

Implementasi FIS Sugeno pada Algoritma C4.5 Berbasis Particle Swarm Optimization (PSO) untuk Prediksi Prestasi Siswa

Muhamad Arief Yulianto

Program Studi Teknik Informatika, Fakultas Teknik Sekolah Tinggi Manajemen Informatika dan Komputer ERESHA Jl. Raya Puspitex No.10, Serpong - Tangerang Selatan
e-mail: arief.kampus@gmail.com

Abstrak

Standar kompetensi kelulusan siswa yang meliputi kemampuan berpikir logis, kritis, kreatif dan inovatif, serta pencapaian prestasi nilai rata – rata UN teori dan praktik merupakan salah satu elemen penilaian akreditasi sekolah menengah. Namun permasalahannya rata – rata hasil ujian nasional di 100 SMK Negeri dan Swasta di Jakarta tahun ajaran 2015 – 2017 mengalami penurunan dari 73,51 menjadi 69,05 berdasarkan data dari situs resmi kementerian pendidikan dan kebudayaan. Oleh sebab itu dalam penelitian ini bertujuan untuk membangun model dengan mengimplementasikan fuzzy inference system metode sugeno pada algoritma decision tree C4.5 berbasis particle swarm optimization (ps) serta membuktikan apakah performa algoritma dilihat dari tingkat akurasi, spesitifitas dan sensitifitas mengalami peningkatan performa algoritma yang dihasilkan oleh algoritma decision tree C4.5 berbasis PSO dalam memprediksi prestasi siswa. Implementasi Algoritma decision tree C4.5 berbasis particle swarm optimization (ps) menggunakan 299 data dengan 32 atribut independen dan 1 atribut dependen menghasilkan pohon keputusan dengan jumlah atribut sebanyak 18 atribut dependen dan 1 atribut dependen yang akan digunakan sebagai rule dalam implementasi fis sugeno. Dalam implementasinya diperoleh hasil bahwa implementasi fis sugeno pada algoritma decision tree C4.5 berbasis PSO (Particle Swarm Optimization) mampu meningkatkan performa tingkat akurasi algoritma C4.5 berbasis PSO dari 94,66% menjadi 96% dan spesitifitas dari 94,24% menjadi 97,80%.

Kata kunci : *fuzzy inference system*, Sugeno, C4.5, PSO, lulusan.

Abstract

The standard of student's graduation competency which includes logical thinking ability, critical, creative and innovative, and achievement of the average value of UN theory and practice is one element of the assessment of secondary school accreditation. However the problem is the average national exam results in 100 SMK Negeri and Private in Jakarta academic year 2015 - 2017 has decreased from 73.51 to 69.05 based on data from the official website of the ministry of education and culture. Therefore this study is aimed to build a model by implementing fuzzy sugeno on the PSO-based C4.5 algorithm as well as to prove whether the accuracy level in fuzzy implementation can be increased from the accuracy level generated by the PSO-based C4.5 algorithm in predicting the achievement of middle students. Implementation of PSO-based Algorithm C4.5 uses 299 data with 32 independent attributes and 1 dependent attribute yields a decision tree with the number of attributes of 18 dependent attributes and 1 dependent attribute that will be used as rule in fuzzy sugeno implementation. The implementation of fuzzy sugeno on the PSO (Particle Swarm Optimization) C4.5 algorithm improved the accuracy of PSObased C4.5 algorithm from 94.66% to 96% with an increase of 1.34%.

Keywords: fuzzy sugeno, C4.5, PSO, graduates.

1. Pendahuluan

Standar kompetensi kelulusan siswa yang meliputi kemampuan berpikir logis, kritis, kreatif dan inovatif, serta pencapaian prestasi nilai rata – rata UN teori dan praktik merupakan salah satu

elemen penilaian akreditasi perguruan tinggi yang dijelaskan dalam Salinan Peraturan Menteri Pendidikan Nasional Nomor 13 Tahun 2009 Tentang Kriteria Dann Perangkat Akreditasi Sekolah Menengah Kejuruan/Madrasah Aliyah

Kejuruan (SMK/MAK). Algoritma C4.5 telah digunakan sebagai teknik klasifikasi dengan membentuk pohon keputusan yang memiliki tingkat keakuratan, kecepatan, dan kehandalan yang lebih baik dibandingkan dengan teknik klasifikasi lainnya seperti *support vector machine* (SVM), naive bayes dan neural network. David Hartanto Kamagi dan Seng Hansun Mengimplementasikan algoritma C4.5 dalam memprediksi kelulusan mahasiswa diperoleh tingkat akurasi sebesar 87,5%. Lain halnya yang dilakukan oleh Rizqi Agung Permana melakukan penelitian dalam proses seleksi atribut metode *Support Vector Machine* untuk menentukan kelulusan mahasiswa *E-Learning* diperoleh tingkat akurasi sebesar 81,25% dan nilai AUC sebesar 0,825. Metode pada Machine learning dan data mining pada saat ini telah dikembangkan untuk digunakan dalam membangun basis pengetahuan pada sistem pakar. Sedangkan Paulo dan Alice Silva dalam penelitiannya yang berjudul *using data mining to predict secondary school student performance* diperoleh hasil dengan tingkat akurasi sebesar *naïve bayes* 91,9%, *Neural Network* 88.3 %, *SVM* 86,3%, *Decision Tree* 90.7 % dan *Random Forest* 91,2%. *Particle Swarm Optimization* (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi optimal secara global dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel. Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut. *Particle swarm optimization* dapat digunakan sebagai teknik optimasi untuk mengoptimalkan subset fitur dan parameter secara bersamaan. Farid Nurhidayat melakukan penelitian dalam menentukan besar akurasi metode klasifikasi menggunakan algoritma C4.5 berbasis *Particle Swarm Optimization* pada prediksi penyakit diabetes menyimpulkan bahwa tingkat akurasi algoritma C4.5 berbasis *PSO (Particle Swarm Optimization)* dengan algoritma C4.5 mengalami kenaikan dari tingkat akurasi 73,56% dengan AUC 0,773 menjadi 76,84% dengan AUC 0,785. Berbeda dengan penelitian yang dilakukan oleh Khoirul Muarif (2013) dalam membandingkan algoritma C4.5 dengan algoritma C4.5 berbasis *PSO (Particle Swarm Optimization)* untuk memprediksi kelulusan mahasiswa membuktikan bahwa algoritma C4.5 berbasis *PSO (Particle Swarm Optimization)* memiliki

