%$ dari total aset sebuah bank. Hal ini menunjukkan bahwa begitu besarnya peranan pemberian kredit dalam sebuah bank yang akan menentukan pertumbuhan dan keuntungan yang diperoleh bank atau bisa dikatakan bahwa kredit memberikan sumbangan pendapatan paling besar di perusahaan perbankan. Di samping keuntungan tersebut, pemberian kredit juga bisa jadi bumerang yang akan memberikan kerugian terbesar untuk perkembangan dunia perbankan berupa kredit macet yang dihasilkan jika nasabah yang diberi pinjaman oleh bank tidak dianalisa dengan baik asal usulnya. Saat ini, banyak lembaga keuangan selain bank yang memberikan kredit kepada nasabah. Salah satunya adalah BMT. BMT merupakan koperasi yang memiliki aktivitas berupa penghimpunan maupun penyaluran pendanaan. BMT memiliki andil yang besar dalam menumbuhkan kembangkan usaha kecil atau mikro dalam memperoleh sumber pembiayaan dana, meskipun dalam pelaksanaannya memiliki beberapa kekurangan.

Permasalahan utama yang dihadapi oleh BMT X terkait dengan pengajuan kredit terjadi pada pembiayaan kredit multiguna. Data tahun 2019 menunjukkan bahwa terdapat 600 ajuan kredit multiguna. Berdasarkan informasi yang diperoleh, hal ini tidak sebanding dengan jumlah analis kredit yang dimiliki oleh BMT X. Analis kredit yang berjumlah dua orang hanya mampu mengevaluasi 5-10 permohonan kredit setiap minggunya. Metode *credit scoring* yang digunakan di BMT X membutuhkan waktu yang relatif lama, sehingga seringkali debitur memberi *feedback* berupa kekecewaan terhadap jangka waktu kepastian persetujuan pengajuan kredit. Permasalahan lainnya adalah tingkat kolektabilitas kredit yang diberikan. Dari jumlah tersebut, hanya sekitar 76% menunjukkan kolektabilitas yang baik. Kondisi kolektabilitas kredit yang tidak maksimal menyebabkan BMT X harus mengeluarkan biaya lebih untuk mengumpulkan angsuran yang harus dibayarkan oleh debitur secara langsung. Sebagai langkah antisipatif, BMT X menetapkan target kolektabilitas kredit sebesar 90%. Kredit macet ini menimbulkan kerugian pada lembaga keuangan yang bersangkutan. Tingkat pembiayaan berkorelasi positif terhadap potensi resiko kredit macet. Resiko tersebut harus diminimalisir agar lembaga keuangan dapat mempertahankan kelangsungan usahanya, sehingga penawaran produk pembiayaan dilakukan kepada debitur yang tepat sasaran.

Berdasarkan latar belakang yang telah diuraikan di atas, maka perlu dilakukan prediksi status kredit calon debitur dengan tujuan meminimalisir resiko kredit macet. Prediksi status kredit merupakan langkah preventif lembaga keuangan dalam usahanya untuk menghindari resiko kredit macet atau tunggakan kredit yang telah sampai jatuh temponya (Sinaga et al., 2019). Permasalahan mengenai analisis keputusan pada persetujuan permohonan kredit dapat diselesaikan dengan menggunakan konsep data mining dengan metode klasifikasi.

2. METODE PENELITIAN

2.1 Tahap Persiapan

Tahap persiapan dimulai dengan pendefinisian masalah yang bertujuan untuk mengetahui penyebab munculnya kredit macet yang dihadapi oleh BMT X. Untuk memudahkan proses identifikasi masalah kredit macet, digunakan pendekatan identifikasi empat elemen permasalahan (Daellenbach et al., 2012) sebagai berikut:

1. Pengambil Keputusan; yang dimaksud dengan pengambil keputusan pada penelitian ini adalah analis kredit pada BMT X.
2. Kriteria Pemilihan; Kriteria pemilihan alternatif keputusan pada permasalahan kredit macet yaitu meminimumkan kesalahan prediksi, meminimumkan jumlah nasabah yang akan melakukan kredit macet
3. Ukuran Performansi; Ukuran performansi pada permasalahan ini adalah, tingkat kesalahan prediksi, jumlah nasabah yang melakukan kredit macet.
4. Kontrol Masukan atau Alternatif Tindakan; Alternatif tindakan dalam permasalahan ini adalah kebijakan sanksi dan kebijakan penghargaan yang dapat diberikan oleh BMT X

