

# Analisis Kelayakan Pemberian Kredit Nasabah Koperasi Menggunakan Algoritma C4.5

Rudi Setiawan  
Universitas Trilogi  
Jakarta Selatan, Indonesia  
rudi@trilogi.ac.id

**Abstract**— Kegiatan Analisa terhadap permohonan pinjaman kredit di koperasi merupakan hal yang penting dilakukan agar tidak terjadi penunggakan pembayaran angsuran dikemudian hari oleh para nasabah, hasil analisa kelayakan pemberian pinjaman menjadi penentu atas permohonan pinjaman yang diajukan, berbagai metode analisa dilakukan untuk memprediksi kelayakan pemberian pinjaman kredit, pada penelitian ini dilakukan analisa kelayakan pemberian pinjaman menggunakan algoritma C4.5, algoritma C4.5 merupakan pengembangan dari algoritma ID3. Pengumpulan data dilakukan di Koperasi simpan pinjam Posdaya melati bukit duri, data diolah menggunakan tools RapidMiner, hasil analisis menunjukkan Area Under Curve yang optimis sebesar 0.971 ini menunjukkan hasil klasifikasi berada pada kategori sangat baik.

**Kata kunci** — Algoritma C4.5, Analisis Kelayakan Kredit

## I. PENDAHULUAN

Indonesia sebagai negara terpadat ke 4 didunia dengan total populasi penduduk mencapai ±287.000.000 [1] dengan fenomena bonus demografi yang dimiliki Indonesia terdapat lebih dari 50% penduduk Indonesia berada pada usia produktif untuk bekerja tentunya menjadi tantangan tersendiri bagi pemerintah dalam menyiapkan lapangan pekerjaan maupun akses permodalan bagi usaha masyarakat. Meskipun menurut Badan Pusat Statistik Republik Indonesia pada bulan Maret 2018, jumlah penduduk miskin (penduduk dengan pengeluaran per kapita per bulan di bawah garis kemiskinan) di Indonesia mencapai 25,95 juta orang, berkurang sebesar 633,2 ribu orang dibandingkan dengan kondisi September 2017 yang sebesar 26,58 juta orang artinya pada angka tersebut menandakan masih banyak masyarakat Indonesia yang belum bebas dari permasalahan ekonomi.

Tuntutan kebutuhan hidup memaksa segenap orang untuk berusaha sekuat tenaga dalam memenuhi kebutuhan ekonomi rumah tangganya, berbagai jalan ditempuh mulai dari bekerja lembur dikantor untuk mendapatkan penghasilan tambahan hingga berupaya untuk berusaha secara mandiri maupun berkelompok, akan tetapi kendala yang sering terjadi dalam memulai usaha adalah minimnya modal kerja, sedikit sekali orang yang dapat berusaha dan menjadi tumbuh serta berkembang tanpa pembiayaan dari pihak lain [2], bagi mereka yang tidak memiliki modal alternative nya adalah dengan mencari pinjaman dari berbagai sumber salah satunya adalah koperasi, koperasi sebagai alternative pemberi pinjaman kredit bagi masyarakat kecil [3].

Melihat kondisi yang demikian koperasi di Indonesia hadir dengan semangat bergotong royong dalam hal pemenuhan kebutuhan ekonomi masyarakat yang menjadi anggotanya. UUD 1945 Pasal 33 Ayat 1 yang berbunyi “Perekonomian disusun sebagai usaha bersama berdasar atas asas kekeluargaan” [4]. Hal ini sejalan dengan tujuan koperasi yaitu mengumpulkan tabungan dan memberikan pinjaman kredit bagi anggotanya dengan bunga yang wajar [5]. Koperasi merupakan sokoguru perekonomian bangsa memiliki makna bahwa koperasi berperan sebagai pilar atau penyangga utama perekonomian di Indonesia untuk mewujudkan kesejahteraan dan kemakmuran rakyat [6]. Dibanyak negara sistem keuangan yang dikelola oleh Koperasi simpan pinjam memiliki peran penting [7]. Dikarenakan koperasi menjangkau kebutuhan masyarakat menengah kebawah dan usaha kecil dan menengah.