tingkat akurasi lebih tinggi yaitu sebesar 86,09 % daripada algoritma C4.5 yang memiliki tingkat akurasi sebesar 84,13 %. Metode fuzzy sudah sering digunakan dalam kasus klasifikasi dan memberikan hasil yang bagus dalam klasifikasi maupun prediksi. Fuzzy memiliki keunggulan dalam hal perhitungannya yang tidak kaku (samar), sehingga mampu memperhitungkan kemungkinan tidak pasti. Adapun kelebihan dari FIS Sugeno yaitu efisien dalam komputasi, cocok untuk pemodelan – pemodelan sistem linier, cocok untuk digabung dengan teknik optimasi dan adaptif, menjamin kontinuitas keluaran dan memungkinkan dilakukan analisis matematis sedangkan FIS Mamdani bersifat intuitif, diterima secara luas dan sangat cocok diberi human input. Penelitian yang dilakukan oleh Meenakshi dan Pankaj Nagar dalam penerapan metode mamdani yang berjudul *Application of Fuzzy Logic for Evaluation of Academic Performance of Students of Computer Application Course* diperoleh hasil dengan tingkat akurasi sebesar 95%. Berbeda dengan yang dilakukan oleh Heri Bambang Santoso dalam membangkitkan aturan pada *Fuzzy Inference System* menggunakan *fuzzy decision tree* untuk memprediksi keberhasilan studi mahasiswa memperoleh hasil tingkat akurasi *fuzzy ID3* sebesar 95,85% sedangkan ID3 sebesar 93,41% (Heri Bambang Santoso, 2015). Abidatul Izzah dan Ratna Widyastuti melakukan prediksi kelulusan mata kuliah menggunakan *Hybrid Fuzzy Inference System* memperoleh hasil dengan tingkat akurasi sebesar 94,33%, sensitifitas 96,55% dan spesifisitas 84,21% (Abidatul Izzah, 2016). Berdasarkan uraian di atas, dengan menggunakan data penelitian yang telah dilakukan oleh Paulo Cortez dan Alice Silva, peneliti akan mencoba untuk mengimplementasikan fuzzy inference sistem metode sugeno pada algoritma C4.5 berbasis *particle swarm optimization*(PSO) dalam memprediksi prestasi siswa dengan meningkatkan performa dilihat dari tingkat akurasi, spesifisitas dan sensitifitas pada algoritma *decision tree c4.5* berbasis *particle swarm optimization* (PSO).

2. Metode Penelitian

A. Data Preparation

Pada penelitian ini, data yang digunakan berasal dari penelitian Paulo Cortez dan Alice Silva yang berjudul “*Using Data Mining to Predict Secondary School Student Performance*” yang telah diupload di repositori UCI. Adapun data penelitian yang

digunakan dalam penelitian ini yaitu data set *student-mat.csv* yang berupa data laporan prestasi siswa berjumlah 395 *record* dan terdiri dari 33 atribut dengan 16 atribut bertipe numerik, 13 atribut bertipe biner dan 4 atribut bertipe nominal. Adapun atribut ataupun variabel independen yang digunakan yaitu (*sex, age, school, address, Pstatus, Medu, Mjob, Fedu, guardian, famsize, famrel, reason, traveltime, studytime, failures, schoolsup, famsup, activities, paidclass, internet, nursery, higher, romantic, freetime, goout, Walc, Dalc, health, absences, G1, G2*). Data akan dikelompokkan ke dalam dua kelas yang merupakan variabel dependen (*G3*), jika nilai *G3* lebih dari 10 maka masuk ke kelas 1 (*PASS*) dan jika nilai *G3* kurang dari sama dengan 10 maka masuk ke kelas 2 (*FAIL*). Adapun deskripsi atribut ataupun variabel yang akan peneliti gunakan untuk algoritma *c4.5* berbasis *PSO(Particle Swarm Optimization)* adalah sebagai berikut :

a. Variabel independen

1. *Sex* (Jenis Kelamin)

Sebagian besar penelitian yang menguji pengaruh gender terhadap prestasi belajar menunjukkan bahwa perempuan cenderung memiliki prestasi akademis yang lebih bagus daripada laki-laki. Penjelasan teoritis mengenai hal ini antara lain karena setelah zaman emansipasi, pendidikan merupakan kunci utama kemajuan, pemberdayaan dan kebebasan bagi kaum perempuan. Selain itu, perempuan dikenal cenderung lebih tekun dalam belajar dan rajin terlibat dalam kegiatan kampus yang menunjang proses belajar, sedangkan laki-laki lebih menyukai kegiatan kampus yang bersifat refreshing dan olah raga.

2. *Age* (Usia)

Richardson (1994) mengemukakan bahwa mahasiswa yang usianya relatif lanjut diduga mengalami *age-related intellectual deficits* sehingga mengalami penurunan dalam hal *basic skills* yang diperlukan untuk belajar efektif pada tingkat pendidikan tinggi.

3. *School* (Asal Sekolah)

Asal sekolah mempengaruhi terhadap prestasi kelulusan siswa, dikarenakan tidak semua sekolah memiliki fasilitas

dan kualitas yang sama antara 1 sekolah dengan sekolah lainnya.

