

Perbandingan Algoritma C.45 dengan C.45 Berbasis Particle Swarm Optimization Untuk Menganalisa Penentuan Kelayakan Kredit BUMDES di Desa Gerung Permai Kecamatan Suralaga Kabupaten Lombok Timur

Fathurrahman¹, Muhammad Wasil²
Fakultas Teknik Universitas Hamzanwadi
fathurrahman.bloger@gmail.com¹, m.wasil@ymail.com²

Abstrak

Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga. Dari permasalahan yang ada digunakan sebuah metode sekaligus memprediksi kelayakan pemberian kredit di Desa Gerung Permai Kecamatan Suralaga yaitu dengan model algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO). Setelah di lakukan pengujian dengan dua model yaitu algoritma C4.5 dan algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO) maka hasil yang didapat adalah untuk nilai akurasi algoritma C4.5 senilai 98,86% sedangkan nilai akurasi algoritma C4.5 berbasis PSO senilai 98,89% sehingga selisih untuk nilai akurasi meningkat sebesar 0,3% dan *ROC curve* untuk keduanya yaitu, nilai AUC berdasarkan *ROC curve* untuk algoritma C4.5 bernilai 0,895, sedangkan untuk algoritma C4.5 berbasis PSO bernilai 0,917 sehingga didapatkan selisih nilai AUC meningkat sebesar 0,22 sehingga dapat disimpulkan bahwa penerapan teknik *Particle Swarm Optimization*(PSO) dapat meningkatkan nilai akurasi pada Algoritma C4.5.

Keyword : Analisa Kredit, Algoritma C4.5, *Particle Swarm Optimization* (PSO)

Abstract

Credit is the provision of money or bills that can be likened to it, based on a loan agreement between the bank and another party which requires the borrower to repay the debt after a certain period of time with interest. By the existing problems, the utilized method is a method while predicting the feasibility of giving credit in Gerung Permai village, Suralaga Subdistrict, namely Particle Swarm Optimization (PSO)-based C4.5 algorithm. After testing with two models, namely C4.5 algorithm and C4.5 algorithm on Particle Swarm Optimization (PSO)-based the results obtained are for the accuracy of C4.5 algorithm worth 98.86% while the accuracy of PSO-based C4.5 algorithm valued at 98.89% so that the difference of the accuracy value increased by 0.3% and the ROC curve for both of them, the AUC value based on the ROC curve for the C4.5 algorithm was 0.895, while for the C4.5 based PSO algorithm it was 0.917 so that the difference in value was obtained AUC increased by 0.22 so it can be concluded that the application of Particle Swarm Optimization (PSO) techniques can increase the accuracy value of the C4.5 Algorithm.

Keyword: *Credit Analysis, C4.5 Algorithm, Particle Swarm Optimized*

1. Pendahuluan

Seiring dengan laju pertumbuhan dan perkembangan bisnis, masalah perkreditan selalu menarik untuk di kaji dan diungkap karena dunia bisnis selalu bergerak, maju, dan berkembang dari waktu ke waktu. Kredit adalah pinjaman uang atau tagihan yang dapat dipersamakan dengan itu berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dan pihak lain yang mewajibkan pihak peminjam untuk melunasi imbalan, atau pembagian hasil Keuntungan. Desa Gerung Permai kecamatan Suralaga adalah salah satu desa yang menyediakan layanan kredit dan bisa dikatan kredit bumdes kepada para nasabahnya. Pengajuan kredit bisa dilakukan oleh pemohon (Calon Nasabah) yang telah memenuhi syarat dari pihak yang mengelola kredit bumdes. Pemberian kredit ini masih menjadi masalah yang krusial yg ada di desa Gerung Permai, karna pihak desa masih sangat sulit untuk menentukan pola debitur yang menyebabkan kredit lancar dan tidak lancar (Macet).

