

Implementasi *Data Mining* untuk Klasifikasi dan Pemerataan Guru pada Sekolah di Tangerang

Eko Sutono^{1*}, Anig Astofa¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ¹dosen02598@unpam.ac.id, ²dosen02360@unpam.ac.id

(* : coresponding author)

Abstrak– Permasalahan dalam dunia pendidikan yaitu kurangnya tenaga guru, guru yang mengajar tidak sesuai dengan latar belakang pendidikan (*mismatch*), kualifikasi rendah, disparitas kompetensi, dan penyaluran guru tidak merata. Penelitian ini bertujuan untuk (1) merancang model algoritma klasifikasi pemerataan kebutuhan guru dengan menggunakan algoritma nearest neighbor (2) Untuk merancang model algoritma klasifikasi pemerataan kebutuhan guru dengan menggunakan algoritma C4.5 (3) membandingkan akurasi model algoritma nearest neighbor dan algoritma C4.5 untuk pemerataan kebutuhan guru. Penelitian ini merupakan penelitian historis dengan menggunakan metode eksperimental yaitu dengan melakukan perancangan dan pembuatan model sistem. Pengumpulan data dilakukan dengan metode studi pustaka (*library research*) dan metode pengumpulan data lapangan (*field research*) dan pembuatan aplikasi berdasarkan analisis hasil dari metode data mining yakni algoritma nearest neighbor dan algoritma C4.5. Adapun hasil dari penelitian ini adalah informasi klasifikasi keadaan kebutuhan guru mata pelajaran apakah lebih, cukup atau kurang pada tiap sekolah. Tingkat akurasi algoritma nearest neighbor mencapai 72% dan tingkat akurasi algoritma C4.5 mencapai 83%.

Kata Kunci: *Data Mining*, Algoritma Nearest Neighbor, Algoritma C4.5

Abstract– Problems in education is the lack of teachers, the teachers are not in accordance with the educational background (*mismatch*), low qualifications, competence disparities and uneven distribution of teachers. This study aims to (1) design a model of classification algorithms equity needs of teachers by using nearest neighbor algorithm (2) To design a classification algorithm model distribution needs of teachers by using algorithm C4.5 (3) compare the accuracy of the model and algorithm nearest neighbor algorithm C4.5 equity requirement for teachers. This study aims to (1) design a model of classification algorithms of equity needs of teachers by using nearest neighbor algorithm (2) To design a classification algorithm models the distribution needs of teachers by using algorithm C4.5 (3) compare the accuracy of the models and nearest neighbor algorithm C4.5 algorithm equity requirement for teachers. The results of this study is the information in the form of pattern classification state whether teachers need more, or less enough of nearest neighbor algorithm and C4.5 algorithms. Nearest neighbor algorithm accuracy rate reaches 72% and the accuracy rate reaches 83%.

Keywords: *Data Mining*, Nearest Neighbor Algorithm, The Algorithm C4.5

1. PENDAHULUAN

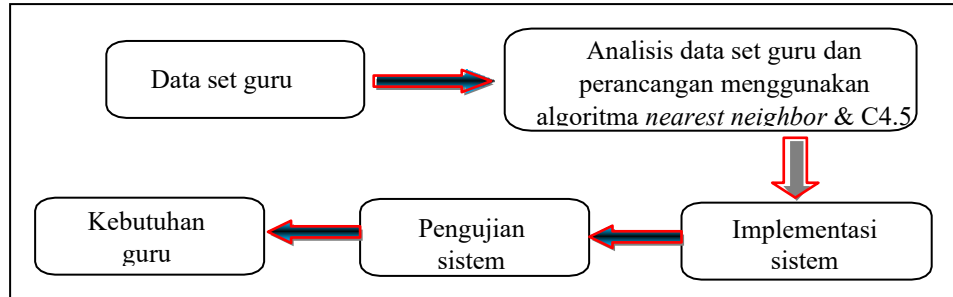
Dalam dunia pendidikan, peran dan fungsi guru merupakan salah satu faktor yang sangat signifikan. Guru merupakan bagian terpenting dalam proses pembelajaran, sesuai dengan Undang-undang Republik Indonesia Nomor 14 Tahun 2005 tentang Guru dan Dosen. Oleh sebab itu, dalam setiap upaya peningkatan kualitas pendidikan di tanah air tidak dapat dilepaskan dari berbagai hal yang berkaitan dengan eksistensi guru itu sendiri.