4. *Address* (Alamat Tempat Tinggal)

Lokasi tempat tinggal siswa akan mempengaruhi terhadap prestasi kelulusan siswa, siswa yang tinggal di pedesaan akan berbeda dengan siswa di perkotaan hal itu terjadi dikarenakan faktor lingkungan dan fasilitas yang disediakan.

5. *Pstatus* (Status Pernikahan Orang Tua)

Status Pernikahan Orang Tua akan mempengaruhi prestasi kelulusan siswa, karena dengan adanya orang tua yang lengkap yang tinggal bersama dengan anak maka anak tersebut akan memperoleh dukungan baik moral ataupun moril yang mempengaruhi emosional maupun karakter dari anak.

6. *Medu* (Pendidikan Ibu)

Tingkat pendidikan Ibu berpengaruh terhadap bagaimana cara mendidik anak di rumah, hal tersebut akan mempengaruhi terhadap prestasi kelulusan siswa.

7. *Mjob* (Pekerjaan Ibu)

Jenis pekerjaan ibu dapat mempengaruhi sikap dan cara mendidik anak di rumah, jika jenis pekerjaan memiliki tingkat stress tinggi maka hal tersebut akan terbawa sampai ke rumah sehingga akan berdampak terhadap sikap ibu kepada anak.

8. *Fedu* (Tingkat Pendidikan Ayah)

Tingkat pendidikan ayah juga berpengaruh terhadap bagaimana cara mendidik anak di rumah selain itu juga sebagai panutan dan motivasi untuk anak sehingga hal tersebut dapat mempengaruhi terhadap prestasi kelulusan siswa.

9. *Fjob* (Pekerjaan Ayah)

Jenis pekerjaan ayah dapat mempengaruhi terhadap prestasi kelulusan siswa, hal tersebut terjadi jika sosok ayah yang menjadi figur dalam sebuah keluarga tidak pernah memiliki waktu luang untuk memberikan waktu bersama kepada keluarga terutama kepada anak sehingga hal tersebut akan mempengaruhi psikologi anak.

10. *Guardian* (Pelindung) Dengan adanya sosok pelindung dalam artian sebagai pengayom buat anak akan berpengaruh terhadap psikologi anak sehingga mempengaruhi prestasi kelulusan siswa.
11. *Famsize* (Jumlah Anggota Keluarga) Jumlah anggota keluarga memiliki pengaruh terhadap prestasi kelulusan siswa, semakin besar anggota keluarganya akan mempengaruhi baik dari segi ekonomi, kasih sayang dan dukungan kepada anak.
12. *Famrel* (Kualitas Hubungan Keluarga) Baik atau buruknya suatu hubungan dalam keluarga sangat berpengaruh terhadap belajar anak. Hubungan yang baik adalah hubungan yang penuh pengertian dan kasih sayang disertai dengan bimbingan dari orang tua dan saudara, karena dapat membuat anak nyaman pada saat belajar (Utami, 2014).
13. *Reason* (Alasan Memilih Sekolah) Dengan adanya alasan yang kuat dan positif terhadap pemilihan sekolah akan mempengaruhi semangat dalam mengikuti proses kegiatan belajar sehingga mempengaruhi prestasi siswa.
14. *Traveltime* (Lama Perjalanan Ke Sekolah) Waktu tempuh perjalanan akan mempengaruhi terhadap kondisi fisik siswa yang akan mempengaruhi tingkat konsentrasi siswa.
15. *Studytime* (Durasi Belajar Selama Seminggu) Durasi belajar selama seminggu akan mempengaruhi kemampuan siswa dalam proses berfikir dan memahami pelajaran sehingga akan mempengaruhi prestasi siswa.
16. *Failures* (Total Tidak Lulus / Naik Kelas Periode Sebelumnya) Histori siswa dalam hal kenaikan ataupun kelulusan akan mempengaruhi terhadap mental ataupun psikologi siswa sehingga mempengaruhi prestasi siswa.
17. *Schoolsup* (Kursus / Les Diluar Sekolah) Dengan adanya kegiatan kursus ataupun belajar di luar sekolah akan membantu siswa dalam melakukan proses pemahaman pembelajaran sehingga akan mempengaruhi prestasi siswa.
18. *Famsup* (Dukungan Keluarga Terhadap Pendidikan) Dengan adanya dukungan keluarga terhadap pendidikan, maka siswa akan menambah kepercayaan diri sehingga dapat mempengaruhi psikologi dan mental siswa.
19. *Activities* (Kegiatan Ektrakurikuler) Dengan adanya kegiatan ekstrakurikuler pada siswa akan dapat mempengaruhi baik kemampuan berpikir maupun kondisi fisik siswa.
20. *Paidclass* (Kursus / Les Bidang Pelajaran Tertentu) Dengan adanya kegiatan kursus ataupun les pelajaran tertentu maka akan membantu siswa dalam melakukan proses pemahaman pembelajaran sehingga akan mempengaruhi prestasi siswa.
21. *Internet* (Akses Internet di rumah) Adanya akses internet akan menambah kemudahan bagi siswa untuk mendapatkan informasi yang sebanyak banyaknya sehingga mampu menambah pengetahuan siswa.
22. *Nursery* (Ketersediaan Taman Di Sekolah) Adanya taman di sekolah dapat menambah kebugaran baik pikiran ataupun jasmani bagi siswa di saat jenuh seharian melakukan aktivitas di dalam kelas, sehingga dapat mengembalikan semangat belajar siswa.
23. *Higher* (Motivasi Melanjutkan Pendidikan Ke Jenjang Lebih Tinggi) Dengan adanya motivasi siswa untuk melanjutkan pendidikan ke jenjang yang lebih tinggi dapat menambah semangat siswa dalam belajar.
24. *Romantic* (Adanya Hubungan Spesial Antara lawan Jenis (Pacaran)) Hubungan spesial terhadap lawan jenis akan mempengaruhi terhadap kondisi psikologi maupun semangat dalam belajar.
25. *Freetime* (Waktu Luang Setelah Sekolah) Ketersediaan waktu luang setelah sekolah dapat mempengaruhi kondisi pikiran dan fisik siswa, karena dengan

adanya waktu luang setelah sekolah siswa dapat memanfaatkan untuk kegiatan yang positif.

