

PENERAPAN DATA MINING ALGORITMA C4.5 DALAM MENGIDENTIFIKASI RISIKO KREDIT PADA LEMBAGA KEUANGAN

Tsalatsatun Nur Rohmah^{1*}, Kurniawati²

^{1,2} Program Studi Sistem Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ivet, Jalan Pawiyatan Luhur IV, Nomor 17, Kota Semarang, Indonesia

Email: tsalatsatun27@gmail.com¹, kurniawati@ivet.ac.id²

ABSTRAK

Untuk mengurangi risiko kredit, lembaga keuangan melakukan analisis data sebelum memberikan kredit. Nasabah dapat merugikan perusahaan apabila mereka tidak membayar kredit dengan lancar. Oleh karena itu, analisis kredit merupakan faktor penting untuk meminimalkan risiko keuangan. Penelitian ini bertujuan untuk mengidentifikasi risiko kredit menggunakan metode data mining dengan membuat pola pohon keputusan menggunakan algoritma C4.5. Pada penelitian ini algoritma C4.5 diimplementasikan dan dianalisa menggunakan software RapidMiner. Hasil analisis data yang dilakukan menggunakan 540 data training dengan 135 data testing, dan 8 variabel dataset serta 1 variabel class menunjukkan hasil akurasi yang cukup besar yaitu sebesar 97.78%.

Kata Kunci: Data Mining, C4.5 Algorithm, Credit Risk.

PENDAHULUAN

Kredit adalah dana yang disediakan oleh suatu lembaga keuangan yang dapat digunakan oleh seorang nasabah baik itu individu maupun badan usaha melalui perjanjian pinjam-meminjam dengan adanya bunga untuk memenuhi kebutuhan finansial [1]. Adapun prosedur memberikan kredit atau pinjaman, lembaga keuangan yang berperan sebagai kreditur harus menganalisis kemampuan finansial nasabah terlebih dahulu untuk mengetahui apakah nasabah tersebut dapat membayar kembali kredit yang telah diberikan beserta bunganya atau tidak. Sehingga, kreditur dapat menentukan kelayakan nasabah tersebut dalam menerima kredit untuk meminimalisir terjadinya kerugian yang disebabkan ketidakmampuan nasabah dalam membayar kembali kredit yang telah diberikan. Analisis kredit dilakukan dengan menggunakan prinsip 5C yaitu *Character* (karakteristik), *Capacity* (kemampuan), *Capital* (modal), *Collateral* (jaminan), dan *Condition of economy* (kondisi perekonomian) [2]. Akan tetapi, menurut Cahyaningtyas (2019), mengungkapkan bahwa analisis kredit dengan menggunakan prinsip 5C masih kurang efektif dan hasilnya kurang akurat [3].

Analisis kredit selain menggunakan prinsip 5C juga bisa dilakukan dengan menggunakan metode *data mining*. *Data mining* merupakan pengembangan dari metode analisis data tradisional dan pendekatan *statistic* (Jackson, 2002). Ada 5 metode pendekatan *data mining* salah satunya adalah *classification*, dimana metode ini bisa menghasilkan pola pengetahuan berbasis *tree*, model *regression* untuk prediksi, dan partisi analisis *cluster* untuk set data tertentu [4]. Terdapat banyak penelitian mengenai analisis kredit menggunakan metode data mining seperti algoritma *Naïve Bayes*, *K-Nearest Neighbor*, *Decision Tree*, dan C4.5. Salah satunya

adalah penelitian oleh Retnosari (2021), dimana penelitian ini menghasilkan hasil analisis kredit kurang efektif. Penelitian ini menggunakan algoritma *Naïve Bayes* untuk menganalisis kelayakan kredit dan menghasilkan tingkat akurasi sebesar 70,00% [5]. Adapun penelitian lain yang dilakukan oleh Santoso (2019), penelitian ini menghasilkan hasil analisis kredit yang lebih efektif dan akurat. Pada penelitian ini digunakan algoritma C4.5 untuk menganalisis kelayakan kredit dan menghasilkan tingkat akurasi sebesar 97,5% [6]. Berdasarkan hasil penelitian di atas, maka data mining adalah salah satu metode yang dapat digunakan untuk melakukan analisis risiko kredit. Oleh karena itu, pada penelitian ini akan menerapkan metode data mining yaitu algoritma C4.5 untuk mengidentifikasi risiko kredit pada lembaga keuangan.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3, hasilnya adalah pohon keputusan yang dapat digunakan sebagai solusi pemecahan masalah pada teknik klasifikasi [7]. Tujuan dari penelitian ini adalah untuk membuat pola pohon keputusan menggunakan algoritma C4.5 untuk mengidentifikasi risiko kredit pada lembaga keuangan.

