

IMPLEMENTASI DATA MINING UNTUK PREDIKSI DAERAH RAWAN PENYAKIT DEMAM BERDARAH MENGGUNAKAN ALGORITMA C45

T. Adi Kurniawan¹, Adyatama Lubis², Ahrizal Ihfandi³
Universitas Satya Negara Indonesia^{1,3}, STIKAP Rokania²
t.adikurniawan@gmail.com¹, Aynatalubis@gmail.com²,
rizalequal42@gmail.com³

ABSTRAK

Sebuah aplikasi yang mengimplementasikan algoritma C4.5 dibangun dengan menggunakan bahasa pemrograman PHP&MYSQL. Tujuan utama penelitian adalah membuat aplikasi untuk memprediksi dan pengecekan data dengan tepat dan tidak memakan waktu yang lama. Pada penelitian yang dilakukan, diimplementasikan algoritma C4.5 untuk membuat suatu pohon keputusan untuk memprediksi data dengan beberapa atribut di dalam data tersebut. Metode yang digunakan antara lain berupa studi literatur, perancangan diagram-diagram, implementasi, serta uji coba dan evaluasi. Hasil penelitian ini mengungkapkan bahwa mengklasifikasikan data dengan algoritma C4.5 dengan membuat pohon keputusan serta atribut data didalamnya akan menghasilkan data yang valid dan tidak memakan proses waktu yang lama.

Kata Kunci : Algoritma C4.5, Demam Berdarah, Kelurahan.

ABSTRACT

An application that implements the C4.5 algorithm is built using the PHP & MYSQL programming language. The main objective of the research is to make an application to predict and check data properly and not take a long time. In the research conducted, C4.5 algorithm is implemented to make a decision tree to predict data with several attributes in the data. The methods used include literary studies, diagramming, implementation, and testing and evaluation. The results of this study reveal that classifying data with C4.5 algorithm by making decision trees and attribute data in it will produce valid data and not take a long time process.

Keywords: Algorithm C4.5, Dengue Fever, Village.

PENDAHULUAN

Demam berdarah merupakan salah satu jenis penyakit menular yang terdapat di wilayah tropis maupun subtropis. Demam berdarah dapat menular karena gigitan Nyamuk Aedes Aegypti. Dampak dari Demam Berdarah dapat membuat suhu tubuh penderita menjadi sangat tinggi dan pada umumnya disertai demam, mual/muntah, sakit kepala, nyeri perut, dan lekopenia. Penyakit Demam Berdarah masih merupakan salah satu permasalahan kesehatan masyarakat. Dinas Kesehatan mempunyai tugas pokok untuk membantu penyelenggaraan kegiatan penyehatan lingkungan.

Semakin meningkatnya kepadatan penduduk di Kabupaten Tangerang pada tahun 2016 berjumlah 3.477.495 jiwa tersebar di 29 Kecamatan, meningkatnya penderita

penyakit Demam Berdarah karena kurangnya kesadaran masyarakat terhadap kesehatan lingkungan. Timbulah masalah yang menyulitkan Dinas Kesehatan untuk mengelompokan daerah yang akan menghasilkan titik-titik pusat penyebaran penderita Demam Berdarah. Agar program kegiatan penyehatan lingkungan berjalan dengan baik dan mudah dilakukan. Maka dilakukan pengelompokan berdasar jumlah penduduk, jumlah penderita. Supaya program kerja Dinas Kesehatan bisa melakukan budi daya tamanan anti nyamuk, program 3M (menguras, menutup, mengubur) yang bertujuan untuk menurunkan angka penderita penyakit Demam Berdarah dan penyuluhan untuk memberi informasi tentang bahaya penyakit Demam Berdarah.

RUMUSAN MASALAH

Berdasarkan latar belakang dapat dirumuskan permasalahan sebagai berikut: “Bagaimana penerapan data mining untuk memprediksi daerah rawan penyakit Demam Berdarah menggunakan algoritma C45?”.