Analisa kredit merupakan hal yang penting dalam lingkup resiko keuangan oleh karena itu perlunya dilakukan analisa. Proses analisa kredit yang ada di kantor desa Gerung Permai masih membutuhkan waktu lama dalam mengidentifikasi data debitur. Analisa kredit bertujuan untuk mengevaluasi nasabah mampu atau tidak dalam memenuhi kewajiban, banyaknya kasus nasabah kredit lancar atau tidak lancar yang terjadi dapat membuat

masalah dan berdampak pada ruginya suatu desa. Penelitian ini bisa dijadikan acuan pihak desa untuk menilai nasabah yang lancar membayar kredit dan tidak lancar membayar kredit dengan melihat data yang sudah ada. Oleh karena itu, data nasabah dari Desa Gerung Permai kecamatan Suralaga menjadi bahan acuan untuk menganalisa pola nasabah kredit bumdes termasuk dalam kategori lancar dan tidak lancar. Dalam kasus ini digunakan tehnik data mining C4.5. Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan salah satu algoritma yang digunakan untuk membentuk decision tree berdasarkan training data. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal ditambah lagi dengan pengembangan yang di lakukan dengan menambahkan Particle Swarm Optimization (PSO) di harap dapat lebih membantu dalam menganalisa di bandingkan hanya menggunakan algoritma C4.5. Particle Swarm Optimization (PSO) adalah teknik optimasi statistik berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy pada tahun 1995, yang terinspirasi oleh perilaku sosial kawan burung atau ikan [1].

2. Tinjauan Pustaka

2.1. Landasan Teori

1. Kredit

Kredit berasal dari bahasa Yunani "Credere" yang berarti kepercayaan. Kredit berasal dari bahasa Latin "Creditum" yang berarti kepercayaan dalam kebenaran. Kredit adalah pinjaman uang atau tagihan yang dapat dipersamakan dengan itu berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dan pihak lain yang mewajibkan pihak peminjam untuk melunasi imbalan, atau pembagian hasil keuntungan.

2. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar^[2]. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data^[3]. Data Mining dapat diklasifikasikan menjadi dua kategori yaitu, deskriptif dan prediktif. Tugas pertambangan deskriptif mengkarakterisasi sifat umum data dalam database pertambangan prediktif, tugas data pada saat ini untuk membuat

prediksi^[4]. Kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor, antara lain^[3]:

1. Perumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang baik.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).

Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Berdasarkan definisi-definisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah :

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Hubungan yang dicari dalam data mining dapat berupa hubungan antara dua atau lebih dalam satu dimensi. Misalnya dalam dimensi produk,

dapat di lihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu, hubungan juga dapat dilihat antara dua atau lebih atribut dan dua atau lebih objek [5].

3. Algoritma C.45

Algoritma C4.5 adalah hasil dari pengembangan algoritma ID3 (Iterative Dichotomiser) yang dikembangkan oleh Quinlan (Han & Kamber, 2006). Algoritma C4.5 atau pohon keputusan mirip sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan dengan mudah dapat di konversi ke aturan klasifikasi. Secara umum keputusan pengklasifikasi pohon memiliki akurasi yang baik, namun keberhasilan penggunaan tergantung pada data yang akan diolah.

Adapun tahapan yang digunakan dalam membuat sebuah pohon keputusan menggunakan algoritma C4.5 yang ada di penelitian ini [6], yaitu :

1. Mempersiapkan data training, dapat diambil dari data historis yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah.

Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus:

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

Keterangan:

i = himpunan kasus

m = jumlah partisi i

f(i,j) = proposi j terhadap i

3. Hitung nilai gain dengan rumus:

$$Entropy\ split = - \sum_{i=1}^p \frac{n_i}{n} \cdot IE(i)$$

Keterangan:

p = jumlah partisi atribut

n_i = proporsi n_i terhadap i

n = jumlah kasus dalam n

4. Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat:
 - a. Semua tupel dalam record dalam simpul m mendapat kelas yang sama
 - b. Tidak ada atribut dalam record yang dipartisi lagi
 - c. Tidak ada record didalam cabang yang kosong.

4. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah teknik optimasi stokastik berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy pada tahun 1995, yang terinspirasi oleh perilaku sosial kawanan burung atau ikan [1].

Particle Swarm Optimization menggunakan teknik perhitungan evolusioner [7].

1. PSO diinisialisasi dengan sekumpulan solusi acak.
2. PSO mencari solusi yang optimum dengan memperbaharui generasi .
3. Perkembangan populasi berdasarkan pada generasi sebelumnya.

Dalam prosesnya, metode PSO dipengaruhi oleh sifat individu dan kelompok dalam solusi optimal. Sebagai pengganti operator evolusi, misalnya untuk persoalan optimasi d-variabel, akan disebar kawanan partikel (misalnya sebanyak n-partikel) dalam ruang d-dimensi secara acak. Masing-masing partikel merupakan kandidat solusi dan mempunyai nilai fitness tertentu. Kemudian tiap partikel akan bergerak dengan kecepatan tertentu yang dipengaruhi oleh pengalaman terbang sendiri ataupun pengalaman partikel lain.

