

PENERAPAN PARTICLE SWARM OPTIMIZATION (PSO) UNTUK KLASIFIKASI DAN ANALISIS KREDIT DENGAN MENGGUNAKAN ALGORITMA C4.5

Saprudin

Teknik Informatika Universitas Pamulang
Jl. Surya Kencana No. 1, Pamulang, Tangerang Selatan-Indonesia
E-mail : dosen00836@unpam.ac.id

ABSTRAK

Teknik klasifikasi saat ini sudah banyak digunakan untuk mengatasi permasalahan yang terkait dengan penggolongan data. Implementasi teknik klasifikasi ini dapat diterapkan pada berbagai bidang. Tujuan utama dari metode klasifikasi adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data. Salah satu metode yang paling baik untuk klasifikasi adalah algoritma C4.5. Untuk itu penulis menggunakan metode klasifikasi C4.5, dan untuk dapat meningkatkan nilai akurasi dioptimasi dengan PSO (Particle Swarm Optimization) dimana data uji yang digunakan adalah Data German Credit. Dari hasil percobaan yang dilakukan model algoritma C4.5 berbasis Particle Swarm Optimization (PSO) mendapatkan hasil terbaik yaitu 70%, sedangkan model algoritma C4.5 tanpa Particle Swarm Optimization (PSO) hanya 68,6%, sehingga selisih untuk nilai akurasi sebesar 1,4%. Dapat disimpulkan bahwa penerapan teknik optimasi particle swarm optimization dapat meningkatkan nilai akurasi pada algoritma C4.5.

Kata kunci: Klasifikasi, Optimasi, Algoritma C4.5

1. PENDAHULUAN

Analisis kredit adalah sistem atau cara yang digunakan oleh Bank atau lembaga pembiayaan lainnya untuk menentukan kelayakan calon debitur untuk mendapatkan pinjaman. Penelitian ini menggunakan *German Credit Dataset* yang diperoleh dari *UCI Machine Learning Repository*, data tersebut merupakan data debitur yang telah melakukan kredit pada Bank Jerman yang terdiri dari 1000 record dengan 20 variabel.

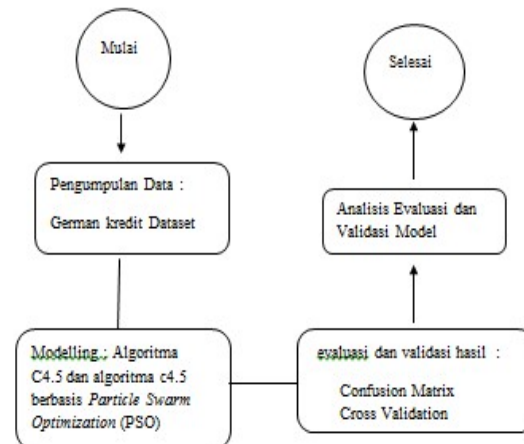
Algoritma klasifikasi akan menggunakan data latihan untuk menghasilkan pengetahuan agar dapat menggolongkan resiko kredit seorang nasabah pada masa mendatang berdasarkan variabel-variabel yang ada.

Analisa dengan metode klasifikasi sendiri adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006).

Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2006).

2. METODE PENELITIAN

2.1 Analisa dan Desain Penelitian



Gambar 2.1 Alur Analisis Penelitian

2.2 Business understanding

Banyaknya kasus nasabah kredit macet yang terjadi dapat membuat masalah dan berdampak pada ruginya suatu bank, oleh karena itu di butuhkan analisa yang tepat sebagai solusinya. Algoritma c4.5 adalah salah satu metode klasifikasi yang sering di gunakan untuk menganalisa kredit dengan tujuan agar analisa lebih akurat, ditambah lagi dengan pengembangan yang dilakukan dengan menambahkan *Particle*

Swarm Optimization (PSO) di harap dapat lebih membantu dalam menganalisa di bandingkan hanya menggunakan algoritma c4.5 saja.

2.3 Data Understanding

Penelitian ini nantinya akan menggunakan data uji yang banyak dipakai dalam permasalahan klasifikasi untuk penilaian kelayakan pemberian kredit yang disebut *German Credit Dataset*, data set ini didonasikan oleh Prof. Hofman dari Hamburg University, Jerman. *Dataset* ini terdiri dari 1000 record dan 20 variabel ditambah dengan sebuah variabel target atau variabel response, dimana 13 variabel diantaranya bertipe kategori dan sisanya sebanyak 7 variabel bertipe numerik. *German Credit Dataset* ini dapat diunduh di *UCI Machine Learning Repository*.