26. *Goout* (Kumpul Bareng Teman (Nongkrong/Bergadang))

Tingkat frekuensi siswa dalam berkumpul ataupun nongkrong dengan teman – temannya akan mempengaruhi kondisi pikiran ataupun fisik siswa sehingga akan mengurangi waktu untuk belajar siswa.

27. *Walc* (Kegiatan Mengonsumsi Alkohol di waktu libur)

Kegiatan mengonsumsi alkohol di waktu *weekend* akan mempunyai pengaruh terhadap kondisi pikiran dan kesehatan siswa.

28. *Dalc* (Kegiatan Mengonsumsi Alkohol di hari kerja)

Kegiatan mengonsumsi alkohol di hari kerja akan mempunyai pengaruh terhadap kondisi pikiran dan kesehatan siswa pada saat mengikuti kegiatan proses belajar di sekolah.

29. *Health* (Kondisi Kesehatan)

Kondisi badan yang sehat akan mempengaruhi kemampuan untuk berfikir baik di dalam kelas maupun di dalam kelas.

30. *Absences* (Jumlah ketidakhadiran siswa di dalam kelas)

Jumlah ketidakhadiran siswa di dalam kelas akan berpengaruh terhadap tingkat pemahaman siswa terhadap materi yang dipelajarinya.

31. *G1*

G1 digunakan untuk mengukur prestasi akademis pada tingkat pertama.

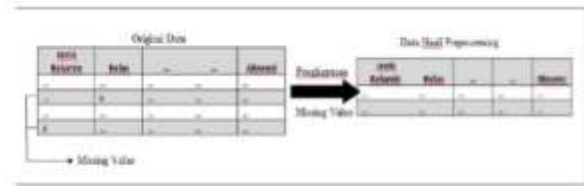
32. *G2*

G2 digunakan untuk mengukur prestasi akademis pada tingkat kedua.

b. Variabel Dependen

G3 merupakan target dari penelitian ini yang akan dikelompokkan dalam dua kelas yaitu kelas *PASS* (Lulus) dan kelas *FAIL* (TidakLulus).

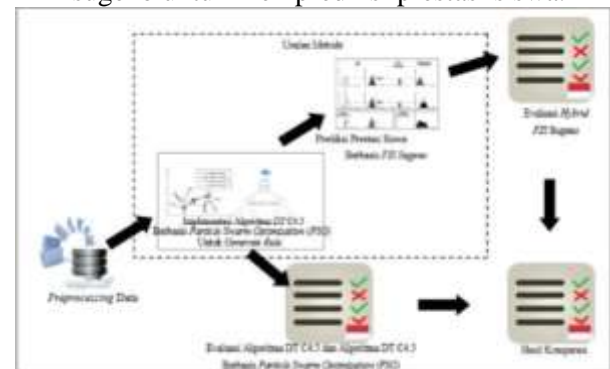
Preparation data dilakukan dengan cara penghapusan missing value. Missing value disebabkan oleh beberapa mahasiswa yang tidak mengikuti atau mengerjakan tes sehingga menyebabkan nilai kosong.



Gambar 1. *Preprocessing Data*

B. Metode Penelitian

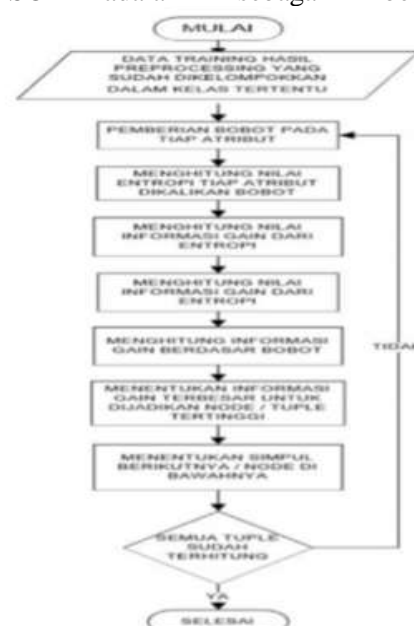
Model yang akan dibangun pada penelitian ini akan digunakan algoritma C4.5 berbasis PSO (particle swarm optimization) yang digunakan untuk membangkitkan aturan – aturan yang akan digunakan oleh fuzzy sugeno untuk memprediksi prestasi siswa.



Gambar 2. *Desain Metode Penelitian*

Model yang akan dibangun pada penelitian ini akan digunakan algoritma C4.5 berbasis PSO (particle swarm optimization) yang digunakan untuk membangkitkan aturan – aturan yang akan digunakan oleh fuzzy sugeno untuk memprediksi prestasi siswa.