Guru mata pelajaran merupakan salah satu faktor penting dalam implementasi kurikulum. Bagaimanapun idealnya suatu kurikulum tanpa ditunjang oleh kemampuan guru untuk mengimplementasikannya, maka kurikulum itu tidak akan bermakna sebagai suatu alat pendidikan, dan sebaliknya pembelajaran tanpa kurikulum sebagai pedoman tidak akan efektif.

Di lain pihak, kondisi dunia pendidikan sekarang ini dihadapkan pada masalah yang kompleks diantaranya persoalan klasik, yakni kurangnya tenaga guru, guru yang mengajar tidak sesuai dengan latar belakang pendidikan (*mismatch*), kualifikasi rendah, disparitas kompetensi, dan pendistribusian guru yang tidak efektif. Hal ini dapat dibuktikan oleh kondisi Indonesia saat ini yang masih kekurangan 200.000 tenaga guru (Ditjen PMPTK, 2010).

2. METODOLOGI PENELITIAN

2.1 Tahapan Rancangan Penelitian



Gambar 1. Tahapan Rancangan Penelitian

2.2 Tahap Analisis Kebutuhan

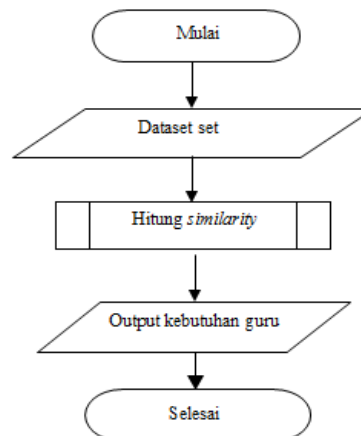
Pada tahap ini akan dilakukan analisis kebutuhan dan spesifikasi (*Requirement Analysis and Spesification*) atas masalah yang akan diselesaikan. Dimulai dengan mengidentifikasi data set guru yang meliputi data sekolah, data guru, data mata pelajaran, data rombongan belajar pada masing-masing sekolah.

2.3 Perancangan *Flowchart* Algoritma

a. Proses klasifikasi algoritma *nearest neighbor*

Langkah-langkah dalam proses klasifikasi metode *nearest neighbor*

1. Dimulai dengan menginput data set guru
2. Melakukan proses perhitungan similarity atau nilai kedekatan kedua kasus
3. Menampilkan output kebutuhan guru
4. Proses selesai



Gambar 2. *Flowchart* Klasifikasi Metode *Nearest Neighbor*

Tabel 1. Hasil Perhitungan *Similarity* dan *Output*

No	Sekolah	Mata Pelajaran	Alokasi jam	Guru Mapel	Jumlah Rombel			Similarity	Output
					1	2	3		
1	SMA Negeri 1	Agama islam	2	2	6	6	6	1.5	Lebih
		PPKN	2	3	6	6	6	1.5	Lebih
		Bahasa Indonesia	4	2	6	6	6	3	Kurang
		Bahasa	4	5	6	6	6	3	Lebih

		Inggris							
		Matematika	4	3	6	6	6	3	Cukup
2	SMA Negeri 2	Agama islam	2	2	6	6	6	1.5	Lebih
		PPKN	2	3	6	6	6	1.5	Lebih
		Bahasa Indonesia	4	2	6	6	6	3	Kurang
		Bahasa Inggris	4	2	6	6	6	3	Kurang
		Matematika	4	4	6	6	6	3	Lebih

Hasil *output* pada tabel 1 dihasilkan menggunakan aturan sebagai berikut :

- Jika Nilai atribut pada guru mata pelajaran lebih besar daripada hasil perhitungan alokasi waktu maka nilai *output* adalah Lebih
- Jika nilai atribut pada guru mata pelajaran lebih kecil daripada hasil perhitungan alokasi waktu maka nilai *output* adalah kurang.
- Jika nilai atribut guru mata pelajaran sama dengan hasil perhitungana lokasi waktu maka nilai *output* adalah cukup.