2.4. Data Preparation

(Sugiyono, 2002) mengemukakan bahwa populasi adalah wilayah yang terdiri atas obyek atau subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya.

Pada tahapan ini di pakai data sebanyak 1000 record dan 20 variabel ditambah dengan sebuah variabel target atau variabel response, sebelumnya akan dilakukan beberapa penyeleksian untuk menghasilkan data yang dibutuhkan, antara lain:

1. *Data Cleaning* untuk membersihkan nilai yang kosong atau tupel yang kosong..
2. *Data Integration* yang berfungsi menyatukan tempat penyimpanan yang berbeda kedalam satu data. Dalam kasus ini hanya ada satu tempat penyimpanan data yaitu status di terima atau tidaknya kredit nasabah.
3. *Data reduction* jumlah atribut yang digunakan

3. HASIL DAN PEMBAHASAN

3.1 Pengukuran Penelitian

3.1.1. Hasil Penelitian

Mengembangkan model yang sudah terbentuk dengan algoritma klasifikasi C4.5.serta melakukan perbandingan menggunakan Algoritma Klasifikasi C4.5 dan Algoritma Klasifikasi C4.5 berbasis *Particle Swarm Optimization*.

3.1.2. Evaluasi dan Validasi Model

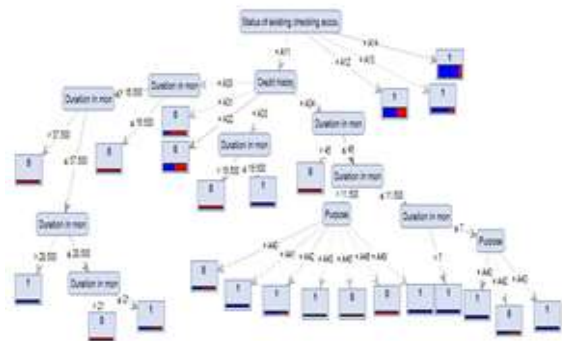
1. Hasil Perhitungan Entropy dan Gain

Contoh perhitungan sederhana memakai tiga atribut :

Tabel 3.1 perhitungan node dengan tiga atribut

NO			JUMLAH KASUS	YES	NO	ENTROPY	GAIN
1	TOTAL		1000	700	300	0.881290899	
		Status of existing checking account					0.095463104
		A 11	274	139	135	0.999846263	
		A 12	268	163	105	0.965945651	
		A 13	62	48	14	0.770629069	
		A 14	396	350	46	0.518225966	
		Credit history					0.043617799
		A 30	40	15	25	0.954434003	
		A 31	49	21	28	0.985228136	
		A 32	530	361	169	0.903146117	
		A 33	88	60	28	0.902393283	
		A 34	293	243	50	0.659186043	
		Purpose					0.023058443
		A 40	233	145	88	0.956388718	
		A41	102	85	17	0.650022422	
		A 42	181	123	58	0.904861457	
		A 43	280	218	62	0.762767467	
		A 44	0	0	0	0	
		A 45	22	14	8	0.945660305	
		A 46	50	28	22	0.989587521	
		A 47	0	0	0	0	
		A 48	23	17	6	0.828055725	
		A 49	97	63	34	0.934527566	
		A410	12	7	5	0.979868757	

Dari tabel 3.1 di atas dapat diketahui bahwa atribut dengan gain tertinggi adalah *Status of existing checking account* yaitu sebesar 0,095463104. Dengan demikian *Status of existing checking account* dapat njadi akar.berikut keterangan dengan gambar pohon keputusannya.



Gambar 3.1 pohon keputusan tiga atribut

Rincian aturan pohon keputusan pada gambar 4.1 di atas adalah sebagai berikut :