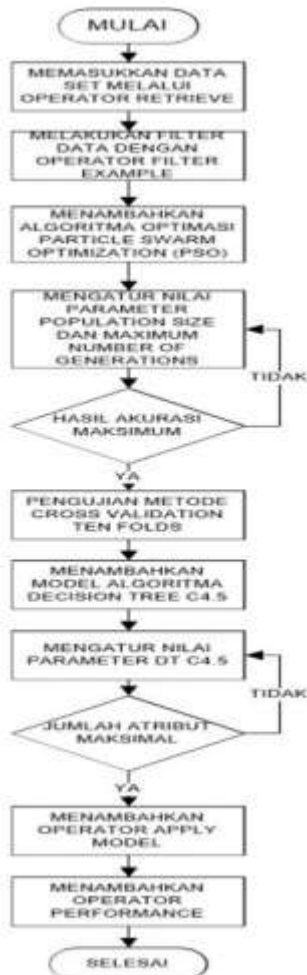
Adapun flowchart algoritma C4.5 berbasis PSO adalah sebagai berikut :



Gambar 3. *Algoritma C4.5 Berbasis PSO*



Gambar 4. Algoritma FIS Sugeno



Gambar 5. Alur Pengujian Algoritma DT C4.5 + PSO pada Rapid Miner

Berikut ini akan diberikan contoh beberapa perhitungan secara manual berdasarkan model

C4.5 berbasis PSO dengan FIS Sugeno sebagai berikut :

1. Model C4.5 Berbasis PSO (Particle Swarm Optimization)

- a. Simulasi pembobotan atribut dengan PSO dihitung berdasarkan persamaan kecepatan dan posisi partikel seperti di bawah ini :

$$V_i(t) = V_i(t-1) + c_1 r_1 [X_{pbest_i} - X_i(t)] + c_2 r_2 [X_{gbest} - X_i(t)]$$

$$X_i(t) = X_i(t-1) + V_i(t)$$

Berdasarkan persamaan di atas, dengan melakukan proses iterasi sebanyak 10 kali pada tiap atribut diperoleh hasil sebagai berikut :

Tabel 1. Hasil Bobot Iterasi 0 Sampai 10

Iterasi	school	sex	age	address	famsize	health	absences	G1	G2
P0	0.03	0.59	0.36	0.75	0.94	0.84	0.87	0.42	0.84
P1	0.03	0.23	0.15	0.29	0.36	0.33	0.34	0.17	0.33
P2	0.03	-0.25	-0.14	-0.34	-0.43	-0.38	-0.39	-0.17	-0.38
P3	-0.27	-0.86	-0.62	-1.03	-1.23	-1.13	-1.16	-0.68	-1.13
P4	-1.19	-1.72	-1.51	-1.87	-2.05	-1.96	-1.98	-1.56	-1.96
P5	-2.68	-2.80	-2.75	-2.84	-2.88	-2.86	-2.87	-2.76	-2.86
P6	-4.29	-3.94	-4.08	-3.84	-3.72	-3.78	-3.76	-4.04	-3.78
P7	-5.91	-5.30	-5.54	-5.12	-4.92	-5.03	-5.00	-5.48	-5.03
P8	-7.52	-7.05	-7.24	-6.90	-6.74	-6.83	-6.80	-7.19	-6.83
P9	-9.14	-9.09	-9.11	-9.08	-9.06	-9.07	-9.07	-9.10	-9.07
P10	-10.75	-11.16	-11.00	-11.29	-11.43	-11.35	-11.37	-11.04	-11.35

- b. Mengalikan nilai bobot dengan probabilitas sampel diperoleh potongan hasil sebagai berikut :

Tabel 2. Hasil Perkalian Bobot dengan Probabilitas Sampel

CABANG	BOBOT	DETAIL ATRIBUT	BOBOT TIAP ATRIBUT	PASS	BOBOT TIAP ATRIBUT FAIL	FAIL	P(PASS) x Bobot	P(FAIL) x Bobot	
1	SCHOOL	GP	0.01	190	0.01	159	0.01	0.01	
		MS	0.00	19	0.00	27	0.00	0.00	
2	SEX	F	0.29	103	0.30	105	0.14	0.15	
		M	0.00	106	0.00	81	0.00	0.00	
3	AGE	15	0.23	51	0.14	31	0.14	0.05	
		16	0.00	61	0.00	43	0.00	0.00	
		17	0.00	50	0.00	48	0.00	0.00	
		18	0.00	35	0.00	47	0.00	0.00	
		19	0.00	10	0.00	14	0.00	0.00	
		20	0.00	2	0.00	1	0.00	0.00	
		21	0.00	0	0.00	1	0.00	0.00	
		22	0.00	0	0.00	1	0.00	0.00	
4	ADDRESS	R	0.33	38	0.43	50	0.14	0.24	
		U	0.00	171	0.00	136	0.00	0.00	
5	FAMSIZE	GT3	0.49	145	0.46	136	0.25	0.22	
		LE3	0.00	64	0.00	50	0.00	0.00	
6	PSTATUS	T	0.34	184	0.31	170	0.17	0.15	
		A	0.00	25	0.00	16	0.00	0.00	
...	
32	G2	0.84	0	0.00	0	0.84	13	0.00	0.84
		1	0.00	0	0.00	0	0.00	0.00	
		2	0.00	0	0.00	0	0.00	0.00	
		3	0.00	0	0.00	0	0.00	0.00	
		4	0.00	0	0.00	1	0.00	0.00	
		5	0.00	0	0.00	15	0.00	0.00	
		6	0.00	0	0.00	14	0.00	0.00	
		7	0.00	0	0.00	21	0.00	0.00	
		8	0.00	2	0.00	30	0.00	0.00	

- c. Menghitung nilai entropi tiap sampel kasus dengan Persamaan entropi sebagai berikut :

$$Entropy(S) = \sum_{i=0}^{|n|} -p_i * \log_2 p_i$$

Tabel 3. Hasil Perhitungan Entropy

CABANG	BOBOT	DETAIL ATRIBUT	$p_i * \log_2(p_i)$ - PAS	$p_i * \log_2(p_i)$ - PRAL	$p_i * \log_2(p_i)$ - PRAK	Entropy (Total) (VPS)
1	SCHOOL	0.03	GP	0.03	0.04	0.03
			MS	0.00	0.00	0.00
2	SEX	0.59	F	0.40	0.45	0.02
			M	0.00	0.00	0.00
...
32	G2	0.84	0	0.00	0.11	0.11
			1	0.00	0.00	0.00
			2	0.00	0.00	0.00
			3	0.00	0.00	0.00
			4	0.00	0.00	0.00
			5	0.00	0.00	0.00
			6	0.00	0.00	0.00

