

KLASIFIKASI DATA KEJADIAN LUAR BIASA CAMPAK MENGUNAKAN METODE *DECISION TREE C4.5*

Sulistyowati

Teknik Informatika, STMIK Palangkaraya
e-mail: sty.kireina@yahoo.com

ABSTRAK

Klasifikasi merupakan sebuah teknik pembelajaran yang ada pada data mining yang dapat membangun sebuah aturan yang digunakan untuk klasifikasi. Pada penelitian ini, klasifikasi yang dilakukan adalah untuk mengetahui daerah yang menderita kejadian luar biasa (KLB) atau non KLB berdasarkan data surveilans penyakit campak di suatu wilayah. Data diperoleh dari Dinas Kesehatan Provinsi DIY dan jumlah data yang dikumpulkan adalah 648 data yang selanjutnya telah dibagi kedalam 2 bagian yaitu data pembelajaran dan data pengujian. Permasalahan dalam penelitian ini adalah bagaimana algoritma *decision tree C4.5* dapat melakukan klasifikasi KLB atau non KLB. Tujuan dari pengklasifikasian yang dilakukan adalah untuk mengetahui kinerja algoritma *decision tree c.45* dalam melakukan klasifikasi data KLB suatu penyakit. Hasil penelitian menunjukkan bahwa dari 500 data diperoleh akurasi sebesar 84.4037%. Berdasarkan hasil data akurat yang diperoleh dari testing tersebut maka dilakukan pengujian menggunakan 10-Fold Cross Validation, algoritma *decision tree c4.5* mampu melakukan klasifikasi sebesar 96%. Sehingga dapat dikatakan algoritma ini dapat digunakan untuk melakukan klasifikasi.

Kata kunci: KLB, Campak, Decision Tree C4.5

ABSTRACT

Classification is a technique of learning on data mining which is given by classifier to build a rule of classification. Classification that has been done for understanding the region is KLB or Non KLB base on surveillance data from measles diseases in a region. The data got from health Departmen DIY province and the data collects are 648 datas, and then divided by 2 path, they are data training and data testing. The problem in a paper is how the Decision Tree C.45 can used to classify in a KLB or Non KLB . The purpose of clasification to understand the performance of Decision tree C.45 algorithm in doing classification of KLB Region a disease. Classification of tthis paper is to clasify Kejadian Luar Biasa Campak of the KLB or Non KLB. The result of classification was done, the level performance is 84.4037%, base on the result of data testing in a 10-Fold Cross Validation, the result is 96%. So the algorithm can be used to clasify for the KLB data.

Keywords : Measless Event Outbreak, Decision Tree C.45, Classification

I. PENDAHULUAN

Saat ini telah banyak bencana yang menimpa kehidupan, baik berupa bencana penyakit, banjir maupun hal hal yang yang menimbulkan wabah sehingga suatu daerah masuk dalam kategori Kejadian Luar Biasa (KLB) merupakan timbulnya atau meningkatnya kejadian Kesakitan dan atau kematian yang bermakna secara epidemiologis pada suatu daerah dalam kurun waktu tertentu [1]. Timbulnya suatu penyakit di suatu daerah adalah dipengaruhi oleh bebrapa faktor yang dapat menyebar ke suatu daerah karena suatu pencegahan yang tidak maksimal. Menurut peraturan menteri bahwa penyakit penyakit yang berpotensi KLB biasanya adalah penyakit menular seperti demam berdarah dengue (DBD), campak, malaria, diare, dan berbagai macam penyakit yang ditimbulkan oleh suatu virus atau bahkan seperti keracunan makanan. Walaupun jenis penyakit tersebut sering menimbulkan KLB di Indonesia, ada beberapa penyakit yang mengalami penurunan seperti diare, campak dan malaria. Namun, ada beberapa penyakit yang kini justru semakin meningkat yaitu DBD, keracunan makanan dan penyakit baru seperti SARS, HFMD, hepatitis E dan sejenisnya. Demikian juga beberapa penyakit yang sudah tidak dianggap masalah namun kini malah timbul kembali seperti differi, chikungunya, leptoperosis dan kolera.