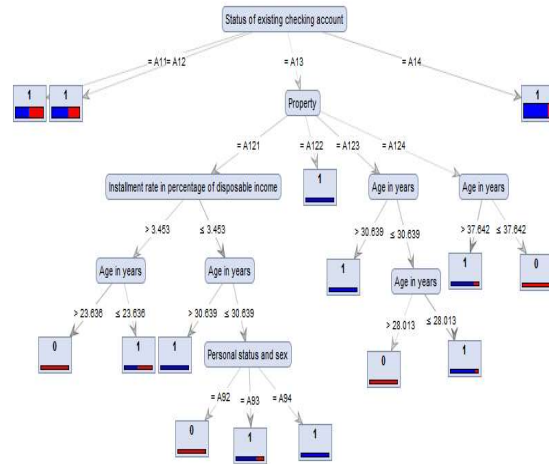
1. R1 = IF *Status of existing checking account* A11, AND *Credit Histori* A30, AND *Duration in month* >16.500, AND *Duration in month* >37.500, THEN class = NO
2. R2 = IF *Status of existing checking account* A11, AND *Credit Histori* A30, AND *Duration in month* >16.500, AND *Duration in*

- month <=37.500, AND Duration in month >28.500 THEN class = YES*
3. R3 = *IF Status of existing checking account A11, AND Credit Histori A30, AND Duration in month >16.500, AND Duration in month <=37.500, AND Duration in month <=28.500, AND Duration in month >21, THEN class=NO*
 4. R4 = *IF Status of existing checking account A11, AND Credit Histori A30, AND Duration in month >16.500, AND Duration in month <=37.500, AND Duration in month <=28.500, AND Duration in month <=21, THEN class=YES*
 5. R5 = *IF Status of existing checking account A11, AND Credit Histori A30, AND Duration in month <=16.500, THEN class = NO*
 6. R6 = *IF Status of existing checking account A11, AND Credit Histori A31, THEN class = NO*
 7. R67= *IF Status of existing checking account A11, AND Credit Histori A32, THEN class = NO*
 8. R8 = *IF Status of existing checking account A11, AND Credit Histori A33, AND Duration in month >19.500, THEN class = NO*
 9. R9 = *IF Status of existing checking account A11, AND Credit Histori A33, AND Duration in month <=19.500, THEN class = YES*
 10. R10 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, THEN class = NO*
 11. R11 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A40, THEN class = NO*
 12. R12 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A41, THEN class = YES*
 13. R13 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A42, THEN class = YES*
 14. R11 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A43, THEN class = YES*
 15. R15 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A44, THEN class = NO*
 16. R16 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A46, THEN class = NO*
 17. R17= *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month >45, AND Duration in month >11.500, AND purpose A49, THEN class = YES*
 18. R18 = *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month <=45, AND Duration in month <=11.500, AND Duration in month >7, THEN class = YES*
 19. R19= *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month <=45, AND Duration in month <=11.500 AND Duration in month <=7, , AND purpose A40, THEN class = YES*
 20. R20= *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month <=45, AND Duration in month <=11.500 AND Duration in month <=7, , AND purpose A42, THEN class = NO*
 21. R21= *IF Status of existing checking account A11, AND Credit Histori A34, AND Duration in month <=45, AND Duration in month <=11.500 AND Duration in month <=7, , AND purpose A43, THEN class = YES*
 22. R22= *IF Status of existing checking account A12, THEN class = YES*
 23. R23= *IF Status of existing checking account A13, THEN class = YES*
R24= *IF Status of existing checking account A14, THEN class = YES*