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1 Variabel Penelitian

Atribut	Variabel	Nilai Atribut (<i>Instance</i>)
Status	Y	Macet; Tidak Macet
Usia	X1	Muda (≤ 30 th) Paruh baya (30-56 th) Tua (≥ 56 th)
Jenis Kelamin	X2	Wanita Pria
Status Pernikahan	X3	Menikah Tidak Menikah
Jangka waktu Pembiayaan	X4	1 tahun (1-12 Bulan) 2 tahun (13-24 bulan)
Jumlah Pembiayaan	X5	Kecil (\leq Rp 4.000.000) Menengah (Rp 4.000.001 – Rp 8.000.000) Besar (\geq Rp 8.000.001)
Jumlah Anak	X6	Sedikit (≤ 2 orang) Sedang (3-4 orang) Banyak (≥ 5 orang)

Atribut	Variabel	Nilai Atribut (<i>Instance</i>)
Peruntukan Kredit	X7	Modal Konsumtif

2.3 Pembuatan Model

Pembuatan model dalam penelitian ini yaitu *preprocessing data* dan rancang bangun model *decision tree*. Tahap *preprocessing data* meliputi :

1. *Data Selection*

Proses penyeleksian dilakukan untuk memilih data atau atribut yang relevan dengan keperluan analisis.

2. *Data Cleaning*

Proses pembersihan data yang dilakukan yaitu untuk menghilangkan *missing value* dari variable – variable yang telah ditentukan sebelumnya.

3. *Data Transformation*

Tranformasi data yang dilakukan pada data penelitian ini adalah mengubah variabel asal menjadi variabel baru dengan tipe data kategorik.

2.4 Decision Tree

Menurut Jiwei H. dan M. Kamber (Han & Kamber, 2001), *Decision Tree* adalah *flowchart* seperti struktur pohon, dimana setiap *internal node* menunjukkan uji pada sebuah atribut, setiap *branch* mempresentasikan kelas-kelas/ distribusi kelas-kelas. *Node* paling puncak pada pohon adalah *root node*, *Internal nodes* ditunjukan oleh segiempat dan *leaf nodes* ditunjukan dengan oval. Dalam rangka mengklasifikasikan sampel yang tidak dikenal, nilai - nilai atribut dari sampel di uji dengan *decision tree*. Apakah diusut dari akar (*root*) sampai *leaf node* yang memegang prediksi kelas untuk sampel tersebut. *Decision Tree* dapat dengan mudah digantikan dengan *classification rules*.

Dalam penyusunan rancang bangun model *decision tree* secara *top-down*, langkah pertama yang dilakukan adalah melakukan evaluasi terhadap atribut yang ada menggunakan teknik statistik. Hal ini bertujuan untuk mengukur efektifitas suatu atribut dalam pengklasifikasian data. Atribut yang menjadi *root node* adalah atribut dengan nilai *information gain* terbesar (Ahmad et al., 2019). Dalam klasifikasi, atribut bersifat kategori sehingga memiliki nilai diskrit. Oleh sebab itu, atribut yang memiliki nilai *continuous* harus diubah ke bentuk diskrit (Zalilia, 2007). Sebelumnya, harus dicari nilai informasi terlebih dahulu dari data yang akan digunakan dalam klasifikasi. Secara matematis dapat dituliskan dengan persamaan (1).

$$I_{(S_1 S_2, \dots, S_n)} = I - \sum_{i=1}^m p_i \text{Log}_2 (p_i) \dots\dots\dots(1)$$

Keterangan :

$I_{(S_1 S_2, \dots, S_n)}$: nilai informasi

m : banyaknya nilai yang berbeda atribut label kelas yang akan mendefinisikan kelas yang berbeda $C_i (i=1, 2, \dots, m)$

s_i : jumlah sampel dalam himpunan sampel S (berisi s sampel) yang masuk kelas C_i

p_i : peluang bahwa suatu sampel akan masuk ke kelas C_i

1. Nilai Entropi

Entropi merupakan parameter yang digunakan untuk mengukur heterogenitas dalam suatu data. Panjang bit untuk menyatakan informasi dalam entropi secara optimal adalah $p \log_2 - bits$ untuk messages yang mempunyai probabilitas p . Apabila kumpulan sampel data yang ada semakin heterogen, maka akan memiliki nilai entropi yang semakin besar. Secara matematis, nilai entropi masing-masing *instance* dirumuskan pada persamaan (2)

$$Entropy(Q) = -p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)} \dots\dots\dots(2)$$

Keterangan :