Klasifikasikan ke dalam daerah yang terjadi KLB atau daerah yang tidak KLB diperlukan agar dapat membantu penyelidikan terhadap terjadinya suatu daerah KLB. Selain itu juga dapat digunakan sebagai penyelidikan awal sebelum dilakukan pengambilan data surveilans kepada penduduk sehingga lebih mudah dalam melakukan pencegahan terhadap daerah-daerah yang memang dekat dengan terjadinya KLB campak tersebut.

Metode pengolahan data yang digunakan klasifikasi ini adalah algoritma *decision tree C4.5*. Metode ini merupakan salah satu metode klasifikasi yang populer karena kemampuannya dalam melakukan klasifikasi dan juga dapat menunjukkan hubungan antar atribut [2]. Selain itu algoritma C4.5 secara rekursif dapat menelusuri setiap *node*, memilih *split* yang optimal, sampai tidak ada pembagian *split* yang mungkin terjadi [3].

Klasifikasi merupakan sebuah teknik pembelajaran pada *data mining*. Data *training* diberikan kepada *clasifier* untuk membangun aturan klasifikasi. Saat ini algoritma C4.5 diganti nama sebagai J48 classifier di *tool WEKA*, yaitu tools yang digunakan dalam pengolahan *data mining* yang bersifat *open source*. Fungsi heuristik yang digunakan dalam klasifikasi ini didasarkan pada konsep informasi entropi. Klasifikasi yang dilakukan adalah untuk

mengetahui daerah yang berpotensi KLB berdasarkan data surveilen suatu penyakit disuatu daerah. Permasalahan yang timbul adalah bagaimana mengelompokkan suatau daerah yang berpotensi KLB atau non KLB. Tujuan dari klasifikasi yang dilakukan adalah untuk mengetahui tingkat performansi dari algoritma *decission tree C.45* dalam melakukan klasifikasi daerah yang berpotensi KLB suatu penyakit.

Penelitian sebelumnya telah dilakukan [4] mengenai peramalan KLB campak yang dilakukan di DI. Yogyakarta. dengan menggunakan metode JST dan CART. Dalam penelitian tersebut hanya memprediksi kejadian campak. Penggunaan *decission tree C4.5* telah dikembangkan dengan teknik atribut *Oriented Induction (AOI)* dan analisis relevansi yang digabungkan dengan konsep hirarki pengetahuan dan algoritma Height Balance Priority untuk pembangunan pohon keputusan bersama dengan *Mining-Multi Level*. Selanjutnya [5] telah membandingkan algoritma klasifikasi antara C4.5 dan C5.0 pada data penderita kanker *teroid*. Hasil yang diperoleh dari perbandingan tersebut adalah pebandingan banyaknya *rule* yang diperoleh. Banyaknya *rule* yang dihasilkan oleh C5.0 lebih sedikit dibandingkan dengan C4.5. Namun algoritma C4.5 [6] memerlukan waktu yang lama jika data yang akan dibuat pohon keputusan sangat besar.

II. METODE

Pada penelitian ini digunakan data surveilen terhadap suatu penyakit menular yang berpotensi KLB. Data yang diperoleh akan dilakukan *pre-prosesing* untuk membersihkan data hingga menghasilkan data *training* dan data *testing*. Adapun tahap-tahap yang dilakukan dalam penelitian ini adalah sebagai berikut.

A. Pengumpulan Data

Pengumpulan dilakukan melalui wawancara dan juga observasi lapangan. Adapun data yang diperoleh adalah data berupa data kasus KLB, jumlah anak, dan data gizi. Selain itu juga data-data yang terkait dengan hal atau faktor-faktor yang mempengaruhi daerah terkena KLB seperti data imunisasi terhadap anak bawah lima tahun (balita) dan batita (bawah tiga tahun), data cakupan gizi, dan cakupan vitamin A yang sudah diberikan oleh puskesmas kepada masyarakat. Data lain yang diperlukan adalah Perilaku Hidup Bersih dan Sehat (PHBS) di setiap kecamatan.