- d. Menghitung nilai informasi gain dari entropi yang telah dihitung dengan persamaan sebagai berikut :

$$Gain(S,A) = Entropy(S) - \sum_{i=0}^{|n|} |S_i| / |S| * Entropy(S_i)$$

Tabel 4. Hasil Informasi Gain

Gain(S,A)	Gain(S,A)	Gain(S,A)	Gain(S,A)
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00

2. Model Fuzzy Sugeno

- a. Menentukan Variabel Input & Output
Variabel input yang akan digunakan dalam proses perhitungan fuzzy Sugeno diperoleh dari hasil perhitungan algoritma C.45 berbasis PSO yang membentuk suatu bentuk pohon keputusan.

Tabel 5. Contoh Tabel Variabel input

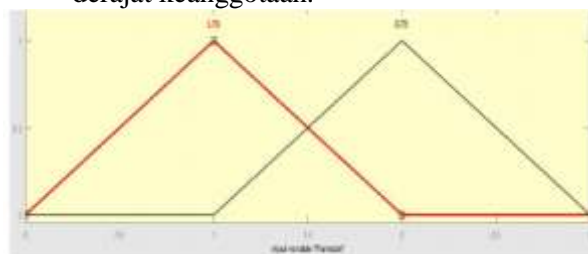
NO	VARIABEL INPUT	NILAI LINGUISTIK	INTERVAL
1.	G2	Fail Sufficient Satisfactory Good Excellent	0-10 9-12 11-14 13-16 15-21
2.	Fedu	None 4 th Education 5 th - 9 th Grade Secondary Education Higher Education	0-2 1-3 2-4 3-5 4-6

Tabel 6. Contoh Tabel Variabel output

NO	VARIABEL OUTPUT	NILAI LINGUISTIK	INTERVAL
1.	G3	FAIL PASS	0-11 10-21

- b. Proses Fuzzifikasi

Proses perubahan nilai variabel input menjadi sebuah nilai dalam bentuk derajat keanggotaan.



Gambar 6. Fuzzifikasi Variabel Famsize Fungsi Derajat Keanggotaannya sebagai berikut :

$$\mu_{L3} = \begin{cases} 0 & u < 0 \text{ atau } u > 2 \\ (u - 0) / (1 - 0) & 0 < u < 1 \\ (2 - u) / (2 - 1) & 1 < u < 2 \end{cases}$$

$$\mu_{R3} = \begin{cases} 0 & u < 1 \text{ atau } u > 3 \\ (u - 1) / (2 - 1) & 1 < u < 2 \\ (3 - u) / (3 - 2) & 2 < u < 3 \end{cases}$$

- c. Sistem Inferensi

Pada proses inferensi ini merupakan sebuah proses pembentukan suatu aturan untuk menghasilkan suatu pengelompokan kelas tertentu. Berikut contoh aturan yang terbentuk :

- IF G2 = SATISFACTORY & FEDU = PRIMARY MAKA G3 = PASS
- IF G2 = GOOD & FEDU = PRIMARY MAKA G3 = PASS
- IF G2 = EXCELLENT & FEDU = PRIMARY MAKA G3 = PASS
-
- IF G2 = FAILED & G1 = FAILED & DALC = VERY LOW & GOOUT = VERY LOW & FAMREL = VERY BAD MAKA G3 = FAILED

- d. Defuzzifikasi

Tahapan defuzzifikasi merupakan suatu tahapan yang memproses perubahan nilai fuzzy menjadi nilai tegas (crisp). Input dari tahap ini

adalah himpunan fuzzy yang diperoleh dari komposisi aturan – aturan fuzzy, sedangkan outputnya adalah suatu bilangan pada domain himpunan fuzzy tersebut. Adapun persamaannya adalah sebagai berikut :

$$Z = \frac{\alpha_1 \cdot z_1 + \alpha_2 \cdot z_2 + \dots + \alpha_n \cdot z_n}{\alpha_1 + \alpha_2 + \dots + \alpha_n}$$

3. Hasil dan Pembahasan

A. Hasil

1. Hasil Algoritma C4.5

Berikut ini merupakan tabel confusion matrix yang dihasilkan oleh sistem rapidminer pada algoritma C4.5 sebagai berikut :

Tabel 7. Confusion Matrix Algoritma C4.5

	true PASS	true FAIL	class precision
pred.PASS(1)	142	3	0,944
pred.FAIL(2)	1	45	0,978
class recall	0,981	0,938	

2. Hasil Algoritma C4.5 Berbasis PSO (Particle Swarm Optimization)

Berikut ini merupakan tabel confusion matrix yang dihasilkan oleh sistem rapidminer pada algoritma C4.5 berbasis PSO (Particle Swarm Optimization) sebagai berikut :

Tabel 8. Confusion Matrix Algoritma C4.5 + PSO

	true PASS	true FAIL	class precision
pred.PASS	100	0	90,00%
pred.FAIL	0	137	94,24%
class recall	90,00%	94,24%	

3. Pohon Keputusan



Gambar 7. Pohon Keputusan Algoritma C4.5

4. Implementasi Fuzzy Sugeno pada C4.5 berbasis PSO (Particle Swarm Optimization)

Berikut ini merupakan tabel confusion matrix yang dihasilkan oleh sistem rapidminer pada algoritma FIS Sugeno pada C4.5 berbasis PSO (Particle Swarm Optimization) sebagai berikut :