TUJUAN PENELITIAN

Ada pun tujuan dari penelitian ini adalah :

- 1) Untuk prediksi sebaran daerah rawan yang terkena penyakit Demam Berdarah se-Kabupaten Tangerang.
- 2) Untuk mempermudah Dinas Kesehatan dalam penentuan daerah rawan penyakit Demam Berdarah.
- 3) Mempercepat proses pelaporan dalam menindak lanjuti penanganan dilokasi yang tingkat penyebaran tinggi terhadap penyakit Demam Berdarah

TINJAUAN PUSTAKA

Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma pohon Keputusan (decision tree). Algoritma ini mempunyai input berupa training samples dan samples. Training samples data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal bisa mengatasi missing data, bisa mengatasi data kontinyu, pruning . Algoritma C4.5 memiliki kelebihan yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan)

Secara umum algoritma C4.5 Untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus:

$$Gain S, A = Entropy S - \sum_{i=1}^n |S_i| |S| * (S_i).....$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|Si|: jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Sedangkan untuk menghitung entropy digunakan rumus:

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

Keterangan:

S: himpunan kasus

A: fitur

n: jumlah partisi S

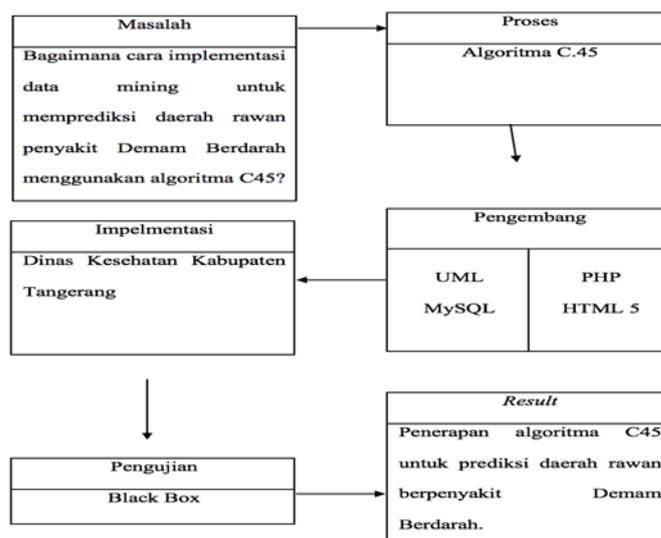
pi: proporsi dari Si terhadap S.

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, dkk. 2005).

Kerangka Pemikiran

Dalam kerangka berfikir tersebut dapat dijelaskan bahwa bagaimana cara mebgimplementasikan datamining untuk memprediksi sebaran daerah rawan penyakit demam berdarah menggunakan algoritma C45. Dengan menggunakan algoritma C45 diharapkan dapat menjadi salah satu alternative solusi untuk mengatasi permasalahan sebaran penyakit demam berdarah di wilayah kota Tangerang. Untuk memudahkan proses implementasi maka program dikembangkan dengan menggunakan UML MySQL dan PHP HTML 5. Selanjutnya Penerapan algoritma C45 untuk prediksi sebaran daerah rawan berpenyakit Demam Berdarah dapat di implementasikan pada dinas Kesehatan Kabupaten Tangerang dan Pengujian system menggunakan Black Box.

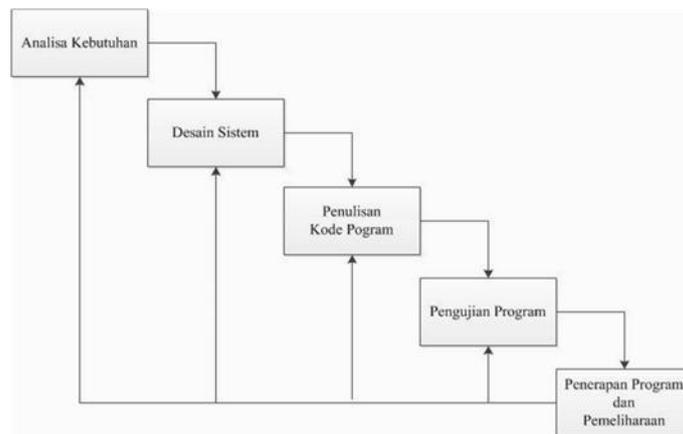


Gambar 1.1. Kerangka berpikir

METODE PENELITIAN

Metode Pengembangan Sistem

Metode yang digunakan penulis adalah menggunakan sistem waterfall. Metode waterfall merupakan metode yang sering digunakan oleh penganalisa sistem pada umumnya. Inti dari metode waterfall adalah pengerjaan dari suatu sistem dilakukan secara berurutan atau secara linear. Jadi jika langkah ke-1 belum dikerjakan, maka langkah 2 tidak dapat dikerjakan. Jika langkah ke-2 belum dikerjakan maka langkah ke-3 juga tidak dapat dikerjakan, begitu seterusnya. Secara otomatis langkah ke-3 akan bisa dilakukan jika langkah ke-1 dan ke-2 sudah dilakukan (Pressman, 2012).