- S = ruang sample yang digunakan untuk *training*.
- P₍₊₎ = jumlah keputusan yang bersolusi positif pada data sample
- P₍₋₎ = jumlah keputusan yang bersolusi negatif pada data sample

Sedangkan perhitungan nilai *entropy* untuk setiap atribut dapat dirumuskan sebagaimana pada persamaan (3).

$$E(A) = \sum_{j=1}^v \frac{S_{ij^+ \dots + S_{mj}}}{S} I(S_{ij}, \dots, S_{mj}) \dots\dots\dots(3)$$

2. Information Gain

Setelah memperoleh nilai *entropy*, langkah selanjutnya adalah melakukan pengukuran efektifitas suatu atribut dalam pengklasifikasian data, yang biasa disebut dengan *Information Gain*. Secara matematis, perhitungan nilai *information gain* dari suatu atribut adalah sebagaimana persamaan (4).

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A) \dots\dots\dots(4)$$

Dengan kata lain, *Gain(A)* adalah reduksi yang diharapkan dalam *entropy* yang disebabkan oleh pengetahuan nilai pada atribut *A*. Atribut yang memiliki nilai *information gain* terbesar dipilih sebagai *root node* (simpul akar) pada *decision tree*. Simpul *A* dibuat dan dilabelkan sesuai dengan atribut, simpul cabang dibuat untuk setiap nilai atribut.

3. Mengubah *tree* menjadi *rule*

Tahap terakhir dalam pemodelan *decision tree* adalah mengubah *tree* menjadi *rules*. *Rules* yang dimaksudkan adalah *if-then rule* sesuai dengan *decision tree* yang terbentuk

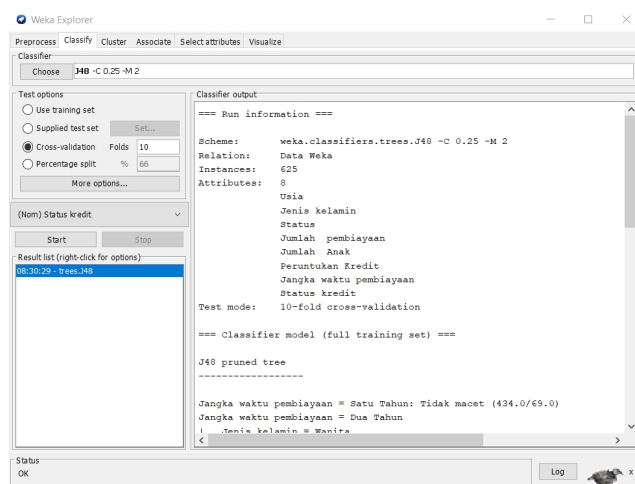
3. HASIL DAN PEMBAHASAN

Data historis yang diperoleh kemudian dilakukan *preprocessing* data, yaitu *data cleaning* dan *data transformation*. Data yang berjenis numerikal diubah menjadi katagorikal. Langkah berikutnya adalah menentukan variabel yang akan digunakan pada pengolahan data selanjutnya. Dari tujuh variabel yang ada kemudian dilakukan seleksi menggunakan uji *chi square*. Dari hasil pengujian tersebut, diperoleh empat variabel yang akan digunakan dalam penelitian ini. Variabel

tersebut adalah jangka waktu pembiayaan, jenis kelamin, jumlah pembiayaan, dan peruntukan kredit.

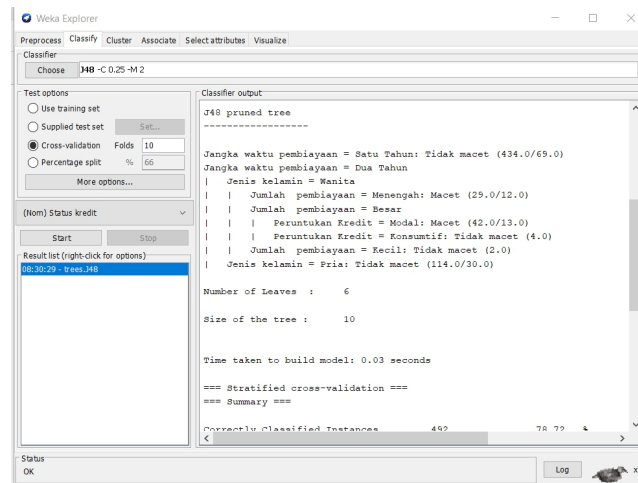
Metode yang digunakan dalam penelitian ini adalah klasifikasi dengan algoritma J.48 atau terkadang juga disebut dengan algoritma C4.5. Output yang dihasilkan berupa pohon keputusan yang memiliki kelebihan, salah satunya adalah mudah dipahami dan dimengerti. Pohon keputusan yang terbentuk kemudian ditransformasikan menjadi *rule*. Dalam penyelesaiannya, penelitian ini menggunakan software WEKA sebagai alat bantu. Langkah penggunaan WEKA adalah sebagai berikut (Mukhibah & Kurniawati, 2019):