METODE PENELITIAN

Penelitian ini menggunakan metode klasifikasi dalam bentuk pohon keputusan. Metode klasifikasi digunakan untuk mengkategorikan atau mengklasifikasi data ke dalam kelompok yang sudah ditentukan, berdasarkan atribut atau karakteristik tertentu. Tujuan dari klasifikasi ini adalah untuk menemukan model dari training set yang membagi atribut ke dalam kelas atau kategori yang sesuai [8]. Pohon keputusan digunakan untuk membagi kumpulan data yang sangat besar menjadi kumpulan data yang lebih kecil untuk mempresentasikan *rule* atau aturan. Algoritma C4.5 merupakan salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan. Secara umum, dalam membuat pohon keputusan menggunakan algoritma C4.5 memiliki beberapa tahap, diantaranya adalah sebagai berikut [9] :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Nilai gain tertinggi dari atribut yang ada, digunakan sebagai dasar untuk memilih atribut sebagai *root* atau akar. Nilai gain dapat dihitung dengan menggunakan persamaan berikut :

1. Mencari Nilai Gain

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

Keterangan :

S = banyaknya himpunan (*dataset*) kasus

A = atribut

n = jumlah partisi Atribut A

|S_i| = jumlah kasus partisi ke-i

|S| = jumlah kasus dalam S

2. Mencari Nilai Entropy

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

Keterangan :

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

Tahapan Penelitian

Proses pelaksanaan penelitian meliputi :

1. Identifikasi Masalah

Proses identifikasi masalah penelitian dilakukan dengan memberikan penjelasan penting tentang masalah tersebut. Dengan demikian, penulis memiliki dasar yang dapat digunakan untuk menjelaskan masalah penelitian.

2. Studi Literatur

Penelitian ini mempelajari teori yang terkait dengan topik penelitian ini serta teori yang sedang berkembang. Selain itu, penelitian ini juga meninjau literatur dari berbagai sumber yang terkait dengan topik metode klasifikasi data mining algoritma C4.5 untuk sistem pengolahan data.

3. Sumber Data

Pada penelitian ini data bersumber dari sebuah *website* yang dapat ditemukan pada alamat website <https://romisatriawahono.net/dm/>. Dataset yang digunakan adalah romi-dataset, dan file yang digunakan bernama CreditRisk.csv [10].

Jumlah *record* pada penelitian ini adalah 675 data dengan jumlah atribut sebanyak 9 atribut, yaitu Application_ID, Credit_Score, Late_Payment, Months_In_Jobs,

Debt_Income_Ratio, Loan_Amt, Liquit_Assets, Num_Credit_Lines, dan Credit_Risk. Pada atribut-atribut tersebut, yang digunakan sebagai label atau variabel output adalah atribut Credit_Risk.

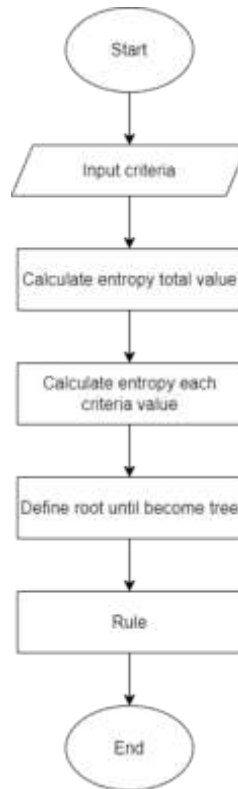
Untuk objek pada penelitian ini adalah calon debitur yang akan diklasifikasikan dalam kelas risiko kredit *low*, *high*, *moderate*, atau *do not lend*. Dengan melakukan proses klasifikasi data peminjam yang memiliki risiko kredit *low*, *high*, *moderate*, atau *do not lend* dengan menggunakan pola data sebelumnya.

4. Split Data Otomatis

Dari total jumlah *record* sebanyak 675 data, pada tahap ini data dibagi menjadi dua yaitu data *training* dan data *testing*. Dengan rasio 8 : 2, dengan jumlah persentase data *training* sebanyak 80% dan data *testing* sebanyak 20%. Data *training* berfungsi sebagai pembentuk pola atau model dan data *testing* berfungsi sebagai penguji model.