Pada tabel 1 juga merupakan tabel informasi mengenai alokasi waktu untuk sekolah pada tiap mata pelajaran sehingga memperoleh input apakah alokasi lebih, kurang atau cukup berdasarkan perhitungan nilai *similarity* atau jarak terdekat, kemudian membandingkan nilai *similarity* dengan guru mata pelajaran. Jika nilai *similarity* besar dibandingkan nilai guru mata pelajaran maka kurang, dan jika nilai *similarity* kecil dibandingkan nilai guru mata pelajaran maka lebih dan jika nilai *similarity* sama dengan nilai guru mata pelajaran maka cukup.

Jadi untuk contoh kasus SMU Negeri 3 dengan inputannya:

- Sekolah :SMU Negeri 3
- Mata Pelajaran : Agama Islam
- Alokasi Jam : 2
- Guru Mata Pelajaran : 4
- Jumlah Rombel 1, 2 dan 3 masing-masing : 3, 4, 3

Setelah dilakukan perhitungan *similarity* maka diperoleh nilai alokasi waktu adalah 2.5 dengan *output* lebih.

Penerapan Algoritma C4.5

Secara umum algoritma C4.5 untuk membangun pohon keputusan memilih atribut sebagai akar dan membuat cabang-cabang untuk tiap-tiap nilai. Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus sepertitertera dalam persamaan 1 dan perhitungan nilai *entropy*. Berikut contoh kasus pada tabel 2. untuk menentukan output kebutuhan guru.

Tabel 2. Kasus Kebutuhan Guru
(Sumber: Dinas Pendidikan Nasional Kota Tangerang)

Sekolah	No	Mata Pelajaran	KebutuhanGuru	JumlahGuru	Output
SMU Negeri 2	1	Agama Islam	Lebih	lebih	kurang
	2	Bahasa Indonesia	Cukup	Kurang	Kurang
	3	PPKN	Cukup	kurang	cukup
	4	Bahasa Inggris	Lebih	kurang	kurang
	5	Matematika	cukup	lebih	lebih
	6	Biologi	cukup	lebih	lebih

Algoritma Nearest Neighbor

Algoritma *nearest neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.

Rumus yang digunakan untuk menghitung kedekatan antara dua kasus yaitu:

$$\text{similarity}(T, S) = \frac{\sum_{i=1}^n f(T_i, S_i) * w_i}{w_i}$$

Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan variabel target. Ada 2 variabel yang dipakai dalam menentukan akar dari pohon keputusan yaitu nilai *entropy* dan nilai *gain*.

Nilai *Entropy* diperoleh dari rumus:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

3. ANALISA DAN PEMBAHASAN

3.1 Penerapan Algoritma *Nearest Neighbor*

Algoritma *nearest neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama.

Tabel 3. Klasifikasi Metode *Nearest Neighbor*

NO	SEKOLAH	MAPEL	ALOKASI JAM (a)	GURU MAPEL (b)	JUMLAH ROMBEL		
					1 (c1)	2 (c2)	3 (c3)
1	SMA NEGERI 1	AGAMA ISLAM	2	2	6	6	6
		PPKN	2	3	6	6	6
		BAHASA INDONESIA	4	2	6	6	6
		BAHASA INGGRIS	4	5	6	6	6
		MATEMATIKA	4	3	6	6	6
2	SMA NEGERI 2	AGAMA ISLAM	2	2	6	6	6
		PPKN	2	3	6	6	6
		BAHASA INDONESIA	4	2	6	6	6
		BAHASA INGGRIS	4	2	6	6	6
		MATEMATIKA	4	4	6	6	6

Untuk memprediksi apakah alokasi waktu cukup, kurang atau lebih untuk setiap sekolah dan masing-masing mata pelajaran pada tabel kasus, tabel 3 maka dilakukan langkah-langkah berikut:

Untuk mata pelajaran agama Gunakan rumus *similarity* pada

$$\begin{aligned}
 \text{similarity} &= \frac{(a * c1) + (a * c2) + (a * c3)}{c1+c2+c3} \\
 &= \frac{(2 * 6) + (2 * 6) + (2 * 6)}{6+6+6} \\
 &= 1.5
 \end{aligned}$$

3.2 Penerapan *Algoritma C4.5*

Secara umum algoritma C4.5 untuk membangun pohon keputusan memilih atribut sebagai akar dan membuat cabang-cabang untuk tiap-tiap nilai. Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 1 dan perhitungan nilai *entropy*. Berikut contoh kasus pada tabel 4.3. untuk menentukan *output* kebutuhan guru.