1. Pada halaman utama WEKA, pilih tombol *explorer*. Setelah muncul jendela WEKA Explorer. Pilih tab *Preprocess* kemudian pilih tombol *open file* untuk membuka data yang akan digunakan.
2. Setelah muncul kotak dialog WEKA, pilih tombol *All* untuk mengeksekusi data.
3. Selanjutnya pada, aktifkan tab *classify*. Kemudian pilih tombol *Choose* pada tab *Classify* untuk memilih algoritma yang akan digunakan.
4. Pada menu *classify* pilih algoritma J.48. Sehingga akan muncul klasifikasi klasifikasi dengan struktur pohon yang merepresentasikan kriteria pembagian kelas.
5. Setelah pemilihan *classifiers*, pemrosesan dapat dimulai dengan mengklik tombol *Start*, user dapat menghentikan proses ini sewaktu-waktu dengan tombol *Stop*. Saat pengolahan *training set* selesai, *output classifiers* di sebelah akan menampilkan teks yang menggambarkan hasil pengolahan *training set* dan *test set*. Sebuah entry baru saja muncul di kotak *Result list*.
6. Untuk memunculkan *decision tree*, maka klik kanan *trees-J.48* pada menu *result list* kemudian pilih *Visualize tree*.



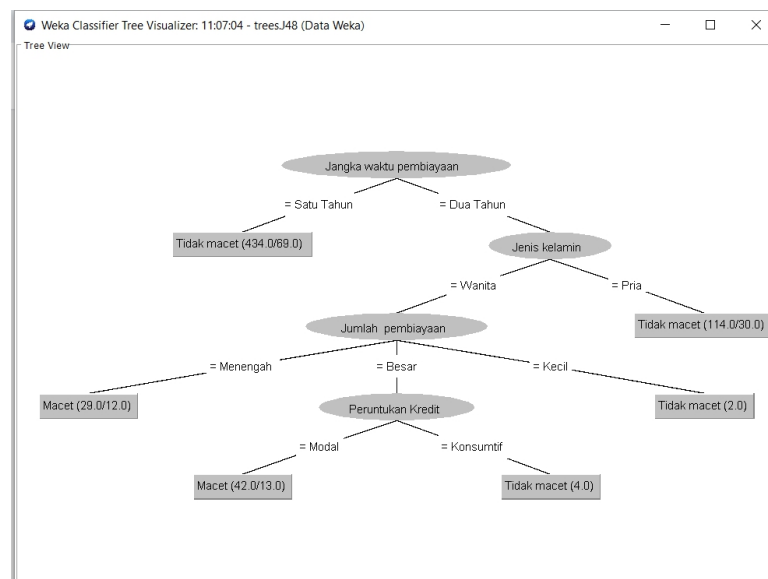
Gambar 1. Hasil dari Klasifikasi Algoritma J.48

Berdasarkan Gambar 1 dapat dilihat pada jendela classifier output bahwa data yang digunakan mempunyai 625 instance dan 8 atribut, dengan 1 label kelas dan 7 atribut. Selanjutnya, hasil dari pemangkasannya algoritma J.48 dan waktu untuk membangun model (kompleksitas waktu) dapat dilihat pada Gambar 2.



Gambar 2. Hasil dari pemangkasan algoritma J.48

Pemangkasan algoritma J.48 menghasilkan 6 *leaf* dan 10 *tree*, dimana definisi *leaf* merupakan simpul yang tidak mempunyai cabang lagi, sedangkan *tree* merupakan simpul yang mempunyai cabang. Setelah tahap pemangkasan, digunakan empat atribut dalam pemrosesan selanjutnya. Pemodelan *decision tree* pada WEKA merupakan tahap untuk memodelkan prediksi status kredit nasabah BMT X dengan menggunakan empat atribut yang ada. Hasil pengolahan klasifikasi *decision tree* dengan menggunakan WEKA dapat dilihat pada Gambar 3.