3. Metodologi Penelitian

3.1. Pengumpulan Data

Data yang sudah dikumpulkan diolah untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Data yang tidak lengkap perlu ditangani secara sistematis agar tidak menimbulkan kesalahan pada saat pengolahan data. Jika memang tidak dapat disesuaikan maka data tersebut bisa dihilangkan. Data yang tidak konsisten muncul karena adanya perubahan aplikasi.

Dalam penelitian ini dikumpulkan sebanyak 175, data dan 7 atribut yang dapat digunakan setelah melalui beberapa tahapan dalam pengolahan data awal (preparation data).

3.2. Model yang diusulkan

Proses penelitian yang dijabarkan dimulai dari proses dataset awal sampai menghasilkan kesimpulan hasil akurasi dan performa dari dataset yang digunakan. Pada tahap ini peneliti mengusulkan Algoritma metode C4.5 Berbasis Particle Swarm Optimization (PSO) yang bertujuan untuk menghasilkan nilai akurasi yang optimal.

3.3. Evaluasi dan Validasi Hasil

Setelah mendapatkan data nasabah kredit dari kantor Desa Gerung Permai maka dilakukan pengolahan data dengan aplikasi Rapid Miner. Data yang di dapat ini terlebih dahulu di tentukan mana saja atribut yang akan digunakan untuk menentukan hasil apakah nasabah tersebut layak untuk mendapatkan kredit atau tidak. Setelah atribut-atribut yang akan digunakan sudah ditentukan selanjutnya data tersebut di olah menggunakan aplikasi Rapid Miner. Model yang digunakan dalam penelitian ini adalah algoritma C4.5 Berbasis *Particle Swarm Optimization* (PSO). Penerapan Algoritma C4.5 ini untuk menentukan nilai akurasinya. Setelah didapatkan nilai akurasi dengan menggunakan algoritma C4.5 ini selanjutnya dari data yang sudah di olah itu bisa dilihat hasil dari

pengolahan data tersebut bahwa nasabah yang ada pada data kredit bumdes ini layak untuk mendapatkan kredit.

4. Hasil Penelitian dan Pembahasan

Pengujian ini dilakukan untuk mengetahui kinerja dari metode C4.5 dalam melakukan prediksi terhadap kelas yang telah ditentukan. Dalam percobaan ini, pengujian dilakukan dengan *K-Fold Validation* dan nilai yang digunakan pada operator *number of validation* adalah 10. *K-Fold Validation* berfungsi untuk membagi jumlah antara data training dan data testing pada data yang diujikan.

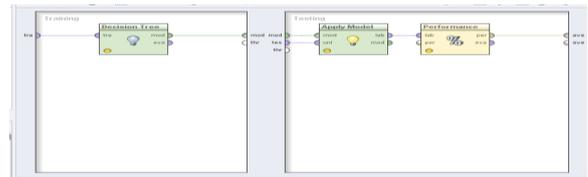
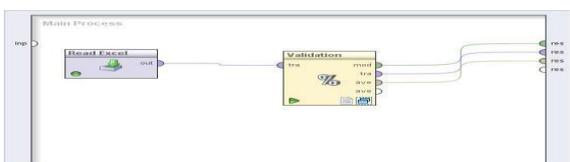
Hasil dari pengujian model berupa *accuracy*, *precision*, dan *recall*. Proses pengujian ini dilakukan *Rapidminer* dalam *building block* untuk analisa data kredit bumdes sebagai ukuran pengujian. Data tersebut akan diterapkan dengan menggunakan algoritma C4.5.

4.1 Hasil Pengujian Algoritma C.45

Hasil dari pengujian yang telah dilakukan dalam untuk mengukur tingkat akurasi dan AUC

a. Evaluasi model dengan *Confusion Matrix*

Dalam menentukan nilai tingkat keakurasi dalam model C4.5. Metode pengujiannya menggunakan cross validation dengan desain modelnya sebagai berikut :



Gambar 1. Area kerja RapidMiner untuk Menentukan hasil accuracy

Berdasarkan data training yang sudah diolah sebanyak 175 record dan diperoleh hasil sebagai berikut :

accuracy: 98.86% +/- 1.40% (mikro: 98.86%)			
	true Lancar	true Tidak Lancar	class precision
pred. Lancar	109	0	100.00%
pred. Tidak Lancar	2	64	96.97%
class recall	99.20%	100.00%	