Tabel 4. Kasus Kebutuhan Guru
(Sumber: Dinas Pendidikan Nasional Kota Tangerang)

Sekolah	No	Mata Pelajaran	Kebutuhan Guru	Jumlah Guru	Output
SMU Negeri 2	1	Agama Islam	Lebih	lebih	kurang
	2	Bahasa Indonesia	Cukup	Kurang	Kurang
	3	PPKN	Cukup	kurang	cukup
	4	Bahasa Inggris	Lebih	kurang	kurang
	5	Matematika	cukup	lebih	lebih
	6	Biologi	cukup	lebih	lebih

Menghitung Jumlah kasus untuk kebutuhan guru yang kurang, cukup dan lebih dan entropy untuk semua kasus dan kasus yang dibagi berdasarkan atribut kebutuhan guru dan jumlah guru. Setelah itu lakukan perhitungan Gain untuk setiap atribut. Hasil Perhitungan ditunjukkan pada tabel perhitungan level 0.

Tabel 5. Perhitungan Level 0

Atribut	Nilai atribut	Jumlah kasus	Jumlah lebih	Jumlah cukup	Jumlah kurang	Entropy	Gain
		6	2	1	3	0.187635764	
Kebutuhan Guru							0.520969097
	lebih	2	0	0	2	0	
	cukup	4	2	1	1	0.5	
	kurang	0	0	0	0	0	
Jumlah Guru							0.187635764
	kurang	3	0	1	2	0	
	lebih	3	2	0	1	0	

4. IMPLEMENTASI

a. Input Data Mata Pelajaran

Tambah Data Mata Pelajaran

Kode MP		max 10 karakter
Nama MP		
Alokasi Waktu Belajar		
Semeeter	-- Pilih --	
Tahun	-- Pilih --	
	Simpan	Batal

Gambar 3. *Form Input Data Mata Pelajaran*

Form input data mata pelajaran untuk menginput kode mata pelajaran, nama mata pelajaran, alokasi waktu belajar, semester dan tahun. Data mata pelajaran yang telah di input kemudian akan di simpan pada database.

Hasil Perhitungan Algoritma C45

Jumlah Kasus = 8

[Jumlah Cukup = 2 || Jumlah Kurang = 3 || Jumlah Lebih = 3 || Entropy => 1.56127812446 || Gain => 6.43872187554]

NO	KODE SEKOLAH	TIPE SEKOLAH	KODE MP	ENTROPY JUMLAH GURU	ENTROPY KEBUTUHAN GURU	GAIN JUMLAH GURU	GAIN KEBUTUHAN GURU	HASIL
1	smn_01	A	BI	0.5	0.5	4.5	2.5	Lebih
2	smn_01	A	MTK	0.53063906223	0.5	3.53063906223	4.5	Kurang
3	smn_01	A	AI	0.53063906223	0.53063906223	3.53063906223	3.53063906223	Cukup
4	smn_01	A	BDNG	0.5	0.5	4.5	2.5	Lebih
5	smn_01	A	IPS	0.5	0.5	2.5	2.5	Cukup
6	smn_01	A	Perjas	0.375	0.5	1.375	2.5	Kurang
7	smn_01	A	IPA	0.53063906223	0.5	3.53063906223	2.5	Lebih
8	smn_01	A	MTK	0.5	0.5	2.5	4.5	Kurang

Gambar 4. Output Sistem Algoritma C4.5

Hasil Perhitungan Algoritma NN

NO	KODE SEKOLAH	TIPE SEKOLAH	KODE MP	JUMLAH GURU	KEBUTUHAN GURU	NILAI KEDEKATAN	NILAI JARAK JUMLAH GURU	NILAI JARAK KEBUTUHAN GURU	HASIL
1	s-001	B	BI	2	1	3	2	1	Lebih

