

Penambahan Pengklasifikasi Fuzzy dengan *Multiobjective Evolutionary Fuzzy Classifier*

Nur Salman¹, Mustikasari^{2*}, Muhammad Nur Akbar³

¹Jurusan Teknik Informatika, Universitas DIPA Makassar, Indonesia

^{2,3}Jurusan Teknik Informatika, Universitas Islam Negeri Alauddin Makassar, Indonesia

¹nursalman.halim@undipa.ac.id, ²mustikasalman@uin-alauddin.ac.id, ³muhnurakbar@uin-alauddin.ac.id

Informasi Artikel

Article history:

Diterima Jan 10, 2022

Revisi Jan 17, 2022

Dipublis Jan 30, 2022

Kata Kunci:

Evolutionary

Fuzzy

Multiobjective

Classifier

ABSTRACT

Classification is one of the key issues in the field of data mining and knowledge discovery. This paper implements a method of constructing a fuzzy rule mining classifier, which is extended in the context of classification. There are three stages of this approach: fuzzy rule set extraction, second; a linguistic labeling process that assigns a linguistic label to each fuzzy set. Owing to many attributes in the database, the feature selection process is also carried out, reducing the complexity to build the final classifier. Third: incorporate strategies to avoid rule redundancy and conflict into process mining. We applied the application Multiobjective Evolutionary Fuzzy Classifier (MOFC), which produced a classifier with satisfactory classification accuracy compared to other classifiers such as C4.5. In addition, in terms of classification based on association rules, MOFC can filter the large of rules and be proven to be able to build compact fuzzy models while maintaining a very good level of accuracy and producing a much smaller set of rules. We examine the performance of fuzzy rule classifiers through computational experiments on three benchmark data sets in the Kaggle and UCI machine learning repository.

*Koresponden Author:

Mustikasari,

Jurusan Teknik Informatika.

Universitas Islam Negeri Alauddin Makassar,

Jl. H.M. Yasin Limpo No. 36 Samata, Kab Gowa, Sulawesi Selatan, Indonesia.

Email: mustikasalman@uin-alauddin.ac.id



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

1. PENDAHULUAN

Model pengklasifikasi yang dapat ditafsirkan memungkinkan validasi eksternal oleh seorang ahli dalam disiplin ilmu tertentu seperti kedokteran, farmasi, pertanian atau bisnis dan memberikan informasi tentang pengambilan keputusan yang sangat penting untuk mengurangi subjektifitas hasil. Sistem fuzzy [1] menjadi salah satu pilihan yang cukup diperhitungkan dan sukses diimplementasikan karena kemampuan dan karakteristiknya dalam menangani ketidakpastian dan

menjelaskan fitur sistem canggih tanpa melibatkan model matematika tertentu. Sistem klasifikasi berbasis aturan fuzzy [2] [3] telah sangat berkembang dalam beberapa tahun terakhir [4], dan sistem tersebut sekarang mengkonsolidasikan alat klasifikasi yang kuat yang juga memungkinkan interpretasi model secara langsung dan jelas karena sistemnya menggunakan linguistik [5] dan sejalan dengan masalah dunia nyata karena sifatnya yang kontinu.

Beberapa target penting pada suatu sistem klasifikasi tentu menjadi pilihan penerapannya. Seperti interpretabilitas dan akurasi. Interpretabilitas sistem klasifikasi mengacu pada kemampuan model untuk mengekspresikan perilaku mereka dengan cara yang mudah dimengerti oleh pengguna. Namun, penggunaan logika fuzzy mungkin tidak cukup untuk membuat model klasifikasi dapat diinterpretasikan atau untuk mencapai beberapa tujuan lain yang diharapkan sebagai hasil. Meskipun sistem klasifikasi berbasis aturan fuzzy dirancang untuk menghasilkan model yang dapat diinterpretasikan; namun, dengan adanya sejumlah besar atribut, pengklasifikasi yang dihasilkan mungkin terlalu rumit untuk ditafsirkan dengan mudah (misalnya, aturan dengan lebih dari enam atribut mungkin tidak dapat diterapkan oleh manusia). Sementara itu, di sisi lain, akurasi juga menjadi parameter yang tidak kalah pentingnya [6]. Kemudian sebuah kenyataan juga hadir mengikuti desain sistem fuzzy bahwa interpretasi dan akurasi adalah sebuah kontradiksi atau bertentangan satu sama lain, yaitu, peningkatan salah satu dari dua opsi menyebabkan penurunan yang lain.