5. Analisis Data

Tahap analisis data dilakukan dengan metode *data mining* algoritma C4.5 dan diimplementasikan menggunakan *software* RapidMiner. Hasil analisis data menghasilkan pohon keputusan atau model yang akan digunakan untuk menguji data *testing*.



Gambar 1. Tahapan Algoritma C4.5

6. Kesimpulan

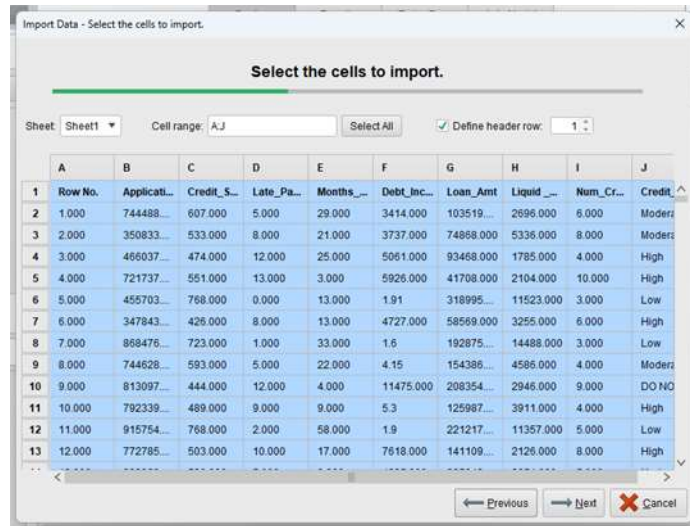
Setelah tahap analisis data selesai dilakukan, tahap selanjutnya adalah penarikan kesimpulan dari hasil analisis data.

HASIL DAN PEMBAHASAN

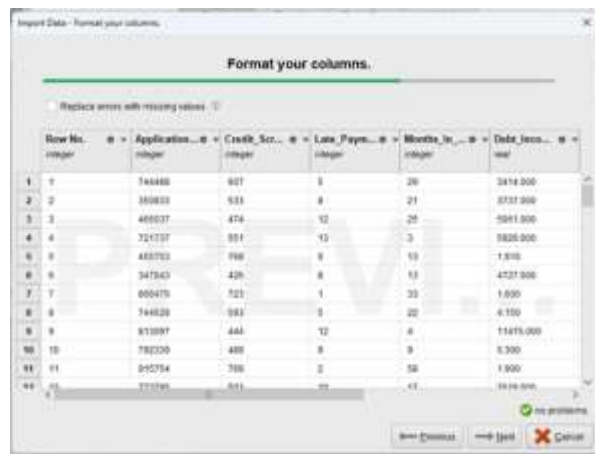
Data yang digunakan pada penelitian ini berjumlah 675 *record*, kemudian data dibagi menjadi 2 dengan jumlah persentase 80% untuk data *training* dan 20% untuk data *testing*. 80% dari jumlah data 675 didapatkan data *training* dengan jumlah 540 dan 20% dari jumlah data 675 didapatkan data *testing* dengan jumlah 135. Variabel data penelitian yang digunakan pada penelitian ini berjumlah 8 variabel *dataset* dan 1 variabel *class*. Untuk variabel *dataset* terdiri atas Application_ID, Credit_Score, Late_Payment, Months_In_Jobs, Debt_Income_Ratio, Loan_Amt, Liquit_Assets, Num_Credit_Lines. Dan untuk variabel *class* nya adalah Credit_Risk.

Pada penelitian ini data diolah menggunakan *software* RapidMiner. Sebelum melakukan proses pengolahan data menggunakan *software* RapidMiner, dataset harus di *import* terlebih dahulu ke dalam *software* RapidMiner. Pada tahap *import* dataset ke dalam *software* RapidMiner terdapat 2 proses yaitu pemilihan *cell* untuk di *import* dan pemilihan variabel

sebagai label seperti yang ditunjukkan gambar 2 dan gambar 3 berikut.

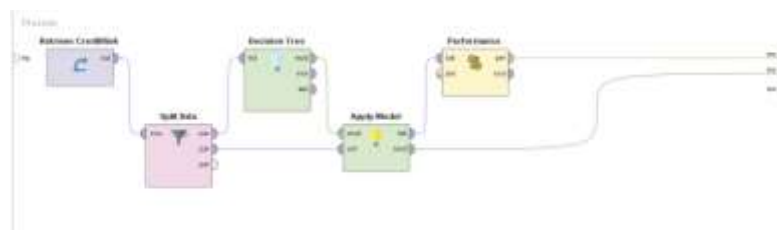


Gambar 2. Pemilihan Cell Untuk Import



Gambar 3. Pemilihan Variabel Sebagai Label

Setelah proses *import dataset* selesai dilakukan, tahap selanjutnya adalah pengolahan data. Pada tahap ini dibuat desain pengujian data seperti yang ditunjukkan pada gambar 4 berikut.