Tabel 3.2 Perhitungan Node Sepenuhnya

No			Jumlah kasus	Yes	No	Entropy	Gain
1	Total		1000	700	300	0.881290899	
	Status of existing checking account						0.095463104
	A 11		274	139	135	0.999846263	
	A 12		268	163	105	0.965945651	
	A 13		62	48	14	0.770629069	
	A 14		396	350	46	0.518225966	
	Credit history						0.043617799
	A 30		40	15	25	0.954434003	
	A 31		49	21	28	0.985228136	
	A 32		530	361	169	0.903146117	
	A 33		88	60	28	0.902393283	
	A 34		293	243	50	0.659186043	
	Purpose						0.023058443
	A 40		233	145	88	0.956388718	
	A 41		102	85	17	0.650022422	
	A 42		181	123	58	0.904861457	
	A 43		280	218	62	0.762767467	
	A 44		0	0	0	0	
	A 45		22	14	8	0.945660305	
	A 46		50	28	22	0.989587521	
	A 47		0	0	0	0	
	A 48		23	17	6	0.828055725	
	A 49		97	63	34	0.934527566	
	A 410		12	7	5	0.979868757	
	Savings account/ bonds						0.029204832
	A 61		603	386	217	0.942573008	
	A 62		103	69	34	0.915025536	
	A 63		63	52	11	0.668127334	
	A 64		47	41	6	0.550984443	
	A 65		184	153	32	0.660209902	
	Present employment since						0.013102323
	A 71		62	39	23	0.951412254	
	A 72		172	102	70	0.974885708	
	A 73		339	235	104	0.88942697	
	A 74		174	135	39	0.767651587	
	A 75		253	189	64	0.81594291	
	Personal status and sex						0.00681055
	A 91		50	30	20	0.970950594	
	A 92		310	201	109	0.935500272	
	A 93		548	402	146	0.836285431	
	A 94		92	67	25	0.843949124	
	A 95		0	0	0	0	
	Other debtors / guarantors						0.004797021
	A 101		907	635	272	0.881156084	
	A 102		41	23	18	0.989245297	
	A 103		52	10	42	0.706274089	
	Property						0.016985186
	A 121		282	222	60	0.746736607	
	A 122		232	161	71	0.888542793	
	A 123		332	230	102	0.889948779	
	A 124		154	87	67	0.987799139	
	Other installment plans						0.00887507
	A 141		139	82	57	0.97653823	
	A 142		47	28	19	0.973385435	
	A 143		814	590	224	0.848805773	
	Housing						0.012753186
	A 151		179	109	70	0.965481126	
	A 152		713	527	186	0.828055725	
	A 153		108	64	44	0.975119065	
	Job						0.001411686
	A 171		23	16	7	0.886540893	
	A 172		200	144	56	0.855450811	
	A 173		630	444	186	0.875391854	
	A 174		147	96	51	0.931304369	
	Telephone						0.015513013
	A 191		586	409	187	0.887951186	
	A 192		404	291	113	0.855045771	
	foreign worker						0.005822991
	A 201		963	667	296	0.890117487	
	A 202		37	33	4	0.494182935	

Pohon keputusan algoritma c4.5 berbasis Particle Swarm Optimization (PSO) dengan minimal gain 0.1



Gambar 4.2 pohon keputusan algoritma c4.5 berbasis Particle Swarm Optimization(PSO)

Rincian aturan pohon keputusan pada gambar 4.3 di atas adalah sebagai berikut :

1. R1 = IF Status of existing checking account A11, THEN CLASS = YES
2. R2 = IF Status of existing checking account A12, THEN CLASS = YES
3. R3 = IF Status of existing checking account A13, AND property A121, AND Installment rate in percentage of disposable income > 3.453, AND age in years >23.636, THEN CLASS = NO
4. R4 = IF Status of existing checking account A13, AND property A121, AND Installment rate in percentage of disposable income > 3.453, AND age in years <= 23.636, THEN CLASS = YES
5. R5 = IF Status of existing checking account A13, AND property A121, AND Installment rate in percentage of disposable income > 3.453, AND age in years >30.639, THEN CLASS = NO
6. R6 = IF Status of existing checking account A13, AND property A121, AND Installment rate in percentage of disposable income > 3.453, AND age in years <=30.639, AND personal status and sex A 92, THEN CLASS = NO
7. R7= IF Status of existing checking account A13, AND property A121, AND Installment rate in percentage of disposable income > 3.453, AND age in years <=30.639, AND personal status and sex A 93, THEN CLASS = YES

8. R8= IF Status of existing checking account A13, AND property A121, AND Installment rate in percentage of disposable income > 3.453, AND age in years <=30.639, AND personal status and sex A 94, THEN CLASS = YES
9. R9 = IF Status of existing checking account A113, AND property A122, THEN CLASS = YES
10. R10 = IF Status of existing checking account A13, AND property A123, AND age in years >30.639, THEN CLASS = YES
11. R11= IF Status of existing checking account A13, AND property A123, AND age in years <=30.639, AND age in years >28.013 THEN CLASS = NO
12. R12 = IF Status of existing checking account A13, AND property A123, AND age in years <=30.639, AND age in years <=28.013 THEN CLASS = YES
13. R13= IF Status of existing checking account A13, AND property A124, AND age in years >37.642, THEN CLASS = YES
14. R14 = IF Status of existing checking account A13, AND property A124, AND age in years <=37.642, THEN CLASS = NO
15. R15 = IF Status of existing checking account A14, THEN CLASS =YES

2. Hasil Pengujian

a. Evaluasi model dengan Confusion Matrix

Tabel 3.4 Konversi confusion matrix algoritma klasifikasi C4.5 berbasis PSO

accuracy: 70.00% +/- 0.00% (mikro: 70.00%)			
	true 1	true 0	class precision
pred. 1	700	300	70.00%
pred. 0	0	0	0.00%
class recall	100.00%	0.00%	