Dalam bentuk usahanya koperasi memiliki berbagai macam kegiatan usaha seperti koperasi produksi, koperasi konsumsi, koperasi serba usaha dan koperasi simpan pinjam, sama seperti usaha bisnis lainnya yang semuanya memiliki resiko [8]. Pada koperasi atau perusahaan yang menyediakan pinjaman kredit, resiko terbesar adalah resiko kredit yang tidak dibayar oleh pelanggan [9].

Riset ini berkaitan dengan koperasi simpan pinjam dalam menentukan pemberian kredit tabur puja kepada para anggota koperasi. Koperasi simpan pinjam berperan memberikan pinjaman kepada para anggotanya dengan bunga yang telah ditentukan, akan tetapi dalam praktiknya masih terdapat adanya permasalahan dalam pembayaran angsuran atau yang dikenal dengan istilah kredit macet.

Koperasi simpan pinjam di Posdaya melati bukit duri telah ada sejak tahun 2014 dengan total dana yang tersalurkan lebih dari 1 Milyar Rupiah dengan jumlah anggota yang melakukan pinjaman 208 anggota dengan jumlah anggota yang menunggak sebanyak 15 anggota. Untuk membantu pengelola dalam mengantisipasi adanya kredit macet pada awal pengajuan pinjaman dikoperasi melati bukit duri, pada penelitian ini akan dilakukan prediksi dengan melakukan simulasi terkait kelancaran anggota koperasi dalam membayar iuran pinjaman berdasarkan sampel data yang ada di koperasi melati bukit duri menggunakan algoritma C4.5. Algoritma C4.5 merupakan algoritma yang umum digunakan untuk keperluan klasifikasi data dengan metode klasifikasi menggunakan pohon keputusan. Algoritma C4.5 berperan dalam pembentukan pohon keputusan sebagai akhir dalam menentukan prediksi kelayakan pengajuan kredit.

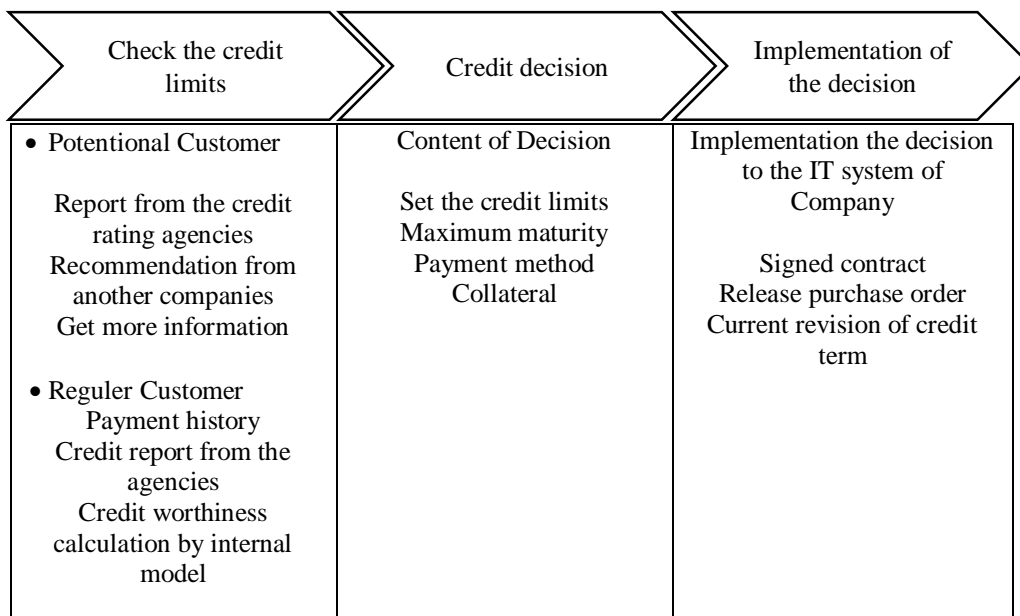
## II. TINJAUAN PUSTAKA

Analisis kredit merupakan hal yang penting dalam mengantisipasi resiko keuangan [10] dan mengidentifikasi data debitur yang bermasalah merupakan hal yang sulit [11]. Hal ini menarik untuk diteliti.