Algoritma evolusioner adalah teknik optimasi stokastik, yang mampu mensimulasikan konsep evolusi alami. Algoritma ini mampu mengatasi masalah terkait ruang pencarian yang besar, kompleks, dan berdimensi tinggi, dan pencapaian berbagai tujuan yang saling bertentangan. Dalam teknik ini, tidak ada solusi optimal, ideal dan tunggal yang dapat diturunkan, sebaliknya, serangkaian solusi dihasilkan karena peningkatan dalam satu tujuan menyebabkan degradasi pada tujuan yang tersisa. Solusi ini dikenal dengan istilah Solusi Pareto-Optimal [7].

Pembahasan dalam makalah ini selanjutnya dibagi menjadi empat bagian. Bagian awal dimana metode MOFC diperkenalkan secara singkat. Bagian 2, memaparkan tentang metode MOFC secara rinci. Bagian 3, berisi pembahasan yang luas tentang hasil yang diperoleh pada penerapan MOFC yang mencakup isu-isu yang berkaitan dengan interpretabilitas serta akurasi. Bagian 4, membahas isu-isu penelitian masa depan yang berkaitan dengan MOFC yang termasuk dalam kesimpulan dan ruang lingkup masa depan.

2. METODE PENELITIAN

Multiobjective Evolutionary Fuzzy Classifier membangun sebuah classifier berbasis aturan fuzzy dengan mengoptimalkan dua tujuan. Yang pertama dapat dikonfigurasi untuk memaksimalkan akurasi, untuk memaksimalkan area di bawah kurva ROC, atau untuk meminimalkan *root mean square error*. Yang kedua adalah meminimalkan jumlah aturan fuzzy dari classifier.

2.1. Skema Umum Pengklasifikasi

Metodologi yang diterapkan pada MOFC adalah proses yang mengikuti tiga langkah berikut yaitu: (1) optimasi multiobjective constrained dari kumpulan dataset, menggunakan algoritma evolusi multiobjective berbasis Pareto untuk memaksimalkan akurasi dan meminimalkan kompleksitas (jumlah aturan) pengklasifikasi, yang tunduk pada batasan interpretasi; langkah ini menghasilkan satu set pengklasifikasi alternatif (Pareto); (2) pelabelan linguistik, yang memberikan label linguistik untuk setiap himpunan fuzzy dari pengklasifikasi; langkah ini penting untuk interpretasi pengklasifikasi; (3) pengambilan keputusan, dimana pengklasifikasi dipilih, jika memuaskan, sesuai dengan preferensi pembuat keputusan.

2.2. Sistem Fuzzy Evolusioner

Algoritma evolusioner adalah teknik optimasi stokastik, yang mensimulasikan konsep evolusi alami. Pendekatan evolusioner terdiri dari metodologi, algoritma genetika, pemrograman evolusioner, dan strategi evolusioner. Dalam Evolutionary Multi-Objective Optimization, tujuan saling bertentangan. Dalam teknik ini, tidak ada solusi optimal, ideal, dan tunggal yang dapat

diturunkan, sebaliknya, serangkaian solusi dihasilkan karena peningkatan dalam satu tujuan menyebabkan degradasi pada tujuan yang tersisa. Solusi ini disebut Solusi Pareto-Optimal. Solusi Optimal Pareto ini dalam hal fungsi tujuan disebut Pareto Front.

