

Teknik Bagging Pada Klasifikasi Algoritma C4.5 Untuk Menentukan Persetujuan Peminjaman di Koperasi

Daryanto^{a,*}, Miftahur Rahman^a, Dedep Vicky Wardiyanto^a

^aUniversitas Muhammadiyah Jember, Jl. Karimata No. 49, Jember, Indonesia

*correspondence email: daryanto@unmuhjember.ac.id

Abstract—Cooperatives are an economic organization that is different from other economic organizations, cooperatives consist of weak groups by making mutual agreements, cooperation of members to achieve mutual welfare and each member is able to carry out his obligations as a cooperative member organization. In this study the bootstrap aggregating technique in 17 tests with 2-fold, 5-fold, and 10-fold divisors obtained the highest accuracy results, namely 56.92% for 2-fold, 65.38% for 5-fold, and 76.92% at 10-fold.

Index Terms—Cooperatives; Data mining; Classification; C4.5; Bootstrap aggregating.

Abstrak—Koperasi merupakan suatu organisasi ekonomi yang berbeda dengan organisasi ekonomi lainnya, koperasi terdiri dari golongan lemah dengan melakukan persetujuan bersama, Kerjasama anggota untuk mencapai kesejahteraan hidup Bersama dan masing-masing anggota sanggup menjalankan kewajibannya sebagai organisasi anggota koperasi. Pada penelitian ini teknik bootstrap aggregating dalam 17 kali pengujian dengan pembagi 2-fold, 5-fold, dan 10-fold mendapatkan hasil akurasi tertinggi yaitu 56,92% pada 2-fold, 65,38% pada 5-fold, dan 76,92% pada 10-fold.

Kata Kunci—Koperasi; Data mining; klasifikasi; C4.5; Bootstrap aggregating.

I. PENDAHULUAN

Koperasi ialah sesuatu sistem ekonomi yang berlainan dengan sistem ekonomi yang lain, koperasi terdiri dari kategori lemas dengan mengadakan persetujuan bersama, kerjasama personel guna menggapai keselamatan hidup bersama serta tiap-tiap bagian mampu melaksanakan kewajibannya selaku badan badan koperasi[1].

Algoritma C4.5 ialah algoritma pengembangan dari algoritma ID3(*Iterative Dichotomiser 3*) yang gampang dipahami serta bisa divisualisasikan dalam wujud tumbuhan ketetapan yang menarik [2]. Banyak penelitian yang menerapkan algoritma C4.5 [3][4][5]. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan dan teknik *Bagging* terbukti dapat digunakan untuk meningkatkan akurasi dari algoritma C4.5. Menurut penelitian [6] berdasarkan Penerapan *Bagging* Untuk Memperbaiki Hasil Prediksi Nasabah Perusahaan Asuransi X menyimpulkan bahwa metode bagging dapat melakukan prediksi dengan tingkat kebenaran/akurasi mencapai 90%, sementara metode yang lain hanya memiliki tingkat akurasi kurang dari 85%. Hasil riset membuktikan kalau tata cara pengelompokan informasi mining dengan memakai algoritma C4.5 buat penemuan angsuran macet sudah menciptakan angka ketepatan, dimana ketepatan itu merupakan algoritma C4.5 sebesar 92,00%, Dari hasil ketepatan itu hingga algoritma C4.5 ialah algoritma yang bagus keakuratannya dipakai buat meprediksi angsuran macet pada koperasi simpan sanggam.

Bersumber pada riset terdahulu bisa di simpulkan kalau metode bagging teruji bisa dipakai buat tingkatkan ketepatan algoritma yang dipakai, riset ini hendak mencoba coba metode bagging dalam tingkatkan ketepatan algoritma C4.5.