Gambar 3 Visualize Tree

Dari *decision tree* yang ditunjukkan pada Gambar 1, dihasilkan 6 *rule* sebagai berikut:

1. **If** Jangka Waktu Pembiayaan Satu Tahun **Then** Kredit Tidak Macet
2. **If** jangka Waktu Pembiayaan Dua Tahun **Then** Jenis Kelamin Pria **Then** kredit Tidak Macet
3. **If** jangka Waktu Pembiayaan Dua Tahun , Jenis Kelamin wanita dan jumlah pembiayaan kecil **Then** kredit Tidak Macet
4. **If** jangka Waktu Pembiayaan Dua Tahun , Jenis Kelamin wanita dan jumlah pembiayaan menengah **Then** kredit Macet

5. **If** jangka Waktu Pembiayaan Dua Tahun , Jenis Kelamin wanita, jumlah pembiayaan besar dan Peruntukan kredit konsumtif **Then** kredit Tidak Macet
6. **If** jangka Waktu Pembiayaan Dua Tahun , Jenis Kelamin wanita, jumlah pembiayaan besar dan Peruntukan kredit Modal **Then** kredit Macet

Penggunaan *chi square* di tahap *preprocessing data* merupakan pendekatan *prepruning*, yaitu penghapusan atribut yang tidak memberikan dampak signifikan terhadap model yang dibangun. Enam *rule* yang dihasilkan tersebut kemudian dapat dimanfaatkan oleh pengambil keputusan sebagai bahan pertimbangan pemberian kredit. Pengolahan data yang diproses untuk pemodelan prediksi status kredit nasabah BMT X, menghasilkan 501 *record* dengan hasil klasifikasi bernilai benar dan 124 *record* dengan klasifikasi bernilai salah. Dari pengolahan data tersebut dapat diketahui bahwa tingkat akurasi model *decision tree* yang terbentuk sebesar 80,2%. Nilai akurasi tersebut menunjukkan bahwa model tersebut dapat diimplementasikan dalam prediksi kredit nasabah baru dan dapat digunakan sebagai referensi kredit analis dalam menentukan pemberian kredit ini melalui variabel terpilih.

4. KESIMPULAN

Berdasarkan hasil pengolahan data, dari 625 *record* data nasabah BMT X yang diproses menggunakan 6 *rule* yang terbentuk untuk memprediksi status kreditnya, diperoleh hasil sebanyak 501 *record* dengan hasil klasifikasi bernilai benar dan 124 *record* dengan klasifikasi bernilai salah. Dari pengolahan data tersebut dapat diketahui bahwa tingkat akurasi model *decision tree* yang terbentuk sebesar 80,2%. Nilai akurasi tersebut menunjukkan bahwa model tersebut dapat diimplementasikan dalam prediksi kredit nasabah baru dan dapat digunakan sebagai referensi kredit analis dalam menentukan pemberian kredit ini melalui variabel terpilih. Oleh karena itu, pada penelitian selanjutnya diharapkan dapat melakukan prediksi jumlah paku kredit yang disetujui oleh BMT X.

DAFTAR PUSTAKA

- Ahmad, W., Ahmad, A., Iqbal, A., Hamayun, M., Hussain, A., Rehman, G., Khan, S., Khan, U. U., Khan, D., & Huang, L., 2019 Intelligent hepatitis diagnosis using adaptive neuro-fuzzy inference system and information gain method, *Soft Computing*, 23(21), 10931–10938.
- Daellenbach, H., McNickle, D., & Dye, S., 2012, *Management science: decision-making through systems thinking*, New York: Macmillan International Higher Education.
- Han, J., and Kamber, M., 2000, *Data Mining : Concept and Techniques*, Morgan Kaufmann Publishers, San Francisco, CA.
- Kasmir, 2014, *Bank dan Lembaga Keuangan Lainnya*. Jakarta: Raja Grafindo Persada.

-
- Mukhibah, D., & Kurniawati, A., 2019, Implementasi Data Mining Dalam Prediksi Performance Software Engineer PT. Emerio Menggunakan Decision Tree, *Jurnal Ilmiah Informatika Komputer*, 22(1).
- Sinaga, K., Buulolo, E., & Nadeak, B., 2019, Implementasi Algoritma Decision Tree J48 untuk Memprediksi Resiko Kredit pada Koperasi Simpan Pinjam (Studi Kasus: Kofipindo Lubuk Pakam), *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1).
- Zalilia, L., 2007, Penerapan data Mining untuk IDS, *Tugas Akhir Mata Kuliah EC7010 Program Studi Teknik Elektro Insitut Teknologi Bandung*.