Gambar 4. Model Waterfall

Analisa Kebutuhan

Langkah ini merupakan analisa terhadap kebutuhan sistem. Pengumpulan data dalam tahap ini bisa melakukan sebuah penelitian, wawancara atau studi literatur. Sistem analis akan menggali informasi sebanyak-banyaknya dari user sehingga akan tercipta sebuah sistem komputer yang bisa melakukan tugas-tugas yang diinginkan oleh user tersebut. Tahapan ini akan menghasilkan dokumen user requirment atau bisa dikatakan sebagai data yang berhubungan dengan keinginan user dalam pembuatan sistem.

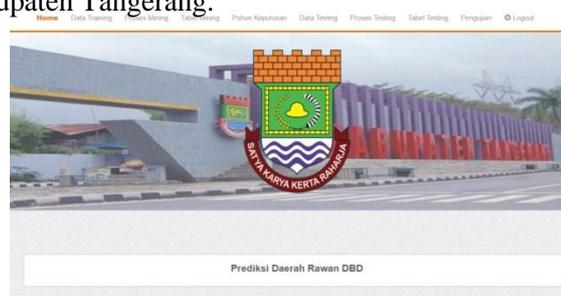
Desain Sistem

Tahapan dimana dilakukan penuangan pikiran dan perancangan sistem terhadap solusi dari permasalahan yang ada dengan menggunakan perangkat pemodelan Unified Modeling Language (UML)

HASIL DAN PEMBAHASAN

Hasil implementasi Program

Pada hasil implementasi program tersebut terlihat sebuah tampilan menu utama dari dinas kesehatan kabupaten Tangerang.



Gambar 5. Tampilan Halaman Utama

Tampilan Halaman Data Testing

Gambar 1 Tampilan Halaman Data Testing

Pada tampilan data testing kita dapat memasukkan ID_Wilayah selanjutnya nama wilayah, jumlah penduduk, jumlah penderita, jumlah rata-rata penduduk per kilometer, dan status.

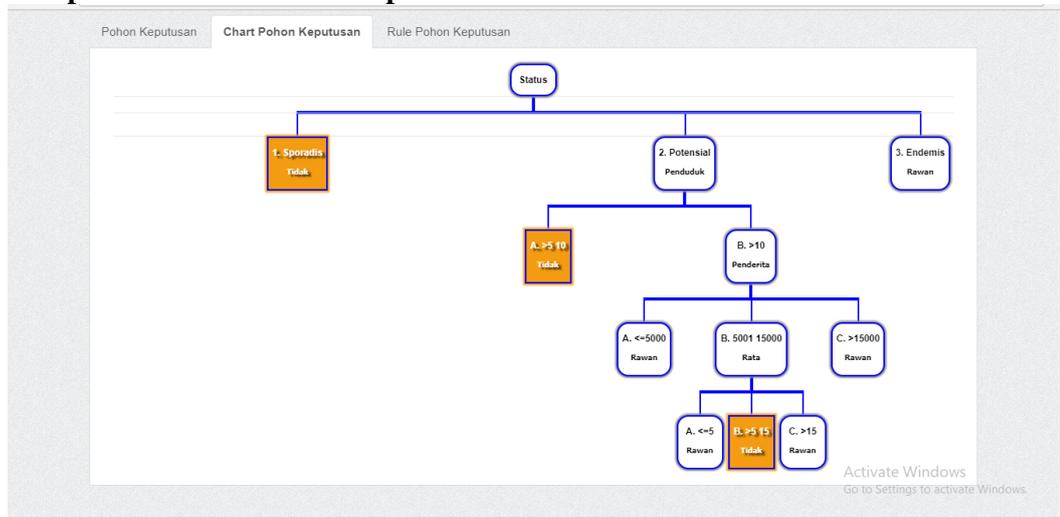
Tampilan Halaman Pengujian

Selanjutnya melalui database system kita dapat lakukan implementasi datamining serta penerapan algoritma C45 untuk memprediksi sebaran penyakit Demam Berdarah di Wilayah Kota Tangerang.