A. Penelitian Sebelumnya

Analisa kinerja keuangan pada nasabah telah banyak dilakukan diantaranya oleh [12] dengan melakukan analisa terhadap data kuantitatif dan kualitatif dari para nasabah yang dinilai dapat menghasilkan informasi yang lebih komprehensif sebagai bahan pengambilan keputusan pemberian kredit tetapi kurang akurat dari segi hasilnya.

Pada penelitian yang dilakukan oleh [13] menyebutkan terdapat 3 fase dalam manajemen kredit yaitu cek batasan credit yang dapat diberikan dengan cara membagi segment customer menjadi 2 (potensial costumer dan reguler customer), pengecekan batasan credit pada kostumer yang potensial dapat dilakukan dengan meminta data dari lembaga pemeringkatan kredit, meminta rekomendasi dari perusahaan yang mengetahui rekam jejak customer dan mencari informasi relevan lainnya mengenai customer. Sedangkan pada reguler customer pengecekan batasan kredit dapat dilakukan dengan melihat history pembayaran customer, laporan kredit dari agency dan melakukan perhitungan kredit yang dilakukan oleh internal perusahaan. Selengkapnnya mengenai 3 fase dalam manajemen kredit ditunjukkan pada Gambar 1.



Gambar 1. Credit Risk Management Process

Penelitian lainnya menggunakan annealing algorithm untuk menganalisa kelayakan pemberian kredit pada nasabah pernah dilakukan disuatu perusahaan kredit di jerman oleh [14] dengan hasil penelitiannya menyatakan bahwa model yang diusulkan efektif.

B. Dasar Teori

a. Algoritma C4.5

Algoritma data mining C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Klasifikasi merupakan salah satu proses pada data mining yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar. Algoritma C4.5 dikembangkan tahun 1993 oleh Quinlan [15] dan merupakan pengembangan dari algoritma ID3. Gambar 2 merupakan pseudocode algoritma C4.5.

```

Check for base case.
For each attribute a
    find the normalized information gain from splitting on a
Let a_best be the attribute with the
    highest normalized information gain
Create a decision node that splits on a_best
Recure on the sublists obtained by splitting on a_best, and add those
    nodes as children of node
    
```

Gambar 2. Pseudocode Algorithm C4.5

Algoritma C4.5 menggunakan gaint untuk menentukan simpul awal berdasarkan informasi dari entropy [16], rumus entropy dan gain dituliskan pada persamaan 1 dan 2.

Tabel 1. Persamaan 1 dan 2

Menghitung Nilai Entropy	Menghitung Gain
$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i$ <p style="text-align: right;">(1)</p>	$Gain (S,A) = Entropy (S) - \sum_{i=1}^n p_i * Entropy (S_i)$ <p style="text-align: right;">(2)</p>

<p>Keterangan:  S : himpunan kasus  n : jumlah partisi S  pi : proporsi dari Si terhadap S</p>	<p>Keterangan:  - S : himpunan  - A : atribut  - n : jumlah partisi atribut A  -   Si   : jumlah kasus pada partisi ke-i  -   S   : jumlah kasus dalam S</p>
--	--

*b. Confusion Matrix*

Hasil kerja dari klasifikasi dicatat kedalam suatu table yang bernama Matriks confusion. contoh matriks confusion untuk 2 kelas klasifikasi, ditunjukkan pada tabel 2.