B. Analisis

Data yang diperoleh dilakukan teknik analisis yaitu data akan dipilih dan dilakukan *pre-prosesing* data. Adapun tahapan data *pre-prosesing* yang dilakukan adalah sebagai berikut :

- 1) Pembersihan data (*cleaning data*), yaitu melakukan pembersihan terhadap data yang telah diperoleh dengan cara menghilangkan data yang dianggap mempunyai nilai yang sama, data yang tidak lengkap, dan juga data yang tidak sesuai.
- 2) Setelah data diperoleh maka data data tersebut akan dilakukan intregasi dan transformasi data, yang dilakukan pada tahapan ini adalah data akan diintegrasikan dengan data yang sesuai dan ditransformasikan kedalam satu kesatuan hingga menjadi data yang saling terintegrasi dan berhubungan antara data yang satu dengan yang lainnya. Adapun hasil data terintegrasi dapat dilihat pada Tabel I. Terdapat 14 kecamatan yang digunakan sebagai sampel data penelitian ini dimana masing-masing kecamatan mempunyai 7 buah atribut non kelas.

TABEL I
3) HASIL PEMBERSIHAN DATA

No	Kecamatan	Vit_A	Imunisasi	Balita	Bayi	PHBS	Pengh. Rumah	Gizi
1	Danurejan	93.33	79.0	29.0	71.0	52.0	3.0	42.1
2	Gedong Tengen	100.0	65.0	22.0	188.0	44.0	3.0	48.5
3	Gondokusuman	97.6	89.0	24.0	163.0	33.0	3.0	33.3
4	Gondomanan	99.33	42.0	26.0	189.0	23.0	5.0	23.3
5	Jetis	100.0	69.0	43.0	76.0	51.0	2.0	45.0
6	Kota Gede	90.37	43.0	36.0	53.0	54.0	5.0	54.0
7	Kraton	100.0	83.0	31.0	56.0	52.0	5.0	33.0
8	Matrijeron	100.0	73.0	10.0	128.0	32.0	5.0	22.0
9	Mergangsan	87.8	72.0	776.0	123.0	20.0	2.0	33.4
10	Gampilan	100.0	93.0	67.0	178.0	56.0	3.0	33.9
11	Pakualaman	100.0	55.0	27.0	226.0	34.0	6.7	20.1
12	Tegalrejo	100.0	86.0	70.0	191.0	70.0	4.3	30.6
13	Umbulharjo	93.41	50.0	52.0	72.0	67.0	5.5	24.0
14	Wirobrajan	100.0	85.0	51.0	60.0	65.0	6.0	22.0

C. Metode klasifikasi

Data yang digunakan sebagai *input* adalah data surveilen KLB campak yang berjumlah 648 record yang akan dibagi ke dalam 2 bagian, yaitu 60% untuk data *training* sehingga menghasilkan suatu *rule*. Kemudian sisanya sebanyak 40% digunakan untuk data *testing*. Data ini untuk menguji hasil *rule* yang telah diperoleh dari proses *training*.

Dengan menggunakan *tool* Weka 3.7 pada fitur J48, *output* yang diharapkan adalah nilai akurasi yang baik untuk melakukan klasifikasi terhadap data KLB. Pengujian akan dilakukan dengan menggunakan *10-fold cross validation* untuk pengujian akurasi dari model yang digunakan. Dalam melakukan evaluasi akurasi data digunakan *confussion matrix*.