Tabel 9. Confusion Matrix FIS Sugeno pada C4.5 berbasis PSO

Akurasi = 0,960	true PASS	true FAIL	class precision
pred.PASS(1)	51	3	0,944
pred.FAIL(2)	1	45	0,978
class recall	0,981	0,938	

B. Pembahasan

1. Perbandingan Algoritma C4.5, Algoritma C4.5 Berbasis PSO dan Penerapan Fuzzy Sugeno pada Algoritma C4.5 berbasis PSO

Analisis yang dihasilkan dengan Algoritma C4.5, C4.5 berbasis PSO diuji menggunakan metode Cross Validation tenfolds. Sedangkan penerapan fuzzy sugeno pada algoritma C4.5 berbasis PSO menggunakan data secara acak sebanyak 100 data. Adapun perbandingan yang dilakukan yaitu nilai akurasi, spesifitas dan sensitifitas sebagai berikut :



Gambar 8. Komparasi Algoritma

Berdasarkan gambar diagram di atas menunjukkan bahwa hasil komparasi antara algoritma C4.5, C4.5 berbasis PSO dan hybrid fuzzy menggambarkan bahwa algoritma hybrid fuzzy menghasilkan nilai pengujian tingkat akurasi dan spesifitas paling tinggi diantara algoritma yang lainnya namun memiliki nilai sensitifitas lebih rendah 1 % dari algoritma C4.5 berbasis PSO, namun masih menunjukkan hasil yang sangat bagus untuk memprediksi kejadian benar. Berdasarkan hasil komparasi di atas menunjukkan bahwa dengan menggunakan pendekatan fuzzy dapat meningkatkan tingkat performa algoritma C4.5 berbasis PSO.

2. Komparasi Hasil Penelitian dengan Penelitian Lainnya

a. Komparasi Algoritma C4.5 dan C4.5 Berbasis PSO

Berikut ini merupakan hasil penelitian yang pernah dilakukan dalam

mengimplementasi algoritma C4.5 dan C4.5 Berbasis PSO untuk memprediksi kelulusan mahasiswa yang kami bandingkan dengan hasil penelitian kami :



Gambar 9. Komparasi Penelitian Algoritma C4.5 dengan C4.5 + PSO

Dari gambar diagram komparasi penelitian algoritma C4.5 dan C4.5 + PSO menunjukkan bahwa dengan menggunakan algoritma optimasi particle swarm optimization (psa) berhasil meningkatkan tingkat akurasi dari algoritma C4.5 itu sendiri, baik itu yang dilakukan peneliti maupun dilakukan oleh peneliti sebelumnya dalam hal ini yaitu Khoiril pada tahun 2013. Dari diagram di atas juga menggambarkan bahwa hasil yang diperoleh peneliti memberikan hasil yang lebih tinggi dibandingkan dengan penelitian yang dilakukan oleh Khoiril. Hal ini bisa terjadi dikarenakan beberapa hal :

1. Dari segi atribut yang digunakan, peneliti menggunakan atribut sebanyak 32 atribut sedangkan atribut yang digunakan oleh Khoiril sebanyak 7 atribut.
 2. Jumlah data yang digunakan, peneliti menggunakan sebanyak 299 data sedangkan Khoiril menggunakan sebanyak 1632 data.
 3. Pengaturan parameter yang ada pada tools rapid miner dalam penentuan pohon keputusan algoritma C4.5 maupun algoritma C4.5 berbasis PSO.
- b. Komparasi Hybrid Fuzzy

Berikut ini merupakan hasil penelitian yang pernah dilakukan dalam mengimplementasi hybrid fuzzy untuk memprediksi kelulusan mahasiswa yang akan dibandingkan dengan hasil penelitian kami:



Gambar 10. Komparasi Hybrid Fuzzy

Dari gambar diagram komparasi hybrid fuzzy menggambarkan hasil penelitian yang dilakukan oleh Heri Bambang pada tahun 2015 dengan menggunakan algoritma ID3 dengan fuzzy ID3 menunjukkan bahwa tingkat akurasi yang diperoleh mengalami peningkatan akurasi dari 93,41% menjadi 95,85% mengalami peningkatan sebesar 2,44%. Pada penelitian ini data yang digunakan sebanyak 403 data, dengan menggunakan 5 atribut yang bersifat kontinu, dan metode training yang digunakan 10-fold cross validation dengan software Weka 3.6 yang digunakan untuk pembentukan pohon keputusan. Penelitian ini dilakukan dengan melakukan perubahan fuzziness control threshold (θ_r) dan leaf decision threshold (θ_n) memperoleh kesimpulan bahwa semakin tinggi fuzziness control threshold (θ_r) dan semakin rendah leaf decision threshold (θ_n) maka tingkat akurasi yang dihasilkan semakin tinggi.

Berbeda dengan penelitian yang dilakukan oleh Abidatul pada tahun 2016 yang menggunakan algoritma fuzzy dengan hybrid fuzzy C4.5 menunjukkan hasil bahwa tingkat akurasi yang diperoleh mengalami peningkatan dari 67,92% menjadi 94,33 % dengan peningkatan sebesar 26,41 %. Adapun penelitian ini data yang digunakan sebanyak 106 record yang terdiri dari 4 atribut yang bersifat kontinu. Berdasarkan dari hasil tersebut menunjukkan bahwa prediksi FIS yang dibangkitkan berdasarkan rule yang dihasilkan oleh Decision Tree dapat membentuk suatu pengetahuan yang sangat baik untuk rule classifier FIS dibandingkan dengan rule yang dibentuk oleh pakar

berdasarkan experiment dan knowledge.