II. METODE PENELITIAN

A. Alur Pelaksanaan

Pemahaman tentang tingkatan riset mengenai aplikasi tata cara C4. 5 dengan Metode Bagging dalam pengelompokan pelanggan koperasi Barokah Berhasil buat memastikan persetujuan peminjaman anggaran. Metodologi ini lewat sebagian jenjang ialah Mendefinisikan, analisa permasalahan, menekuni *literature*, mengumpulkan informasi, praproses informasi, aplikasi serta penilaian. Selanjutnya ini merupakan bagan metodologi riset yang bermuatan jenjang riset yang hendak dicoba selaku selanjutnya [7]:



Gambar 1. Tahapan Metode Penelitian

1. Identifikasi Masalah

Rekognisi permasalahan ialah langkah permulaan dari kemampuan permasalahan dimana subjek dalam sesuatu hubungan khusus dapat dikenali selaku sesuatu permasalahan. Langkah mengenali permasalahan yang dicoba pada koperasi Barokah berhasil dengan metode melaksanakan observasi(observation) yang ialah Metode pengumpulan informasi dengan melaksanakan observasi dengan cara langsung di alun- alun kepada subjek yang hendak di cermat.

Observasi dalam rangka pengidentifikasian masalah pada penelitian ini adalah observasi non-participan dimana peneliti tidak terlibat langsung dan hanya sebagai pengamat independent dengan teknik pengamatan terbuka dimana keberadaan pengamat diketahui oleh subjek yang diteliti dan subjek memberi kesempatan kepada pengamat untuk mengamati peristiwa yang diteliti.

2. Studi Pustaka

Riset pustaka merupakan mencari rujukan filosofi yang relevan dengan permasalahan ataupun kasus yang ditemui. Riset pustaka yang dicoba oleh periset ialah dengan melaksanakan pencarian kepada jurnal- jurnal dari riset terdahulu yang esoknya bisa mensupport serta bisa dijadikan referensi yang menguatkan argumentasi- argumentasi yang terdapat.

3. Pengumpulan Data

Pengumpulan informasi bisa didapat langsung dari koperasi Barokah berhasil tahun 2021. Informasi ini melingkupi sebagian ciri, antara lain: jenis kelamin, rentang usia, kisaran angsuran, lama tempo, kisaran pendapatan, bidang pekerjaan, status pernikahan, status tempat tinggal dan keterangan. Kemudian *Dataset* diolah menggunakan metode *algoritma C4.5*.

Tabel 1. Data Atribut dan Tipe Data

Atribut	Tipe Data
Jenis Kelamin	Nominal
Rentang Usia	Numerik
Kisaran Angsuran	Numerik
Lama Tempo	Numerik
Kisaran Pendapatan	Numerik
Bidang Pekerjaan	Nominal
Status Pernikahan	Nominal
Status Tempat Tinggal	Nominal
Keterangan	Nominal

B. Perhitungan

1. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma untuk membentuk pola pohon keputusan, metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan, dimana aturan tersebut dapat dengan mudah dipahami dengan bahasa alami dan juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query language* untuk mencari *record* pada kategori tertentu[8].

Dalam membuat pohon keputusan menggunakan *algoritma C4.5* langkah-langkah yang harus dilakukan adalah sebagai berikut:

- Baca Data Latih Koperasi Barokah Jaya.
- Mencari angka *entropy* memakai metode *entropy* diterima setelah itu mencari angka gain dari tiap atribut yang terdapat. Ciri yang mempunyai angka gain terbanyak hendak di manfaatkan selaku pangkal tumbuhan. Selanjutnya merupakan metode mencari angka entropy serta gain[9].

$$\text{Entropy}(S) = \sum_{j=1}^k -P_j \log_2 P_j \quad (1)$$

Keterangan:

S : Himpunan kasus

K : Jumlah partisi S

P_j : Probabilitas yang didapat dari jumlah (YA/TIDAK) dibagi total kasus

Berikut adalah rumus untuk mencari nilai Gain[9]:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i) \quad (2)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

N = jumlah partisi attribute A

|S_i| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

2. Bagging

Bagging yakni kependekan dari *bootstrap aggregating* memakai *subdataset (bootstrap)* buat menciptakan set penataran pembibitan L (*learning*), L melatih bawah berlatih memakai metode penataran yang tidak normal, serta setelah itu sepanjang pengetahuan mengutip pada umumnya (Breiman, 1996).