Gambar 4. Desain Pengujian Data

telah dibuat dan menampilkan hasil pada *views Result*. Hasilnya adalah pohon keputusan seperti yang ditunjukkan pada gambar 5 berikut.



Gambar 5. Pohon Keputusan

Pohon keputusan yang dihasilkan adalah *rules* atau aturan yang dapat digunakan untuk mengidentifikasi risiko kredit pada lembaga keuangan.

Analisis hasil pengujian dari 135 data *testing* dengan 540 data *training*, dengan setiap data memiliki 8 variabel *dataset* dan 1 variabel *class* menghasilkan tingkat akurasi yang cukup besar yaitu 97,78% seperti yang ditunjukkan pada gambar 6 berikut.

accuracy: 97.78%

	true Moderate	true High	true Low	true DO NOT LEND	true Very Low	class precision
pred. Moderate	35	1	0	0	0	97.22%
pred. High	0	40	0	1	0	97.56%
pred. Low	0	0	43	0	0	100.00%
pred. DO NOT LEND	0	0	0	1	0	100.00%
pred. Very Low	0	0	1	0	13	92.86%
class recall	100.00%	97.56%	97.73%	50.00%	100.00%	

Gambar 6. Hasil Akurasi

KESIMPULAN

- Berdasarkan analisis dan hasil pengujian yang telah dilakukan, maka dapat disimpulkan metode data mining algoritma C4.5 dapat diterapkan untuk menganalisis risiko kredit pada lembaga keuangan. Algoritma C4.5 dengan jumlah dataset sebanyak 675 record menghasilkan tingkat akurasi yang cukup besar dan akurat yaitu sebesar 97.78%.
- Meskipun hasil penelitian menunjukkan tingkat akurasi yang cukup besar, untuk selanjutnya bisa dilakukan perbandingan antara algoritma C4.5 dengan algoritma klasifikasi lain seperti

algoritma Naïve Bayes dan algoritma K-Nearest Neighbor. Agar menghasilkan rules atau aturan yang lebih akurat, penelitian selanjutnya sebaiknya menggunakan lebih banyak data.

DAFTAR PUSTAKA

- [1] I. Ubaedi and Y. M. Djaksana, “Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection dan Stratified Sampling untuk Prediksi Kelayakan Kredit,” *Sistem Informasi*, vol. 9, no. 1, 2022.
- [2] A. Fauzi, A. Kredit..., and A. Fauzi, “Peran Analisis Kredit terhadap NPL pada PT X,” vol. 5, no. 2, pp. 75–83, 2018.
- [3] R. A. Cahyaningtyas and A. Darmawan, “Pengaruh 5C (Character, Capacity, Capital, Collateral, dan Condition of Economy) terhadap Pemberian Kredit (Studi Kasus Koperasi Pegawai Telkom Purwokerto),” 2019. [Online]. Available: <http://jurnalnasional.ump.ac.id/index.php/kompartemen/>
- [4] J. Jackson, “Data Mining; A Conceptual Overview,” *Communications of the Association for Information Systems*, vol. 8, 2002, doi: 10.17705/1cais.00819.
- [5] R. Retnosari, P. Studi, S. Informasi, S. Nusa, and M. Jakarta, “Analisis Kelayakan Kredit Usaha Mikro Berjalan pada Perbankan dengan Metode Naive Bayes,” *PROSISKO*, vol. 8, no. 1, 2021.
- [6] R. A. Cahyaningtyas and A. Darmawan, “Pengaruh 5C (Character, Capacity, Capital, Collateral, dan Condition of Economy) Terhadap Pemberian Kredit (Studi Kasus Koperasi Pegawai Telkom Purwokerto),” 2019. [Online]. Available: <http://jurnalnasional.ump.ac.id/index.php/kompartemen/>
- [7] Suyanto, *Machine Learning Tingkat Dasar dan Lanjut*. Informatika Bandung, 2018.
- [8] S. Suyanto, *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Informatika Bandung, 2017.
- [9] Mardi Y, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Jurnal Edik Informatika*, vol. 2, no. 2, pp. 213–219, 2017.
- [10] R. S. Wahono, “Data Mining.” Accessed: Jul. 13, 2023. [Online]. Available: <https://romisatriawahono.net/dm/>