Gambar 2. Nilai akurasi algoritma C.45

Model Confusion Matrix akan membentuk matrix yang terdiri dari jumlah True Positive (TP) adalah 109, False Negative (FN) adalah 0, untuk False Positive (FP) adalah 2, dan untuk True Negative adalah 64. Dari confusion matrix di atas, terlihat bahwa akurasi dengan menggunakan algoritma C4.5 adalah sebesar 98,86%. Berikut adalah perhitungan akurasi, sensitivity, specificity, ppv, dan npv.

$$acc = \frac{tp + tn}{tp + tn + fp + fn} = \frac{109 + 64}{109 + 64 + 2 + 0} = 0.988$$

$$Sensitivity = \frac{tp}{tp + fn} = \frac{109}{109 + 0} = 1$$

$$Specificity = \frac{tn}{tn + fp} = \frac{64}{64 + 2} = 0.969$$

$$PPV = \frac{tp}{tp + fp} = \frac{109}{109 + 2} = 0.981$$

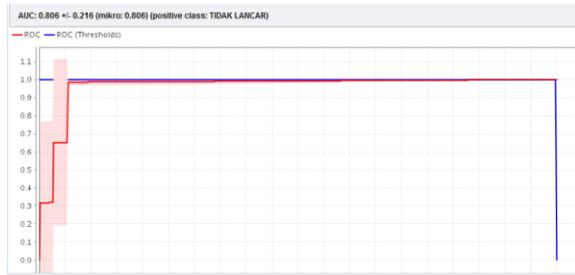
$$NPV = \frac{tn}{tn + fn} = \frac{64}{64 + 0} = 1$$

Hasil perhitungan dari persamaan di atas terlihat pada tabel 1. Tabel Nilai Accuracy, Sensitivity, Specitivity, Ppv dan Npv.

Accuracy	0,988
Sensitivity	1
Specificity	0,969
PPV	0,981
NPV	1

b. Evaluasi ROC Curve

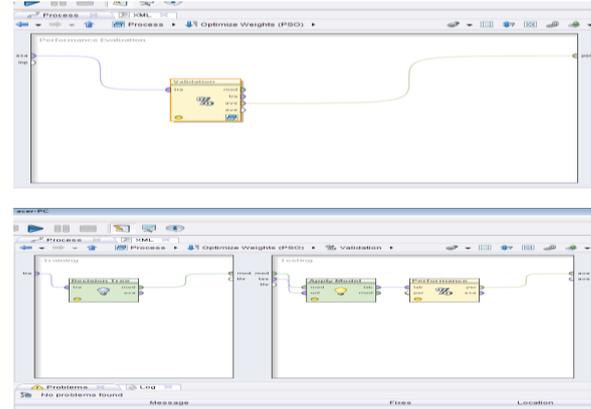
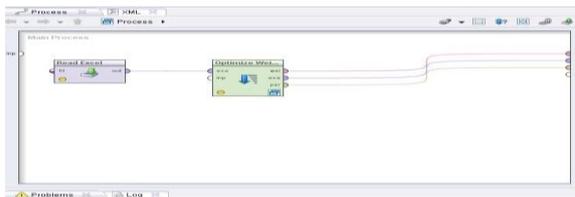
Dari gambar 3. terdapat grafik ROC Dengan nilai AUC sebesar 0.806 dengan nilai akurasi for classification.



Gambar 3. Nilai AUC dalam grafik ROC algoritma C4.5

4.2 Hasil Pengujian Algoritma C.45 Berbasis Particle Swarm Optimization

Dalam menentukan nilai tingkat keakurasian dalam model C4.5 berbasis Particle swarm optimization. Metode pengujiannya menggunakan cross validation dengan desain modelnya sebagai berikut :



Gambar 4. Tampilan Area Kerja Rapidminer dengan C4.5berbasis PSO

Hasil dari pengujian yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC.

a. Confusion Matrix

Berdasarkan data training yang sudah diolah sebanyak 175 record dan di peroleh hasil sebagai berikut :

accuracy: 98.89% +/- 2.48% (mikro: 98.89%)			
	true Lancar	true Tidak Lancar	class precision
pred. Lancar	109	0	100.00%
pred. Tidak Lancar	2	64	96.97%
class recall	98.20%	100.00%	