Data *bootstrap* diciptakan oleh resampling contoh seragam dengan penggantian dari data pelatihan asli. Pengklasifikasi dapat dilatih secara paralel dan klasifikasi akhir dihasilkan dengan menggabungkan klasifikasi ensemble. *Bagging* telah dianggap sebagai teknik pengurangan varians untuk classifier yang diberikan.

Bagging dikenal sangat efektif ketika pengklasifikasi tidak stabil, yaitu ketika perturbasi set belajar dapat menyebabkan perubahan signifikan dalam perilaku klasifikasi, karena *bagging* meningkatkan kinerja generalisasi karena pengurangan varians (*noise*) tetap terjaga atau hanya sedikit meningkatkan bias (Kim & Kang, 2012).

3. RapidMiner

Rapidminer ialah fitur lunak yang bertabiat terbuka (*open source*). *Rapidminer* merupakan suatu pemecahan buat melaksanakan analisa kepada informasi *mining*, *text mining* serta analisa perkiraan.

Rapidminer memakai bermacam metode deskriptif serta perkiraan dalam membagikan pengetahuan pada konsumen alhasil bisa membagikan ketetapan yang sangat bagus. Dalam hal ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai dengan menggunakan *framework RapidMiner*[10].

4. Confusion Matrix

Confusion Matrix merupakan tata cara yang menggunakan bagan *matrix* dimana masing- masing kolom pada *matrix* merupakan kategori perkiraan, sebaliknya masing- masing baris menggantikan peristiwa kategori sesungguhnya. Dengan cara pendek *Confusion Matrix* membagikan spesifikasi perinci hal *misclassification* [8].

Tabel 2 Confusion Matrix

CLASSIFICATION	REAL CLASS	
	Class= Yes	Class= No
PREDICTED CLASS	Class= Yes	A (true positive-TP)
	Class= No	C (false negative-FN)
		B (false positive-FP)
		D (true negative-TN)

Sehabis informasi percobaan dimasukkan pada bagan *confusion matrix* jumlah nilai- nilai yang sudah dimasukkan buat mengenali angka ketepatan serta presisinya [8]. Kemudian untuk menghitung nilai tersebut menggunakan persamaan dibawah ini:

$$\text{Presisi} = \frac{TP}{FP+TP} * 100 \tag{3}$$

$$\text{Akurasi} = \frac{TP+T}{TP+FN+FP+T} * 100 \tag{4}$$

III. HASIL DAN PEMBAHASAN

A. Gambaran *Dataset*

Data yang akan digunakan dalam penelitian ini merupakan data dari koperasi simpan pinjam Barokah Jaya pada tahun 2021. Data ini berupa data peminjaman uang yang terdiri dari 9 atribut yaitu Jenis Kelamin, Rentan Usia, Kisaran Angsuran, Lama Tempo, Kisaran Pendapatan, Bidang Pekerjaan, Status Pernikahan, Status Tinggal, dan Keterangan. Data ini memiliki 2 output yaitu lancar dan tidak lancar, jumlah data awal memiliki 200 *record*, namun karena data tidak sinkron, maka data awal berubah menjadi 130 data.