Misalnya, l (jumlah) masalah maksimisasi objektif dapat dirumuskan dalam persamaan 1, yaitu:

$$\text{Maximize } g_i(x) = (g_1(x), \dots, g_l(x), \quad i = 1, \dots, l \quad (1)$$

Solusi Pareto-Optimal adalah solusi yang tidak didominasi oleh solusi lain, dan Pareto Front dari setiap masalah adalah himpunan semua Solusi Pareto-Optimal dalam hal fungsi tujuan. Misalnya, akurasi dan kompleksitas dengan tujuan maksimalisasi akurasi dan maksimalisasi interpretabilitas (minimalisasi kompleksitas aturan). Beberapa issue terkait muncul dalam rangka menghindari dominasi setiap tujuan yang ingin dicapai. Bagaimanapun, issue tersebut dapat diartikan setara dengan usaha meminimalkan error dan meminimalkan kompleksitas dari aturan yang tercipta, yang dituliskan dalam persamaan 2 berikut ini:

$$\{\text{Minimize } (f_{\text{error}}(M), f_{\text{complexity}}(M))\} \quad (2)$$

Sebagai usaha mendapatkan tujuan ini, beberapa kriteria, seperti jumlah aturan fuzzy yang dipilih, jumlah aturan yang diklasifikasikan dengan benar, penyetelan keanggotaan, granularitas partisi seragam, dan lainnya turut dipertimbangkan. Dengan demikian, strategi pencarian yang dipilih adalah mengidentifikasi solusi yang tidak didominasi untuk masalah optimasi multi-tujuan, yang dapat dirumuskan sebagai contoh dari masalah (l) dengan $l = 2$ (dua tujuan).

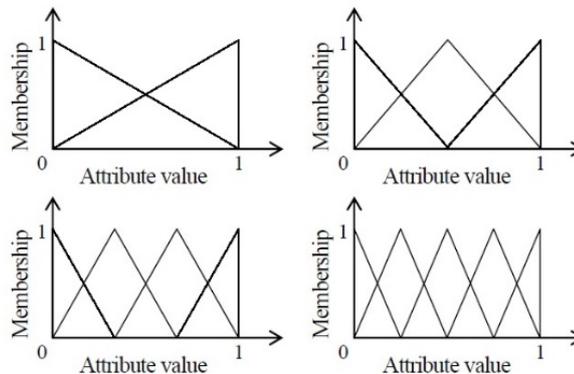
2.3 Fuzzy Rule-Based Classification Systems

Pada bagian ini, secara singkat fuzzy berbasis aturan dijelaskan secara singkat dengan mengasumsikan bahwa kita memiliki m pola pelatihan $\mathbf{x}_p = (x_{p1}, x_{p2}, \dots, x_{pn})$, $p = 1, 2, \dots, m$ dari kelas M di unit n -dimensi hypercube $[0, 1]^n$. Artinya, masalah klasifikasi pola adalah kelas- M masalah dengan m pola pelatihan dalam ruang pola n -dimensi $[0, 1]^n$.

Dimana aturannya menggunakan aturan fuzzy IF-THEN dari jenis berikut:

Aturan R_q : IF x_1 adalah A_{q1} dan ... dan x_n adalah A_{qn} THEN $Class_q$ with $Class_{Fq}$, (3)

di mana R_q adalah label dari aturan fuzzy ke- q , $x = (x_1, \dots, x_n)$ adalah vektor pola berdimensi- n , A_{qi} adalah himpunan fuzzy antecedent untuk atribut ke- i , $Class_q$ adalah kelas konsekuen, dan $Class_{Fq}$ adalah tingkat kepastian (yaitu, bobot aturan). Persamaan (1) menyatakan aturan fuzzy sebagai " $A_q \Rightarrow Class_q$ ". Sebagai himpunan fuzzy anteseden A_{qi} , kita menggunakan salah satu dari himpunan fuzzy pada Gambar 1. Himpunan fuzzy anteseden A_{qi} dapat juga dalam status "don't care". Jadi jumlah total kombinasi dari n himpunan fuzzy antecedent pada (3) adalah 15^n .



Gambar 1. Antecedent himpunan Fuzzy

Pemilihan aturan fuzzy dalam makalah ini adalah untuk menemukan himpunan aturan tak terdominasi dari aturan fuzzy 15^n bentuk dalam (1) sehubungan dengan akurasi dan kompleksitas.