No	ID Wilayah	Nama Daerah	Jumlah Penduduk			Jumlah Penderita		Rata-Rata Penduduk			Status	Keputusan Asli	Keputusan C45	ID Rule
			<=5000	5001-15000	>15000	<=10	>10	<=5	>5-15	>15				
1	1	Mekar Wangi	4500	-	-	7	-	-	7.77	-	2. Potensial	Tidak	Tidak	3
2	2	Dangdang	-	-	25000	-	20	-	-	19.56	2. Potensial	Tidak	Rawan	10
3	3	Budimulya	-	-	27800	-	13	-	14.24	-	3. Endemis	Rawan	Rawan	11
4	4	Bunder	-	13000	-	-	12	3.98	-	-	1. Sporadis	Tidak	Tidak	1
5	5	Dukuh	-	7540	-	-	25	-	9.49	-	2. Potensial	Tidak	Tidak	2

Gambar 7. Tampilan Halaman Pengujian

Tampilan Halaman Pohon Keputusan C4.5



Gambar 8. Tampilan Halaman Pohon Keputusan C4.5

Perhitungan C4.5

Pada rancangan proses akan dilakukan bagaimana menghitung data penderita demam berdarah sesuai dengan algoritma yang digunakan oleh peneliti yaitu dengan algoritma C4.5, menggunakan set data training penduduk seperti pada tabel 4.1 dibawah ini :

Tabel 1. Tabel Data Penderita Dbd

NO	KELURAHAN	JUMLAH PENDERITA	JUMLAH PENDUDUK	RATA-RATA PENDUDUK PER KM	status	keterangan
1	Gembong	15	12685	12,24	potensial	tidak
2	Cangkudu	31	19036	11,94	endemis	rawan
3	Sentul	12	10744	10,24	potensial	tidak
4	Sentul Jaya	10	9508	9,11	potensial	tidak
5	Talaga Sari	25	11561	6,06	endemis	rawan
6	Balaraja	9	12577	9,11	sporadis	tidak
7	Tobat	23	15804	15,05	endemis	rawan
8	Suka Murni	8	6477	9,39	sporadis	tidak
9	Saga	30	33174	16,22	endemis	rawan
10	Mekar Wangi	8	4517	1038	sporadis	tidak
11	Budimulya	15	4892	1482	potensial	rawan
12	Bojong	19	14980	5233	potensial	tidak
13	Suka Mulya	34	30386	10420	endemis	rawan
14	Cikupa	21	25978	12691	endemis	rawan
15	Dukuh	10	17589	5314	potensial	tidak
16	Bitung Jaya	18	14060	8053	potensial	tidak
17	Bunder	9	16373	3149	sporadis	tidak
18	Suka Damai	30	22944	5093	endemis	rawan
19	Pasir Jaya	20	34088	14086	potensial	rawan

20	Pasir Gadung	23	24681	9420	potensial	rawan
21	Talaga Sari	10	25244	9329	potensial	tidak
22	Talaga	20	20607	5007	potensial	rawan
23	Sukanagara	26	11692	5443	endemis	rawan
24	Cibadak	29	12203	4358	endemis	rawan
25	Dangdang	9	4435	8,65	sporadis	tidak
26	Jeungjing	22	13362	5273	endemis	rawan
27	Cisoka	10	9482	3533	potensial	tidak
28	Sukatani	25	13812	6472	endemis	rawan
29	Cempaka	9	7562	3092	sporadis	tidak
30	Karangharja	7	5578	2441	sporadis	tidak
31	Carenang	8	6645	1282	sporadis	tidak
32	Bojongloa	8	6619	2233	sporadis	tidak
33	Caringin	9	9290	3473	sporadis	tidak
34	Slapajang	25	11493	3501	potensial	rawan
35	Cibugel	20	10273	3962	potensial	rawan
36	Sukatani	14	13812	6472	Potensial	Rawan

Menghitung Nilai Total Kasus Pertama

Lakukan perhitungan untuk jumlah kasus atribut peringkat, jumlah kasus untuk hasil beasiswa dan juga jumlah kasus untuk hasil tidak. Kemudian lakukan perhitungan untuk mencari nilai *Entropy* dari setiap nilai atribut dan nilai *Gain* untuk setiap atribut.