B. Skenario Uji

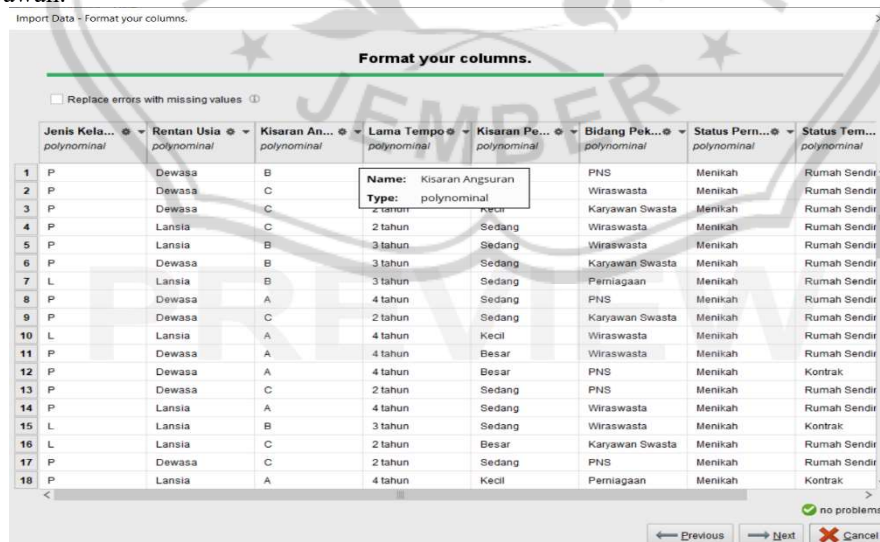
Pengetesan hendak dicoba dengan memakai *tool rapid miner sanggar tipe 9.5* selaku pengaplikasiannya. Saat sebelum dipakai pada *tool rapid miner* informasi terlebih dulu dipecah jadi informasi *training* serta informasi *testing*. Dimana dengan jumlah informasi yang berlainan cocok dengan *k-fold cross validation*. Informasi hendak dipecah jadi k serupa dengan 2, 5, serta 10.

C. Hasil Pengujian Teknik Bagging pada Algoritma C4.5

Hasil yang hendak di perlihatkan ialah hasil terbaik dari tiap- tiap *k-fold cross validation*. Sebab angka output terdapat 2 ialah mudah serta tidak mudah hingga hendak dicoba pengetesan memakai *rapid miner* kemudian hasil yang diperoleh hendak dihitung dengan *confusion matrix* buat memperoleh hasil ketepatan serta akurasi.

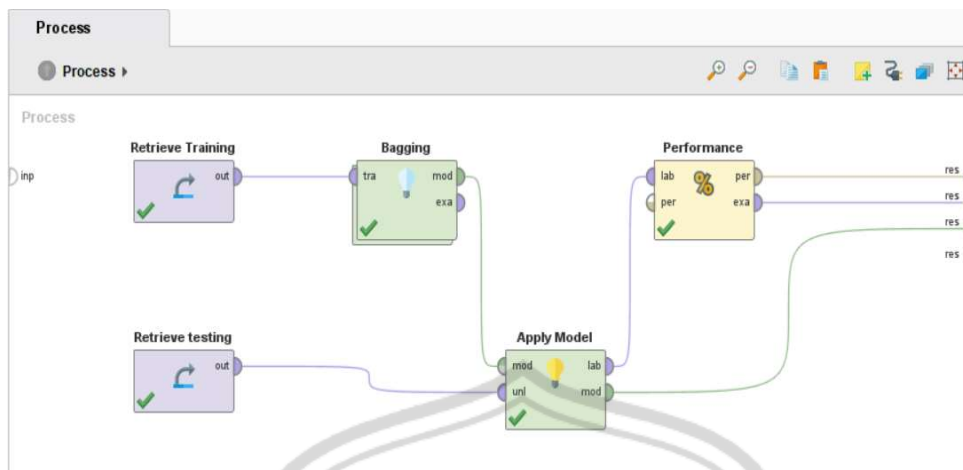
D. Hasil Uji 2-Fold

Bersumber pada *scenario* percobaan pada pengetesan dengan k=2 ialah memilah informasi *training* sebesar 65 serta lebihnya jadi informasi *testing*. Sehabis seluruh informasi sudah disiapkan hingga masukan informasi kedalam *rapid miner* buat diproses, selanjutnya cara pengerjaan informasi memakai metode *bagging* pada algoritma C4. 5. Langkah- langkah dalam *aplikasi rapid miner* hendak dipaparkan pada lukisan dibawah:



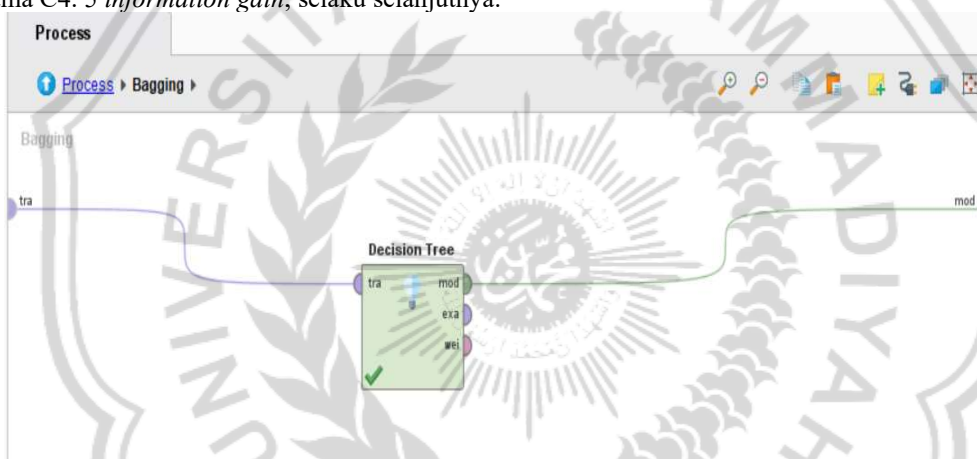
Gambar 2. Proses *Import Data*

Pada cara ini informasi yang di import cocok dengan penjabaran serta buat ciri kendala yang dialamirode dirubah jadi “merek” supaya *rapid miner* mengenali selaku output yang dipakai. Kemudian dicoba pemodelan semacam selanjutnya:



Gambar 3. Proses Pengujian

Pada cara ini hendak dicoba pengetesan pada informasi *training* buat memperoleh bentuk serta informasi testing buat memperoleh angka ketepatan serta akurasi. Kemudian didalam kotak *bagging* kita masukkan algoritma C4.5 *information gain*, selaku selanjutnya:



Gambar 4. Proses Bagging

Sehabis berakhir, hingga hendak dicoba cara buat memperoleh ketepatan serta akurasi serta diperoleh hasil selanjutnya:

accuracy: 56.92%

	true Tidak Lancar	true Lancar	class precision
pred. Tidak Lancar	16	10	61.54%
pred. Lancar	18	21	53.85%
class recall	47.06%	67.74%	

Gambar 5. Hasil akurasi dan presisi percobaan ke-1

Hasil pengetesan pada Lukisan di atas angka dari ketepatan langsung dihitung memakai dengan 2 *output* tanpa mengganti informasi serta diperoleh ketepatan sebesar 56.92%. Hingga hendak dihitung dengan *confusion matrix* 2 output dengan 2 hasil, dengan determinasi dimana hendak dibagi jadi 2 output yang berlainan satu serupa lain. Output yang dipakai ialah mudah dengan tidak mudah selaku selanjutnya:

Tabel 3 Confusion Matrix

CLASSIFICATION	KLASIFIKASI	
	Lancar	Tidak Lancar
	Lancar	TP=16
Tidak Lancar	FP=10	TN=21

Berdasarkan tabel diatas dan presisi yang akan didapat sebagai berikut:

$$akurasi = \frac{16 + 21}{16 + 10 + 18 + 21} * 100\% = 56,92\%$$

$$presisi = \frac{16}{16 + 10} * 100\% = 61,54\%$$

E. Hasil Uji 5-Fold

Bersumber pada sekenario percobaan pada pengetesan dengan k=5 ialah memilah informasi *training* sebesar 26 serta lebihnya jadi *informasi testing*. Sehabis seluruh informasi sudah disiapkan hingga masukan informasi kedalam *rapid miner* buat diproses. Cara dilakuka serupa dengan pengetesan *2-fold*. Ketepatan serta akurasi yang diperoleh pada *5-fold* selaku selanjutnya:

accuracy: 65.38%

	true Lancar	true Tidak Lancar	class precision
pred. Lancar	12	8	60.00%
pred. Tidak Lancar	1	5	83.33%
class recall	92.31%	38.46%	