Karena setiap subset dari aturan fuzzy 15^n dapat diwakili oleh string biner dengan panjang 15^n , ukuran ruang pencarian adalah 2^{15^n} . Kecuali untuk kasus masalah dimensi rendah, sangat sulit untuk menangani ruang pencarian yang begitu besar. Jadi dua tahap pendekatan pemilihan aturan fuzzy dapat digunakan. Tahap pertama adalah ekstraksi aturan heuristik di mana sejumlah aturan kandidat yang menjanjikan diekstraksi dari data numerik menggunakan ukuran evaluasi aturan heuristik dengan cara yang sama seperti penambahan data. Tahap kedua adalah pemilihan aturan evolusioner di mana algoritma optimasi evolusioner digunakan untuk menemukan himpunan bagian yang tidak didominasi dari aturan kandidat yang diekstraksi sehubungan dengan akurasi dan kompleksitas yang menjadi tujuan.

2.4. Pemilihan Fitur (Feature selection) dan Strategi Pencarian

Seleksi fitur didefinisikan sebagai proses menghilangkan fitur dari kumpulan data yang tidak relevan dengan tugas yang akan dilakukan. Ini memfasilitasi pemahaman data, mengurangi persyaratan pengukuran dan penyimpanan, mengurangi waktu komputasi, dan mengurangi ukuran dataset, sehingga pembelajaran model menjadi proses yang lebih mudah.

Metode seleksi pada dasarnya adalah strategi pencarian dimana kinerja subset kandidat diukur dengan evaluator yang diberikan. Ruang pencarian untuk subset kandidat memiliki kardinalitas $O(2n)$, di mana n adalah jumlah fitur. Kriteria penghentian menetapkan kapan proses pemilihan fitur harus selesai. Ini dapat didefinisikan sebagai prosedur kontrol yang memastikan bahwa tidak ada penambahan atau penghapusan fitur yang menghasilkan subset yang lebih baik, atau dapat sesederhana penghitung iterasi.

Metode pemilihan fitur yang digunakan adalah wrapper multivariat klasifikasi fuzzy. Adapun strategi pencarian, diterapkan metode *Multi Objective Evolutionary Search* dengan algoritma evolusioner multi-tujuan ENORA. Kinerja ENORA umumnya lebih baik daripada NSGA-II dalam hal hypervolume, dan lebih baik daripada strategi pencarian tujuan tunggal lainnya [8]. Pencarian yang dipandu ACC telah memberikan hasil yang lebih baik dalam percobaan pendahuluan daripada pencarian yang menggunakan area di bawah kurva ROC (AUC), yang memberikan nilai ACC yang sangat buruk. Oleh karena itu, untuk mengevaluasi subset atribut kandidat, algoritme evolusi multi-tujuan dijalankan untuk mengekstrak set aturan fuzzy pada basis data yang direduksi, yang dievaluasi dengan validasi silang pada metrik akurasi.

2.5. Dataset

Data yang digunakan adalah 3 dataset yang telah disediakan oleh penyedia repositori data terkemuka untuk penelitian data mining dan machine learning, yaitu Kaggle dan UCI Machine Learning repository. Dimana data tidak lengkap (*missing value*) diganti dengan nilai mean atau median. Dan pemberlakuan fitur selection pada data dengan tanda asterisk (*).

Tabel 1. Benchmark Dataset

ID	Dataset (source)	Fitur	Record	Class-label
#1	WaterPotability (Kaggle)	10	3276	2
#2	Iris (UCI ML)	5	150	3
#3	Credit (UCI ML)	4*	1000	4

3. HASIL PENELITIAN DAN PEMBAHASAN

Serangkaian eksperimen pada bagian ini, dilakukan untuk menunjukkan efektivitas metodologi yang diusulkan. Tiga blok percobaan telah dilakukan: yang pertama bertujuan untuk menemukan jumlah generasi yang optimal dari kedua strategi pencarian dan evaluator untuk pertukaran yang memadai antara kinerja dan jumlah aturan; di blok kedua, membandingkan pengklasifikasi fuzzy yang diperoleh dengan metode yang diusulkan dengan pengklasifikasi fuzzy yang diperoleh dengan metode yang sudah dikenal lainnya yaitu (C.45) [9].