Tabel 2. Confusion Matrix

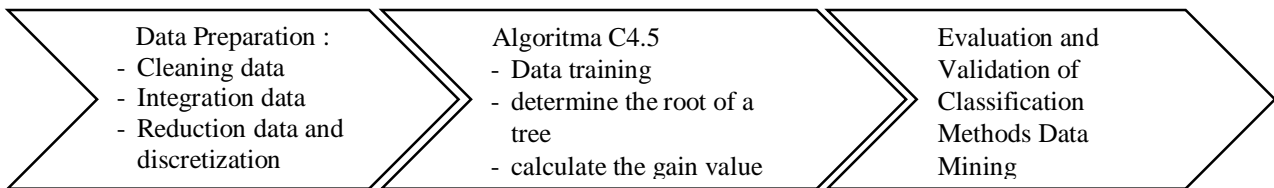
		Class Prediction Results	
		Positive	Negative
Class Prediction	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Validitas sistem dinilai dengan cara menghitung nilai TP, TN, FP, dan FN. Kemudian akurasi dapat dihitung menggunakan persamaan 3 sebagai berikut :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

III. METODOLOGI

Pada penelitian ini, terdapat 6 atribut yang digunakan untuk melakukan prediksi menggunakan algoritma C.45, atribut tersebut adalah pendapatan, tanggungan, jenis usaha, jangka waktu, kepemilikan rumah dan satu atribut class yaitu status pembayaran. Tahapan-tahapan yang dilakukan pada penelitian ini digambarkan pada Gambar 3.



Gambar 3. Tahapan Penelitian

Langkah pertama pada data preparation adalah cleaning data yaitu membersihkan data dari nilai yang tidak valid, selanjutnya integrasi data yaitu menggabungkan beberapa data yang dianggap untuk diproses, proses berikutnya adalah reduction data dan discretization yaitu mengambil beberapa data yang akan digunakan sebagai data training. Langkah kedua yaitu memproses data menggunakan algoritma C4.5, data training yang telah melewati tahap preparation akan dihitung nilai gain dari masing-masing atribut, nilai gain tertinggi pada suatu atribut akan menjadi akar pertama. Sebelum menghitung gain hitung terlebih dahulu entropy dari masing-masing atribut. Langkah kedua dilakukan berulang sampai semua record terpartisi, proses partisi pohon keputusan akan terhenti apabila semua record dalam simpul N terdapat kelas yang sama atau tidak ada atribut yang dapat di partisi atau tidak ada record didalam cabang yang kosong. Langkah ketiga adalah melakukan evaluasi dan validasi dari hasil klasifikasi yang telah dilakukan oleh algoritma C4.5 dengan menghitung confusion matrix dan kurva Receiver Operating Characteristic.

IV. HASIL DAN PEMBAHASAN

Pada penelitian dilakukan pengukuran terhadap data confusion matrix yang berupa nilai accuracy, precision, recall dan kurva Receiver Operating Characteristic

a. Accuracy

Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual, accuracy dari ujicoba yang dilakukan pada penelitian ini menunjukkan nilai sebesar 73.42% yang didapat dari total smooth predictions dengan hasil true smooth sebanyak 53 record dan problematic predictions dengan hasil true problematic sebanyak 5 record. Selengkapnya mengenai detail Accuracy ditunjukkan pada tabel 3.

Table 3. Accuracy Predictions

precision: 73.61% (positive class: Lancar)			
	True Smooth	True Problematic	Class Predictions
Smooth Predictions	5	2	71.43%
Problematic Predictions	19	53	73.61%
Class Recall	20.83%	96.36%	
accuracy: 73.42%			
	True Smooth	True Problematic	Class Predictions
Smooth Predictions	53	19	73.61%
Problematic Predictions	2	5	71.43%
Class Recall	96.36%	20.83%	

## b. Precision

Precision merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem dengan data berupa persentase, nilai precision dari ujicoba yang dilakukan pada penelitian ini menunjukkan hasil sebesar 73.61% dari total problematic predictions dengan hasil true problematic sebanyak 53 record dan problematic predictions dengan hasil true problematic sebanyak 53 record dan true smooth 19 record. Selengkapnya mengenai detail precision ditunjukkan pada tabel 4.

Table 4. Precision Predictions

precision: 73.61% (positive class: Lancar)			
	True Smooth	True Problematic	Class Predictions
Smooth Predictions	5	2	71.43%
Problematic Predictions	19	53	73.61%
Class Recall	20.83%	96.36%	

## c. Recall

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. nilai recall dari ujicoba yang dilakukan pada penelitian ini menunjukkan hasil sebesar 96.36% dari total problematic predictions dengan hasil true problematic sebanyak 53 record dan smooth predictions dengan hasil true problematic sebanyak 2 record. Selengkapnya mengenai detail recall ditunjukkan pada tabel 5.