Gambar 5. Nilai akurasi Algoritma C4.5 Berbasis PSO

Model Confusion Matrix akan membentuk matrix yang terdiri dari jumlah *True Positive* (TP) adalah 109, *False Negative* (FN) adalah 0, untuk *False Positive* (FP) adalah 2, dan untuk *True Negative* adalah 64. Berdasarkan gambar 4.16, tingkat akurasi dengan menggunakan algoritma C4.5 berbasis PSO dengan *Cross Validation* adalah sebesar 98,89%, dan untuk menghitung nilai *accuracy*, *sensitivity*, *specificity*, *ppv* dan *npv* hasilnya pada persamaan di bawah ini :

$$acc = \frac{tp + tn}{tp + tn + fp + fn} = \frac{109 + 64}{109 + 64 + 2 + 0} = 0.988$$

$$Sensitivity = \frac{tp}{tp + fn} = \frac{109}{109 + 0} = 1$$

$$Specificity = \frac{tn}{tn + fp} = \frac{64}{64 + 2} = 0.969$$

Tabe 2. Nilai *Accuracy*, *Sensitivity*, *Specitivity*, *Ppv* dan *Npv*

<i>Accuracy</i>	0,988
<i>Sensitivity</i>	1
<i>Specificity</i>	0,969
<i>PPV</i>	0,981
<i>NPV</i>	1

b. Evaluasi ROC Curve

Dari gambar 6. terdapat grafik ROC dengan nilai AUC sebesar 0,917 dengan nilai akurasi Excellent Classification.



Gambar 6. Nilai AUC dalam grafik ROC algoritma C4.5 Berbasis *Particle Swarm Optimization*

5. Kesimpulan

Penelitian yang telah dilakukan di atas dengan melakukan evaluasi menggunakan *confusion matrix* dan *ROC curve*, Berdasarkan hasil penelitian yang telah dilakukan, algoritma C4.5 berbasis PSO bisa dimanfaatkan oleh seorang

analisis untuk membantu dalam menentukan layak tidaknya seorang nasabah mendapatkan kredit serta memprediksi apakah nasabah tersebut akan macet atau lancar dalam membayar kredit. Dalam penelitian ini dapat disimpulkan bahwa data kredit atau hasil penelitian untuk nilai algoritma C4.5 senilai 98,86% sedangkan nilai akurasi algoritma C4.5 berbasis PSO (*Particle Swarm Optimization*) senilai 98,89% sehingga selisih untuk nilai akurasi meningkat sebesar 0.3% dan untuk nilai AUC untuk algoritma C4.5 senilai 0,895 dan untuk algoritma C4.5 berbasis PSO (*Particle Swarm Optimization*) 0,917 yang berarti nilai AUC nya meningkat 0,22. 0,700% dengan for clasification

6. Daftar Pustaka

[1]. X Hu, R. Eberhart, and Y. Shi. *Recent advances in particle swarm*, IEEE Congress on Evolutionary Computation 2004, Portland, Oregon, USA.

[2]. Turban, E., dkk. 2005. *Decicion Support System and intelligent Systems*. Yogyakarta: Andi Ofset

[3]. Larose , Daniel T, 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Willey & Sons. Inc

[4]. Han, J., & Kamber, M. (2007). *Data Mining Concepts and Technique*. Morgan Kaufmann publisher

[5]. Ponniah, P., 2001, *Datawarehouse Fundamentals : A comprehensive Guide*

- for IT Professional, John Willey & Sons. Inc
- [6]. F. Gorunescu, (2011). Data Mining: Concepts, Models and Techniques. Berlin: Springer-Verlag Berlin Heidelberg
- [7]. X. Hu, R. Eberhart, and Y. Shi. *Particle swarm with extended memory for multiobjective optimization*, IEEE Swarm Intelligence Symposium 2003, Indianapolis, IN, USA.
- [8]. Firmansyah (2011). Penerapan Algoritma Klasifikasi C4.5 untuk Penentuan Kelayakan Pemberian Kredit Koperasi
- [9]. Sigit Abdullah (2012). Penerapan Algoritma Decision Tree C4.5 untuk Diagnosa Penyakit Stroke dengan Klasifikasi Data Mining pada Rumah Sakit Santa Maria Pemalang
- [10]. Siti Masripah (2011). Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization untuk Evaluasi Penentuan Kelayakan Kredit Koperasi Syariah
- [11]. Khafi Heryandi Suradiradja (2012). Deteksi Transaksi Pencucian Uang Dengan Algoritma Klasifikasi C4.