3.1. Setting Parameter Blok Percobaan Pertama

Blok percobaan pertama untuk menguji kemampuan pencarian ENORA dan NSGA-II algoritma, dilakukan pemeriksaan set aturan terbaik yang sehubungan dengan akurasi diantara aturan yang diperoleh dari setiap reportfrequency. Akurasi rata-rata dari aturan terbaik yang ditetapkan lebih dari 20 reportfrequency ditunjukkan pada Tabel 2. Dan Tabel 3.

Tabel 2. Percobaan pertama (ENORA) with 10-FoldCV

Dataset	Jumlah Generasi	Evaluation Measure	Jumlah Aturan	Dataset	Jumlah Generasi	Evaluation Measure	Jumlah Aturan
<i>ACC</i>				<i>AUC</i>			
#1	20	62.88	11	#1	20	52.97	12
	50	63.16	11		50	59.40	4
	100	64.01	7		100	60.93	5
#2	20	92.67	3	#2	20	95.33	5
	50	93.33	6		50	92.67	8
	100	93.33	5		100	94	6
#3	20	70.4	2	#3	20	64.7	3
	50	70	12		50	66.1	8
	100	71	11		100	67.7	8

Tabel 3. Percobaan pertama (NSGA-II) with 10-FoldCV

Dataset	Jumlah Generasi	Evaluation Measure	Jumlah Aturan	Dataset	Jumlah Generasi	Evaluation Measure	Jumlah Aturan
<i>ACC</i>				<i>AUC</i>			
#1	20	63.67	2	#1	20	59.98	12
	50	63.86	10		50	60.68	7
	100	60.92	11		100	60.80	9
#2	20	94	4	#2	20	95.33	4
	50	95.33	4		50	94.67	4
	100	94.67	4		100	94.67	4
#3	20	71.8	3	#3	20	70.4	2
	50	72.8	4		50	70	12
	100	72.7	4		100	71	11

3.2. Setting Parameter Blok Percobaan Kedua

Tabel 3. Percobaan Kedua MOFC (ENORA, gen=20) versus C.45

Dataset	Classifier		Jumlah Aturan MOFC	Jumlah Aturan C.45
	MOFC	C.45		
#1	63.67	62.33	2	7
#2	94.67	94	4	3
#3	71.8	71.5	4	17

3.3. Himpunan Aturan Fuzzy

Selanjutnya kita fokuskan untuk mengamati tabel yang berisi himpunan aturan fuzzy yang diperoleh dari dataset *water_potability*. Aturan fuzzy pengklasifikasi mengganti nilai label class ke dalam 5 area yaitu: *low*, *moderately_low*, *medium*, *moderately_high*, *high*. Dari hasil pengamatan pada aturan yang terbentuk kita dapat melihat bahwa aturan ini sangat rasional dalam menempatkan setiap linguistic label untuk memenuhi ketentuan konsekuen rule masing-masing, yang dapat dilihat dari besaran dan rentang nilai center sebagaimana tersaji pada Tabel 4.

Tabel 4. Fuzzy Rule-Based Classification Models (#1, Gen=20, ACC, NSGA-II)

Himpunan Aturan Fuzzy				
RULE 1:				
IF		ph	IS	Low
	AND	solids	IS	Low
	AND	sulfate	IS	High
	AND	conductivity	IS	Low
	AND	organic_carbon	IS	High
	AND	trihalomethanes	IS	Medium
THEN		potability	IS	0
RULE 2:				
IF		ph	IS	Moderately High
	AND	solids	IS	Medium
	AND	sulfate	IS	Low
	AND	conductivity	IS	Moderately Low
	AND	organic_carbon	IS	Medium
	AND	trihalomethanes	IS	Low
THEN		potability	IS	1
Attribute	Center		Linguistic Label	
ph	1.0314 - 5.3072		Low, Moderately High	
solids	1398.7452 - 5929.0954		Low, Medium	
sulfate	458.4438 - 92.6294		High, Low	
conductivity	197.7393 - 104.6102		Low, Moderately Low	
organic_carbon	25.5511 - 5.1496		High, Medium	
trihalomethanes	62.9626 - 10.365		Medium, Low	