Gambar 6. Hasil akurasi dan presisi percobaan ke-2

Hasil pengetesan pada Lukisan 6 angka dari ketepatan langsung dihitung memakai dengan 2 *output* tanpa mengganti informasi serta diperoleh ketepatan sebesar 65.38%. Hingga hendak dihitung dengan *confusion matrix* 2 *output* dengan 2 hasil, dengan determinasi dimana hendak dibagi jadi 2 *output* yang berlainan satu serupa lain. *Output* yang dipakai ialah Mudah dengan Tidak Mudah selaku selanjutnya:

Tabel 4 Confusion Matrix

CLASSIFICATION	KLASIFIKASI	
	Lancar	Tidak Lancar
AKTUAL	Lancar	TP=12 FN=1
	Tidak Lancar	FP=8 TN=5

$$akurasi = \frac{12 + 5}{12 + 1 + 8 + 5} * 100\% = 65.38\%$$

$$presisi = \frac{12}{12 + 8} * 100\% = 60.00\%$$

F. Hasil Uji 10-Fold

Bersumber pada sekenario percobaan pada pengetesan dengan k=10 ialah memilah *informasi training* sebesar 13 serta lebihnya jadi *informasi testing*. Sehabis seluruh informasi sudah disiapkan hingga masukan informasi kedalam *rapid miner* buat diproses. Cara dilakuka serupa dengan pengetesan *2-fold*. Ketepatan serta akurasi yang diperoleh pada *10-fold* selaku selanjutnya:

accuracy: 76.92%

	true Lancar	true Tidak Lancar	class precision
pred. Lancar	6	1	85.71%
pred. Tidak Lancar	2	4	66.67%
class recall	75.00%	80.00%	

Gambar 7. Hasil akurasi dan presisi percobaan ke-3

Hasil pengetesan pada Lukisan 7 angka dari ketepatan langsung dihitung memakai dengan 2 *output* tanpa mengganti *informasi* serta diperoleh ketepatan sebesar 76.92%. Hingga hendak dihitung dengan *confusion matrix* 2 *output* dengan 2 hasil, dengan determinasi dimana hendak dibagi jadi 2 *output* yang berlainan satu serupa lain. *Output* yang dipakai ialah Mudah dengan Tidak Mudah selaku selanjutnya:

Tabel 5 Confusion Matrix

CLASSIFICATION	KLASIFIKASI	
	Lancar	Tidak Lancar
AKTUAL	Lancar	TP=6 FN=2
	Tidak Lancar	FP=1 TN=4

$$akurasi = \frac{6 + 4}{6 + 1 + 4 + 2} * 100\% = 76.92\%$$

$$presisi = \frac{6}{6 + 1} * 100\% = 85.71\%$$

Kesimpulan dari pengujian yang sudah kita jalani dengan membagi jumlah totalitas informasi sebesar 17 kali memakai *metode cross validation* yang dipakai buat menilai kemampuan bentuk ataupun *algoritma* dimana informasi dipisahkan jadi 2 sub ialah informasi memberi pelajaran serta informasi percobaan. Angka k yang diperoleh dalam tiap pengujian menciptakan angka yang berlainan tetapi kebanyakan ada hasil yang serupa hingga diseleksi angka ketepatan serta akurasi yang tertinggi, hingga hasil dari percobaan skrip sebesar 3 kali dimana jumlah totalitas informasi dihitung 17 kali alhasil menciptakan angka ketepatan yang paling tinggi pada k -fold 10 eksperimen ke 2 ialah angka ketepatan 76.92% serta akurasi 85.71%.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Penerapan metode teknik *Bagging* pada penelitian ini dari 130 data dilakukan percobaan sebanyak 17 kali dengan pembagi *2-fold*, *5-fold*, dan *10-fold* menggunakan masing-masing data *training* dan data *testing* yang akan menghasilkan masing-masing akurasi yang berbeda yang mana akan dipilih hasil yang terbaik ada setiap percobaan.
2. Hasil yang diambil dari masing-masing percobaan ini adalah hasil dengan nilai akurasi yang tertinggi, yaitu dari *2-fold* dengan akurasi 56.92%, *5-fold* dengan akurasi tertinggi 65.38%, *10-fold* dengan akurasi tertinggi 76,92%