III.HASIL

Berikut disajikan contoh hasil dari tahapan proses yang telah dilakukan. Bentuk data hasil *preprocessing* adalah seperti pada Tabel II. Pada Tabel II, ditambahkan sebuah atribut yaitu status yang berisi data KLB dan non KLB. Selain itu, atribut nama kecamatan dihilangkan karena digunakan sebagai identitas saja. Gambar 1 merupakan bentuk *decision tree* yang terbentuk dari data *training* yang telah diolah. Selanjutnya dari pohon keputusan pada Gambar 1 dapat dibentuk *rule* seperti pada Tabel III. Aturan keputusan masuk dari If Atecedent then consequent, seperti terlihat pada Tabel II. Untuk keputusan dari aturan yang terbentuk, atesendent terdiri dari nilai atribut dari cabang yang diambil dengan cara mengambil jalur melalui pohon, sementara consequent terdiri dari klasifikasi untuk variabel target yang diberikan node daun.

TABEL II
CONTOH DATA KLB

Cakupan VitaminA	Cakupan imunisasi	Jumlah Balita	Jumlah Bayi	PHBS	Penghuni Rumah	Cakupan Gizi	status
78	43	194	448	41.1	5	34	KLB
87.06	70	276	366	29.4	6	34	KLB
83.26	54	170	300	42.2	5	34	KLB
100	75	52	100	36.9	5	60	Non KLB
100	70	51	107	84.7	5	71	Non KLB
100	57	62	109	31.9	5	60	Non KLB
100	75	167	230	45.1	5	60	Non KLB
100	71	125	175	30.6	2	50.5	Non KLB
79	64	210	330	25.8	6	34	KLB
100	56	288	290	33	4	34	KLB
100	55	110	200	41.3	7	60	KLB
100	41	108	263	35.2	7	34	KLB
100	81	94	116	54.1	3	60	Non KLB
99.06	79	253	319	39.8	7	34	KLB
98.8	57	337	290	42.1	7	24	KLB

Aturan yang terbentuk pada Tabel III merupakan aturan (*rule*) yang terbentuk dari 500 data *training* yang telah ditentukan sebelumnya, kemudian dilakukan pengujian terhadap data uji yang berjumlah 109 data. Dari hasil pengujian 109 data, diperoleh hasil bahwa metode ini berhasil mengklasifikasikan 92 data yang benar. Dengan demikian diperoleh akurasi keberhasilan sebesar 84,40% dan nilai RSE sebesar 0,352 .

TABEL III
ATURAN (RULE) DARI GAMBAR 1

Antesenden	Konsekuen	Support	Confidence
If imunisasi <=64	then KLB	180/15	0.05
If imunisasi >64 and gizi <=50.1 and PHBS <=63	then KLB	76/12	0.05
If imunisasi >64 and gizi <=50.1 and PHBS > 63	then Non KLB	5/1	0.05
If Gizi > 50.1 and imunisasi <=69 and Vit_A <= 96.7	then KLB	9/0	0.05
If Gizi > 50.1 and imunisasi <=69 and Vit_A <= 96.7	then KLB	3/0	0.05
If Gizi > 50.1 and imunisasi <=69 and Vit_A >90.08	then Non KLB	2/0	0.05
If Gizi > 50.1 and imunisasi >69 and balita >235	then KLB	4/0	0.05
If Gizi > 50.1 and imunisasi >69 and balita <=235 and vit_A <=98.41	then KLB	4/0	0.05
If Gizi > 50.1 and imunisasi >69 and balita <=235 and vit_A > 98.41	then Non KLB	3/1	0.05
If Gizi > 50.1 and imunisasi >69 and balita <=235 and vit_A > 98.41	then Non KLB	5/2	0.05
If Imunisasi >73 and Penghunirumah<=5	then KLB	14/20	0.05
If Penghunirumah<=5 and Imunisasi <=69	then NonKLB	19/5	0.05
If Penghunirumah<=5 and Imunisasi <=69			

Untuk melakukan pengujian terhadap metode yang digunakan disini digunakan *10-fold cross validation*. Adapun hasil proses validasi yang dilakukan terlihat pada Tabel IV. Hasil *10-fold cross validation* memiliki rata rata 96,43%. Hasil tersebut diperoleh dari hasil rata-rata dari pengujian disetiap atributnya.