Keuntungan utama dari formulasi multiobjective dibandingkan dengan single-objective adalah bahwa beberapa set aturan yang tidak didominasi diperoleh dari single run-nya, yang secara visual menunjukkan keseimbangan hasil akurasi versus kompleksitas Hasil eksperimen menunjukkan bahwa perolehan dengan pelatihan pengklasifikasi non-fuzzy mungkin tidak selalu memberikan hasil yang lebih baik.

4. KESIMPULAN

Serangkaian percobaan dalam penelitian ini menerapkan teknik penambangan aturan pengklasifikasi dimana kedua strategi pencarian dan *classifier* didasarkan pada komputasi *evolutioner multiobjective*. Dalam makalah ini juga dirancang serangkaian eksperimen untuk menetapkan setelan yang dapat diterima sehubungan dengan jumlah evaluasi yang diperlukan oleh strategi penelusuran serta pengklasifikasi, dan kami menguji strategi kami pada *benchmark dataset*. Metode ini dapat membantu memecahkan masalah pemilihan atribut terbaik untuk tugas pembelajaran klasifikasi yang sangat spesifik berdasarkan aturan fuzzy. Pada kenyataannya, kinerja pengklasifikasi yang diberikan sangat masuk akal untuk atribut yang dipilih dan menggunakan pemilihan filter, yang didasarkan pada nilai statistik umum, dimana peningkatan jumlah tujuan memiliki kemungkinan untuk meningkatkan kemampuan generalisasi sistem klasifikasi berbasis aturan fuzzy yang dapat membuat pencarian set aturan yang tidak didominasi menjadi lebih sulit.

DAFTAR PUSTAKA

- [1] A. Fernández, V. López, M. J. del Jesus, and F. Herrera, "Revisiting Evolutionary Fuzzy Systems: Taxonomy, applications, new trends and challenges," *Knowledge-Based Systems*, vol. 80, 2015, doi: 10.1016/j.knosys.2015.01.013.
- [2] S. Birtane and H. Korkmaz, "Rule-based fuzzy classifier for spinal deformities," in *Bio-Medical Materials and Engineering*, 2014, vol. 24, no. 6. doi: 10.3233/BME-141154.
- [3] F. R. Hariri, "Klasifikasi Jenis Golongan Darah Menggunakan Fuzzy C-Means Clustering (FCM) dan Learning Vector Quantization (LVQ)," *MATICS*, vol. 10, no. 1, 2018, doi: 10.18860/mat.v10i1.5356.
- [4] O. Cordón, "A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems," *International Journal of Approximate Reasoning*, vol. 52, no. 6. 2011. doi: 10.1016/j.ijar.2011.03.004.
- [5] H. Ishibuchi, Y. Kaisho, and Y. Nojima, "Design of linguistically interpretable fuzzy rule-based classifiers: A short review and open questions," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2–3, 2011.
- [6] S. Elhag, A. Fernández, A. Altalhi, S. Alshomrani, and F. Herrera, "A multi-objective evolutionary fuzzy system to obtain a broad and accurate set of solutions in intrusion detection systems," *Soft Computing*, vol. 23, no. 4, 2019, doi: 10.1007/s00500-017-2856-4.
- [7] M. Galende-Hernández, G. I. Sainz-Palmero, and M. J. Fuente-Aparicio, "Complexity reduction and interpretability improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection," *Soft Computing*, vol. 16, no. 3, 2012, doi: 10.1007/s00500-011-0748-6.
- [8] J. Novaković, P. Strbac, and D. Bulatović, "Toward optimal feature selection using ranking methods and classification algorithms," *Yugoslav Journal of Operations Research*, vol. 21, no. 1, 2011, doi: 10.2298/YJOR1101119N.
- [9] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.