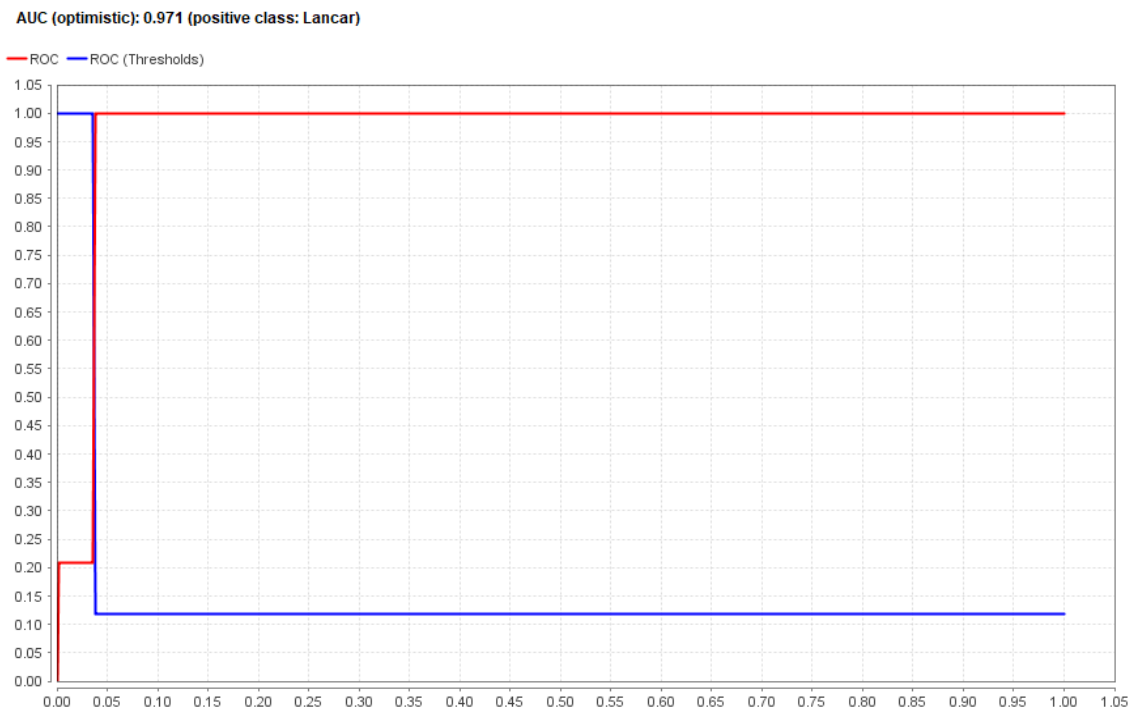
Table 5. Recall Predictions

recall: 96.36% (positive class: Lancar)			
	True Smooth	True Problematic	Class Predictions
Smooth Predictions	5	2	71.43%
Problematic Predictions	19	53	73.61%
Class Recall	20.83%	96.36%	

## d. Kurva Receiver Operating Characteristic

Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengekspresikan data confusion matrix. Garis horizontal mewakili nilai false positives (FP) dan garis vertikal mewakili nilai true positives (TP).

Akurasi Area Under Curve (AUC) dari kurva ROC dikatakan sempurna apabila nilai AUC mencapai 1.000 dan akurasi buruk jika nilai AUC dibawah 0.500. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok [17]. 0.900-1.000 klasifikasi sangat baik, 0.800-0.900 klasifikasi baik, 0.700-0.800 klasifikasi cukup, 0.600-0.700 klasifikasi buruk, 0.500-0.600 klasifikasi salah. Dari gambar 4. dapat diketahui bahwa nilai Area Under Curve yang optimis dari penelitian ini sebesar 0.971, hal ini menunjukkan bahwa model algoritma C4.5 mencapai klasifikasi sangat baik.