TABEL IV
TEN CROSS FOLD VALIDATION

Fold training	Fold testing	Akurasi
1,2,3,4,5,6	7	100%
1,2,3,4,5,7	6	100%
1,2,3,4,6,7	5	90%
1,2,3,4,6,7	4	95%
1,2,4,5,6,7	3	100%
1,3,4,5,6,7	2	100%
2,3,4,5,6,7	1	90%
Rata-rata		96.43%

TABEL V
NILAI AKURASI PENGUJIAN

Algoritma	Akurasi
JST dan CART	98.438%
Decision Tree C. 4.5	96.43%

IV. PEMBAHASAN

Pada penelitian sebelumnya dengan menggunakan data yang sama menggunakan *classification and regression tree* (CART) dan jaringan syaraf tiruan (JST) telah diperoleh hasil akurasi sebesar 88,52%. Hasil ini lebih baik dibandingkan dengan dengan hasil klasifikasi yang dilakukan dengan menggunakan *Decision Tree* C4.5. Peramalan yang dilakukan dengan menggunakan *JST backpropagation* dan *decision tree* 4.5, hasil yang ditunjukkan untuk melakukan pengklasifikasian CART akurasinya lebih tinggi jika dibandingkan dengan klasifikasi yang diperoleh dengan menggunakan *decision tree* C4.5. Pengujian yang dilakukan dengan menggunakan *10-fold cross validation* terhadap berbagai metode yang digunakan ditunjukkan pada Tabel V. Hasil akurasi menunjukkan bahwa gabungan JST dan CART mempunyai akurasi yang lebih baik dari *decision tree* C4.5.

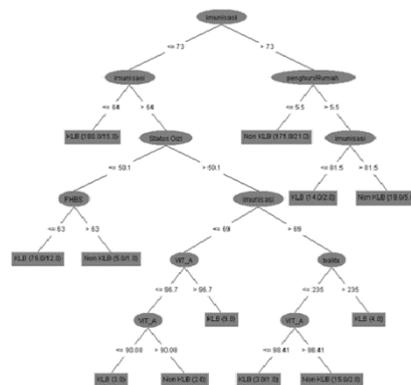
V. SIMPULAN

Adapun simpulan yang bisa diperoleh dari penelitian ini adalah:

- 1) Metode *decision tree* C4.5 dapat digunakan untuk melakukan klasifikasi daerah KLB.
- 2) Hasil akurasi yang diperoleh dengan metode tersebut adalah 84,40%.

REFERENSI

[1] Pedoman Penyelenggaraan Sistem Kewaspadaan Dini Kejadian Luar Biasa (KLB), Peraturan Menteri Kesehatan Republik Indonesia Nomor 949/Menkes/Sk/Vt/2004, Jakarta, 2004.
 [2] T. Larose, *Discovering Knowledge in Data Mining*, Wiley Interscience, Canada, 2005.
 [3] Ali Mahmood Mohd., Qaseem Mohd. S., Lakshmi Rajamani, and A. Govardhan, "Extracting useful rules through improved decision tree induction using information entropy", *International Journal of Information Sciences and Techniques (IJIST)* vol.3, No.1, January, India, 2013.
 [4] Sulistyowati dan E. Winarko, "Peramalan KLB campak menggunakan gabungan metode JST Backpropagation dan CART", *Indonesian Journal of Computing and Cyberetics System*, vol.8, no.1, pp. 49-58, ISSN: 1978-1520, 2014.
 [5] P. Sharma and R. Bhartiya, "Implementation of decision tree algorithm to analysis the performance", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 1, issue 10, December, India, 2012.
 [6] Jei Wei Han, "A map reduce implementation of C4.5 decision tree algorithm", *International Journal of Database Theory and Application*, vol. 7, no.1 (2014), pp. 49-60, 2014.
 [7] B. Santoso, *Data mining: teori dan teknik pemanfaatan data mining untuk keperluan bisnis*, Graha Ilmu, Yogyakarta, 2012.
 [8] Kusri dan Lutfi, *Algoritma data mining*, Andi, Yogyakarta, 2009.



Gambar 1. Hasil pohon keputusan KLB campak