Pada Confusion Matrix di atas, Actual/ sebenarnya ada 1000 predikat 1(positive), namun system memprediksi ada 700 predikat 1(positive) , 300 predikat 0 (negative), 0 predikat 0(positif), dan 0 predikat 0(negative).

$$\begin{aligned} \text{Accuracy} &= \frac{\text{true 1} + \text{true 0}}{\text{True 1} + \text{true 0} + \text{false 1} + \text{false 0}} \\ &= \frac{700 + 0}{700 + 0 + 0 + 300} \\ &= 0,7 * 100\% = 70\% \end{aligned}$$

$$\text{Error rate} = \frac{300}{1000} * 100\% = 30\%$$

Berdasarkan Tabel 3.4 menunjukkan bahwa, tingkat akurasi menggunakan algoritma klasifikasi C4.5 berbasis PSO adalah sebesar 70% dan error ratenya sebesar 30%.

3.2 Analisis Evaluasi dan Validasi Model

Dari hasil pengujian diatas, dengan dilakukannya evaluasi secara *confusion matrix* ternyata terbukti bahwa pengujian yang dilakukan dengan menggunakan algoritma klasifikasi C4.5 berbasis PSO memiliki nilai akurasi yang lebih tinggi dibanding hanya menggunakan algoritma klasifikasi C4.5.

Nilai akurasi untuk model algoritma klasifikasi C4.5 sebesar 68,6 % dan nilai akurasi algoritma klasifikasi C4.5 berbasis PSO (*Particle Swarm Optimization*) sebesar 70% dengan selisih akurasi sebesar 1,4%, dapat dilihat pada Tabel 3.3 dibawah ini.

Tabel 3.3 Pengujian Algoritma klasifikasi C4.5 dan C4.5 berbasis PSO

Accuracy	
C4.5	68,6 %
C4.5 berbasis PSO	70%

4. KESIMPULAN

Hasil penelitian untuk nilai akurasi algoritma klasifikasi C4.5 senilai 68,6 % sedangkan untuk nilai akurasi algoritma klasifikasi C4.5 berbasis PSO senilai 70% sehingga selisih untuk nilai akurasi sebesar 1,4 %. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi *particle swarm optimization* dapat meningkatkan nilai akurasi pada algoritma C4.5

5. SARAN

Berdasarkan proses pengujian dan kesimpulan yang telah dilakukan, maka ada beberapa saran dalam penelitian ini adalah:

1. Menambahkan jumlah data dan atribut lebih banyak lagi, sehingga bisa menentukan keakuratan suatu metode dengan lebih baik lagi.
2. Menggunakan metode prediksi dan optimasi lain sehingga dapat memperkaya pengetahuan khususnya di bidang metode datamining.
3. Menggunakan program software rapiminer yang lebih terbaru.
4. Meningkatkan lagi sistem analisa kredit untuk penentuan kelayakan pemberian kredit

DAFTAR PUSTAKA

- [1] Sugiyono. (2002). *Metode Penelitian Bisnis*. Bandung: CV. Alfabeta.
- [2] Olson, David dan Yong Shi (2007) *Introduction to Business Data Mining*, Boston:McGraw-Hill.
- [3] Sulianta, Feri dan Dominikus Juju (2010) *Data Mining Meramalkan Bisnis Perusahaan*, Jakarta: Elex Media Komputindo.
- [4] Hair, Joseph, et al (1995.) *Multivariate Data Analysis with Readings*, Englewood Cliffs: Prentice
- [5] Hall.Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques*. San Francisco: Diane Cerra.
- [6] Larose, D. T. (2005). *Discovering Knowledge In Data*. Canada: WileyInterscience.
- [7] Fayyad, Usama. (1996) *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [8] X. Hu, R. Eberhart, and Y. Shi. *Particle swarm with extended memory for multiobjective optimization*, IEEE Swarm Intelligence Symposium 2003, Indianapolis, IN, USA
- [9] X. Hu, R. Eberhart, and Y. Shi. *Recent advances in particle swarm*, , IEEE Congress on Evolutionary Computation 2004, Portland, Oregon, USA
- [10] Gorunescu, F. (2011). *Data Mining Concepts, Model and Techniques*. Berlin: Springer.