B. Saran

Berdasarkan penelitian yang telah dilakukan, diajukan saran untuk peneliti selanjutnya diharapkan mengembangkan penelitian dengan *platform* yang lebih canggih, bisa *website* ataupun *android* agar semakin mudah digunakan untuk seluruh kalangan.

DAFTAR PUSTAKA

- [1] I. N. Syahfitri, "Penerapan Data Mining untuk Menentukan Besar Pinjaman pada Koperasi Simpan Pinjam dengan Algoritma C4.5 (Studi Kasus: Koperasi Wanita XYZ)," *J. Inform. dan Sist. Inf.*, vol. 3, no. 2, pp. 18–27, 2017.
- [2] T. B. Santoso and D. Sekardiana, "Penerapan Algoritma C4.5 untuk Penentuan Kelayakan Pemberian Kredit (Studi Kasus : Koperia - Koperasi Warga Komplek Gandaria)," *J. Algoritm. Log. dan Komputasi*, vol. 2, no. 1, pp. 130–137, 2019, [Online]. Available: <https://journal.ubm.ac.id/index.php/alu>
- [3] F. Hadi, "Penerapan Metode Algoritma C4.5 dalam Menganalisa Pegajian Kredit pada Koperasi Jasa Keuangan Syariah Kelurahan Limau Manis Selatan," *Indones. J. Comput. Sci.*, vol. 7, no. 1, pp. 28–42, 2018, doi: 10.33022/ijcs.v7i1.58.
- [4] I. T. A. Nur, N. Y. Setiawan, and F. A. Bachtiar, "Prediksi Kredit Macet Berdasarkan Preferensi Nasabah Menggunakan Metode Klasifikasi C4.5 pada Koperasi Simpan Pinjam Mitra Raya Wates," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6118–6127, 2018.
- [5] A. Sucipto, "Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Alogaritma Klasifikasi C4.5," *DISPROTEK*, vol. 6, no. 1, pp. 75–87, 2015.
- [6] W. Ari and A. Purwarianti, "Penerapan Bagging Untuk Memperbaiki Hasil Prediksi Nasabah Perusahaan Asuransi X," *Inform. Politek. Negeri*, vol., no., pp. 1–8, 2012.
- [7] M. Rahman, "Prediksi Pembayaran Tagihan Listrik Menggunakan Model Artificial Neural Network," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 4, no. 1, p. 7, Feb. 2019, doi: 10.32528/justindo.v4i1.2417.
- [8] T. F. P. Siallagan, "Pencarian Nasabah dengan Menggunakan Data Mining dan Algoritma C4.5 Koperasi Maduma Subang," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 3, pp. 221–228, 2015, doi: 10.28932/jutisi.v1i3.399.
- [9] R. Nofitri and J. Eska, "Implementasi Data Mining Klasifikasi C4.5 Dalam Menentukan Kelayakan Pengambilan Kredit," in *Seminar Nasional Royal (SENAR)*, 2018, vol., no., p.
- [10] A. R. Wicaksono and Y. Rahayu, "Klasifikasi Data Mining Untuk Menentukan Potensi Primkoveri Waleri Menggunakan Algoritma Decision Tree C4.5," *Univ. Dian Nuswantoro Semarang*, vol., no., p., 2014.

Daryanto Meraih gelar sarjana teknik (S.Kom) dari STMIK Darmajaya Lampung pada tahun 2004. Kemudian meraih gelar Master (M.Kom) dari Sekolah Tinggi Teknik Surabaya pada tahun 2012. Saat ini Penulis menjadi dosen program studi Informatika di Universitas Muhammadiyah Jember.