Gambar 5. Output Sistem Nearest Neighbor

Tabel 6. Pengujian Akurasi Algoritma Nearest Neighbor

No	Nama Sekolah	Mata Pelajaran	Data Prediksi	Data Asli	PrediksiBenar
1.	SMA Negeri 31	Agama Islam	Lebih	Lebih	Benar
		Bahasa Indonesia	Lebih	Lebih	Benar
		PPKn	Lebih	Lebih	Benar
2.	SMA Negeri 4	Agama Islam	Cukup	Cukup	Benar
		Bahasa Indonesia	Cukup	Lebih	Salah
		PPKn	Cukup	Cukup	Benar
3.	SMA Negeri20	Agama Islam	Lebih	Lebih	Benar
		Bahasa Indonesia	Cukup	Cukup	Benar
		PPKn	Kurang	Cukup	Salah
4.	SMA Negeri 1	Agama Islam	Kurang	Cukup	Salah
		Bahasa Indonesia	Cukup	Cukup	Benar
		PPKn	Cukup	Cukup	Benar
5.	SMA Negeri 2	Agama Islam	Cukup	Cukup	Benar
		Bahasa Indonesia	Lebih	Lebih	Benar
		PPKn	Lebih	Kurang	Salah
6.	SMA Negeri 7	Agama Islam	Cukup	Cukup	Benar
		Bahasa Indonesia	Kurang	Kurang	Benar
		PPKn	Kurang	Cukup	Salah

Berisi hasil implementasi ataupun pengujian.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut

1. Model klasifikasi dari algoritma *nearest neighbor* dan C4.5 diperoleh bahwa kebutuhan guru pada kota Makassar dan kabupaten Maros tidak merata penyebarannya.
2. Dari hasil perbandingan algoritma *nearest neighbor* dan algoritma C4.5, dengan masing-masing nilai keakuratan 72% dan 83%, maka nilai keakuratan algoritma C4.5 lebih besar sehingga hasil prediksi dari algoritma C4.5 lebih akurat dibandingkan dengan algoritma *nearest neighbor*.

REFERENCES

- Petunjuk Teknis (Juknis) *Pelaksanaan Peraturan Bersama Lima Menteri tentang penataan dan Pemerataan Guru PNS*, November 2011.
- Sanjaya, Wina. DR.M.Pd.2006. *Pembelajaran Dalam Implementasi Kurikulum Berbasis Kompetensi*. Jakarta: Kencana.
- Kusrini, Emha T. Luthfi. "Algoritma Data Mining". Penerbit ANDI, 2009
- Budi Santosa, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis". Penerbit 2007.
- Ari Kurniawan, Mochamad Hariadi. "Klasterisasi kompetensi guru menggunakan hasil penilaian portofolio sertifikasi guru dengan metode data mining". Jurnal Teknik Elektro (Telematika), Institut Teknologi Sepuluh Nopember Surabaya.
- Priadi Surya. "Pemetaan Pendidikan (*Education Mapping*) Sebagai Dasar Meningkatkan Layanan Pendidikan". Universitas Negeri Yogyakarta dimuat dalam Makalah ICEMAL (*International Conference Educational Management, Administration and Leadership*), 4-5 Juli 2012.
- Hukmiah Arif. "Pemetaan Mutu Pendidik dan Tenaga Kependidikan Berbasis Spasial". Tesis Program Pasca Sarjana Universitas Hasanuddin Makassar. 2011.
- Ninki Hermaduanty, Sri Kusumadewi. "Sistem Pendukung Keputusan berbasis SMS untuk menentukan status gizi dengan metode k-nearest neighbor". Fakultas Teknologi Industri Universitas Islam Indonesia.
- Kusrini, Sri Hartati, Retantyo Wardoyo, Agus Harjoko "Perbandingan metode Nearest Neighbor dan algoritma C45 untuk menganalisis kemungkinan pengunduran diri calon mahasiswa di STMIK AMIKOM Yogyakarta", Yogyakarta, Jurnal DASI Vol. 10 No. 1 Maret 2009, ISSN:1411-3201.
- Sunjana. "Klasifikasi Data Nasabah sebuah Asuransi menggunakan Algoritma C4.5". SNATI 2010 Yogyakarta, 19 Juni 2010.
- Maulani Kapiudin. "Data mining untuk klasifikasi pelanggan dengan AntColony Optimization". Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra, yang dimuat pada paper informatika Petra.
- Roger S. Pressman. "Rekayasa Perangkat Lunak Pendekatan Praktisi (buku satu)". Penerbit Andi Yogyakarta. 2002.
- Yuni Sugiarti. "Analisis dan Perancangan UML (Unified Modelling Language) Generated VB.6". Penerbit Graha Ilmu Yogyakarta. 2013.