Gambar 4. Kurva Receiver Operating Characteristic

## V. KESIMPULAN

Dari hasil penelitian ini diketahui nilai akurasi perlu ditingkatkan, banyak faktor yang mempengaruhi nilai akurasi diantaranya adalah jumlah data training dan kompleksitas isi data training, data training dengan jumlah yang lebih banyak dan kompleks dapat mempengaruhi hasil klasifikasi, saran penelitian kedepan perlu dilakukan percobaan pengujian dengan algoritma yang lain seperti naïve bayes atau support vector machine maupun algoritma lainnya sehingga dapat diketahui hasil klasifikasi terbaik untuk kasus ini.

## DAFTAR PUSTAKA

- [1] BPS, 2018. Profile Kemiskinan di Indonesia Maret 2018, Berita resmi statistik No.57/07/Th.XXI.
- [2] J. R. Brown, J. A. Cookson and R. Z. Heimer, "Growing up without finance", *Journal of Financial Economics*, vol. 134, issue 3, pp. 591-616, December 2019.
- [3] T. P. Silva, *et al*, "Financial and economic performance of major Brazilian credit cooperatives", *Contaduría y Administración*, vol. 62, issue 5, Pages 1442-1459, December 2017.
- [4] UUD 1945 Pasal 33 Ayat 1
- [5] M. S. Henock, "Financial sustainability and outreach performance of saving and credit cooperatives: The case of Eastern Ethiopia", *Asia Pacific Management Review*, vol. 24, issue 1, pp. 1-9, March 2019.
- [6] R. Setiawan, U. A. Faruq, "Sistem Informasi Koperasi Taburpuja Berbasis Tanggung Renteng", *Jurnal Sistem Informasi dan Sains Teknologi*, vol. 1, issue. 1, Februari 2019.
- [7] D. McKillop, *et al*, "Cooperative financial institutions: A review of the literature", *International Review of Financial Analysis*, vol. 71, October 2020, doi:10.1016/j.irfa.2020.101520
- [8] S. Anna, K. Boris and W. Ivana, "Impact of credit risk management", 4th World Conference on Business, Economics and Management, *Procedia Economics and Finance*, vol. 26, pp. 325 – 331, 2015.
- [9] P. Adamko, T. Klietstik, M. Birtus, "History of credit risk models". 2nd international conference on economics and social science, *Information Engineering Research Institute*, pp. 148-153.
- [10] K. K. Lai, L. Yu, L. Zhou, S. Wang, "Credit Risk Evaluation with Least Square Support Vector Machine", *International Conference on Rough Sets and Knowledge Technology, RSKT 2006: Rough Sets and Knowledge Technology*, vol. 4062, pp 490-495.
- [11] O. O, Odeh, A.M, Featherstone, S. Das. "Predicting credit default: Comparative results from an artificial neural network, logistic regression and adaptive neuro-fuzzy inference system", *International Research Journal of Finance and Economics* vol. 42, pp. 7-18, January 2010.
- [12] G. Caruso, *et al*, "Cluster Analysis for mixed data: An application to credit risk evaluation", *Socio-Economic Planning Sciences*, Available online 13 April 2020, Article 100850
- [13] D. Buc, T. Klietstik, "Aspects of statistics in terms of financial modelling and risk". *Proceeding of the 7th International Days of Statistics and Economics*, pp. 215- 224, Prague. 2013.
- [14] Y. Jiang, "Credit Scoring Model Based on Decision Tree and the Simulated Annealing Algorithm". *World Congress on Computer Science and Information Engineering*, pp. 18 – 22. Los Angeles: IEEE Computer Society. 2009.
- [15] J. R. Quinlan, *C4.5: Programs for Machine Learning*, USA: Morgan Kaufmann, 1993.
- [16] J. Han and M. Kamber, *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. 2006.
- [17] F. Gorunescu, *Data Mining: Concept, Models and Techniques*, Springer-Verlag Berlin Heidelberg, 2011. DOI. 10.1007/978-3-642-19721-5