Dalam penelitian ini menunjukkan bahwa hasil yang diperoleh memperoleh tingkat akurasi dan spesifitas paling tinggi dibandingkan dengan penelitian yang dilakukan oleh peneliti – peneliti sebelumnya. Hal ini membuktikan bahwa dengan adanya penerapan algoritma C45 berbasis PSO untuk membangkitkan rule classifier FIS Sugeno mampu meningkatkan tingkat akurasi dari algoritma C4.5 berbasis PSO. Dalam penelitian ini data yang digunakan sebagai sebanyak 100 data yang terdiri dari 18 atribut yang bersifat kontinu dan diskrit.

4. Simpulan

Pada penelitian ini telah diimplementasikan fuzzy inference system (fis) metode sugeno pada algoritma decision tree C4.5 berbasis particle swarm optimization (pso) untuk memprediksi prestasi siswa. Data yang digunakan berasal dari UCI Repository hasil penelitian dari Paulo Cortez dan Alice Silva yang terdiri dari 32 atribut independen dan 1 atribut dependen. Dalam membangkitkan rule classifier FIS Sugeno peneliti menggunakan algoritma C4.5 berbasis PSO (Particle Swarm Optimization) dengan bantuan tool Rapid Miner Studio 8.0 menghasilkan sebuah pohon keputusan yang memiliki atribut 18 independen dan 1 atribut dependen. Kemudian rule yang diperoleh akan diimplementasikan pada FIS Sugeno untuk melakukan prediksi prestasi siswa dengan menggunakan tool matlab R2016b. Data yang diujikan pada hybrid fuzzy sebanyak 100 data dengan menghitung tingkat akurasi, spesifitas dan sensitifitas.

Berdasarkan hasil pengujian diperoleh hasil bahwa implementasi FIS metode Sugeno pada algoritma C4.5 berbasis PSO (Particle Swarm Optimization) mampu meningkatkan performa tingkat akurasi algoritma C4.5 berbasis PSO dari 94,66% menjadi 96% dan tingkat spesifitas dari 94,24% menjadi 97,8 %. Fuzzy memiliki keunggulan dalam hal perhitungannya yang tidak kaku (samar), sehingga mampu memperhitungkan kemungkinan tidak pasti sehingga memberikan hasil yang bagus dalam klasifikasi ataupun prediksi. Berdasarkan hasil tersebut membuktikan bahwa FIS Sugeno mampu diimplementasikan pada algoritma C4.5 berbasis PSO (Particle Swarm Optimization) dalam meningkatkan tingkat akurasi algoritma.

5. Penelitian Selanjutnya

Untuk mengembangkan penelitian ini, peneliti menghimbau kepada peneliti selanjutnya dapat mencoba dengan menambahkan algoritma optimasi pada *fuzzy inferency system* sebagai contoh yaitu genetika algoritma maupun algoritma *hill climbing* dan bisa juga dengan algoritma optimasi yang lainnya. Untuk penelitian selanjutnya dapat dilakukan proses seleksi atribut seperti dengan menggunakan jumlah atribut yang lebih komplek lagi dan tipe – tipe data digunakan lebih banyak menggunakan data kontinu yang digunakan dalam memprediksi prestasi siswa. Sehingga performa dari algoritma yang diterapkan dapat mengalami peningkatan.

Daftar Pustaka

- Fitri, D.L., & Darmanto, E. (2014). Sistem Pendukung Keputusan untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes. Seminar Prosiding SNATIF, Universitas Muria Kudus.
- Izzah, A., & Widyastuti, R. (2016). Prediksi Kelulusan Mata Kuliah Menggunakan Hybrid Fuzzy Inference System. Jurnal Ilmiah Teknologi Sistem Informasi.
- Kamagi, D.H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. Jurnal Ultimatics.
- Kamsyakawuni, A., Gernowo, R., & Sarwoko, E.A. (2012). Aplikasi Sistem Pakar untuk Diagnosa Penyakit Hipertiroid dengan Metode Inferensi Fuzzy Mamdani. Jurnal Sistem Infomasi Bisnis.
- Kemendiknas Pendidikan Nasional. (2009). Kriteria dan Perangkat Akreditasi Sekolah Menengah Kejuruan/Madrasah Aliyah Kejuruan (SMK/MAK)(Permendiknas Nomor 13 tahun 2009). Jakarta.
- Kemendiknas Pendidikan Nasional. (2009). Perangkat Akreditasi Program Studi Sarjana (S1) (Permendiknas Nomor 73 tahun 2009). Jakarta.
- Mauriza, A.F., & Nugorho, Y.S. (2014). Implementasi Data Mining untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan Metode Naïve Bayes. Skripsi, Teknik Informatika, Universitas Muhammadiyah Surakarta.
- Menakshi, & Nakar, P. (2015). Application of Fuzzy Logic for Evaluation of Academic Performance of Students of Computer Application Course. Ijrasnet.
- Mohammad, B. (2014). Perbandingan Algoritma C4.5 dan Neural Network untuk Memprediksi Hasil Pemilu Legislatif DKI Jakarta. Jurnal Piral Nusa Mandiri.

- Nurhidayat, F. (2013). Penentuan Besar Akurasi Metode Klasifikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm Optimization pada Prediksi Penyakit Diabetes. Skripsi, Teknik Informatika UDINUS.
- Permana, R.A. (2016). Seleksi Atribut Pada Metode Support Vector Machine untuk Menentukan Kelulusan Mahasiswa E-Learning. Jurnal Evolusi.
- Puji, A. (2016). Komparasi Penerapan Algoritma C4.5, KNN dan Neural Network dalam Proses Kelayakan Penerimaan Kredit Kendaraan Bermotor. Jurnal Faktor Exacta.
- Retno, A.S. (2013). Fuzzy Mamdani dalam Menentukan Tingkat Keberhasilan Dosen Mengajar. Seminar Nasional Informatika, UPN Veteran Yogyakarta.
- Sofi, D. (2015). Penentuan Peminatan Tugas Akhir Mahasiswa Teknik Informatika UNISKA. Jurnal Ilmiah Solusi.