Jumlah kasus atribut kasus : 36

Jumlah kasus dengan status tidak rawan : 18

Jumlah kasus dengan status rawan : 18

Hitung nilai *entropy* dari total kasus :

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

Entropy(S)total

$$= \left(\left(-\frac{\text{total tidak}}{\text{total kasus}} \right) * \log_2 \left(\frac{\text{total tidak}}{\text{total kasus}} \right) \right) + \left(\left(-\frac{\text{total rawan}}{\text{total kasus}} \right) * \log_2 \left(\frac{\text{total rawan}}{\text{total kasus}} \right) \right)$$

$$Entropy(S) \text{ total} = ((-18/36)*\log_2 (18/36)) + ((-18/36)*\log_2 (18/36)) = 1$$

Tabel 2 Atribut Kasus Pertama

NODE	Atribut	Nilai	Jml Kasus	Tidak	Rawan	Entropi	Gain
			(S)	(S1)	(S2)		
1	Total		36	18	18	1	
	JUMLH PENDERITA						0.581977
		>5-10	14	14	0	0	

		>10	22	4	18	0.6840 38	
	JUMLH PENDUDUK						0.0395 62
		<=5000	3	2	1	0.9182 96	
		5001- 15000	19	11	8	0.9819 41	
		>15000	14	5	9	0.9402 86	
	RATA-RATA PENDUDUK						0.1115 23
		<=5	12	8	4	0.9182 96	
		>5-15	21	10	11	0.9983 64	
		>15	3	0	3	0	
	STATUS						0.5846 7
		sporadis	11	0	11	0	
		Potensial	15	8	7	0.9967 92	
		endemis	10	10	0	0	

KESIMPULAN DAN SARAN

Kesimpulan

Kesimpulan yang dapat diperoleh pada pengembangan sistem dengan menggunakan metode prediksi c45 yaitu dapat digunakan untuk memprediksi daerah rawan atau tidak rawan di kabupaten tangerang, Hasil tingkat keberhasilan prediksi daerah rawan secara keseluruhan yang telah diukur adalah sebesar 91% valid. Yaitu untuk daerah Pasir Gadung, Talaga Dan Suka Nagara.

Saran

Adapun saran dari penyusunan Penelitian ini adalah :

1. Pengimplementasian ini seharusnya mengambil data yang lebih besar untuk mendapatkan hasil yang lebih akurat.
2. Pada penelitian selanjutnya dapat dikembangkan dengan perhitungan algoritma C4.5 lebih dari 4 atribut.
3. Dapat dikembangkan dengan macam-macam metode selain C4.5.

DAFTAR PUSTAKA

- Abdillah, S. (n.d.). Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pematang.
- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes . *Jurnal Techno Nusa Mandiri Vol. XIII, No. 1*, 50-59.
- Munawar. (2005). *Pemodelan Visual dengan UML*. Yogyakarta: Graha Ilmu.

- Nugroho, B. (2008). *Aplikasi Pemrograman Web Dinamis dengan PHP dan MySQL*. Yogyakarta: Gava Media.
- Prasetyo, E. (2012). *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB*. Gresik: Andi.
- Pressman. (2012). *Rekayasa Perangkat Lunak*. Yogyakarta: Andi.
- Rusda Wajhilla, Ihsan. (2015). Penerapan Algoritma C4.5 Terhadap Diagnosa Penyakit Demam Tifoid Berbasis Mobile . *Swabumi Vol III NO. 1*, 24-30.
- Tan P, Steinbach M, Kumar V. (2006). Introduction to Data Mining. In E. Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB* (p. 2). Gresik: ANDI Yogyakarta.
- Widoyono. (2008). Demam Berdarah. 60-63.