

Implementasi Algoritma *Decision Tree* (ID3) Untuk Penyakit Campak

Inda Srimenganti¹, Ichsan Taufik², Edi Mulyana³

^{1,2}Teknik Informatika, UIN Sunan Gunung Djati Bandung

³Teknik Elektro, UIN Sunan Gunung Djati Bandung

Jl. A.H. Nasution No. 105, 022-7800525

indasrimenganti@student.uinsgd.ac.id¹, ichsan@uinsgd.ac.id², edim@uinsgd.ac.id³

Abstract – Community Health Centre (*Puskesmas*) is a place for the outpatients. Often, a Community Health Centre does not provide a complete set of medical service. The medical workers –especially doctors often find it challenging to determine a patient's disease. One of the challenges is to diagnose measles. Iterative Dichotomizer (ID3) is one of the algorithms to help diagnose measles by obtaining data from several attributes. The attributes of the algorithms may vary depends to patients' need. The algorithm works by calculating the entropy of every attribute and the Information Gain (IG) of every attribute to construct a decision tree. The decision tree helps doctors to diagnose measles. There are 118 data at this research, most of the respondents were diagnosed with measles. The accuracy test of the ID3 algorithm-based application showed the number of 89.83%, categorized as excellent. This concludes that the ID3 algorithm is effective to diagnose measles.

Keywords: Measles, Decision Tree (ID3)

1. Pendahuluan

Penyakit campak adalah penyakit yang ditandai dengan gejala awal demam, batuk, pilek, dan konjungtivitas, yang kemudian diikuti dengan bercak kemerahan pada kulit (*rash*). Campak disebabkan oleh virus campak (*paramyxovirus*) yang disebarkan melalui udara. Virus campak berasal dari family *paramyxovirus*, genus morbill virus. Virus ini adalah virus RNA yang dikenal hanya mempunyai satu antigen. Struktur virus mirip dengan virus penyebab parotitis epidemik dan para influenza. Virus campak dapat bertahan selama beberapa hari pada temperature 0° C dan selama 15 minggu pada sediaan beku. Diluar tubuh manusia virus ini mudah mati pada suhu kamar sekalipun, virus ini akan kehilangan infektifitasnya sekitar 60% selama 3-5 hari. Setiap kasus campak memerlukan variable data yang banyak beberapa data yang harus dicatat diantaranya no epid, nama pasien, nama orang tua, alamat tempat tinggal, imunisasi campak didapatkan sebelumnya, kapan mulai panas kapan mulai timbul *rash*, dan sebagainya. Kondisi seperti ini juga menambah lamanya waktu dalam *entry data* sistem *surveilans* campak berbasis kasus [1][2]

Pusat kesehatan masyarakat adalah sebuah tempat pengobatan rawat jalan, yaitu sebuah instansi yang memiliki banyak data pasien dan alat yang berkecukupan, sehingga petugas atau dokter di puskesmas kesulitan untuk menentukan penyakit Pasien. Kesulitan yang dirasakan oleh pihak balai pengobatan adalah dalam memprediksi penyakit campak seorang pasien menderita penyakit campak atau tidak. Pengembangan sistem *surveilans* campak dimaksudkan untuk melengkapi kekurangan yang ada baik pada komponen input, proses, output sehingga sistem tersebut yang dikembangkan dapat menghasilkan informasi yang lebih lengkap dan berkualitas agar keputusan yang diambil lebih tepat, kegiatan *surveilans* epidemiologi lebih efektif dan efisien sehingga mampu memutus mata rantai penularan dan bisa mengeradikasi sehingga tidak ada lagi transmisi virus campak.

Metode yang digunakan dalam merancang aplikasi prediksi penyakit campak tersebut adalah metode *decision Tree* (ID3), yaitu pohon keputusan yang memperlihatkan faktor-faktor

kemungkinan atau probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi akhir yang akan didapatkan apabila kita mengambil alternatif keputusan tersebut.

2. Metodologi Riset

A. CAMPAK

Penyakit Campak dikenal juga dengan istilah *morbili* dalam bahasa latin dan *measles* dalam bahasa Inggris. Campak pada masa lalu dianggap sebagai suatu hal yang harus dialami oleh setiap anak, mereka beranggapan bahwa penyakit Campak dapat sembuh sendiri bila ruam sudah keluar, sehingga anak yang sakit Campak tidak perlu diobati. Tanda dan gejala yang menjadi ciri ciri campak meliputi:

- Demam
- Batuk kering
- Hidung Ingusan (pilek)
- Sakit tenggorokan
- Mata merah meradang (konjungtivitis)
- Peka terhadap cahaya
- Diare
- Koplik's spot (Bintik-bintik kecil berwarna putih dengan warna putih kebiruan di tengahnya, ditemukan pada lapisan dalam pipi).
- Ruam kulit berwarna merah kecil-kecil rapat dan merata, hampir seluruh tubuh[3].

B. Decision Tree

Menurut Santosa [4], decision tree sesuai digunakan untuk kasus-kasus dimana outputnya bernilai diskret. Walaupun terdapat banyak variasi model *decision tree* dengan tingkat kemampuan dengan syarat yang berbeda, pada umumnya beberapa ciri kasus berikut cocok untuk diterapkan pada *decision tree*:

1. Data dinyatakan dengan pasangan atribut dan nilainya. Misalnya atribut satu data adalah temperatur dan nilainya adalah dingin.
2. Label atau output data biasanya bernilai diskret.
3. Untuk membuat *decision tree*, perlu diperhatikan:
 1. Atribut mana yang akan dipilih untuk pemisahan obyek
 2. Urutan atribut mana yang akan dipilih terlebih dahulu.
 3. Struktur *tree*
 4. Kriteria pemberhentian[5].

C. Algoritma Iterative Dichotomiser 3 (ID3)

Iterative Dichotomiser 3 (ID3) adalah algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh (greedy) pada semua kemungkinan pohon keputusan. Salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun decision tree (pohon keputusan) secara top-down (dari atas ke bawah) [6].

D. Entropy

Menurut Rokach dan Maimoon (2008), information gain atau biasa disebut gain info adalah kriteria pemisahan yang menggunakan pengukuran entropy. Untuk mendapatkan information gain dari suatu atribut dibutuhkan entropy keseluruhan kelas atau Entropy(S) Menurut Han et al. (2011), secara matematis entropy dirumuskan sebagai berikut[7]

$$Entropi (S) = \sum_{j=1}^k - p_j \log_2 p_j$$

formula entropi

Persamaan 1 Entropy

dengan, S adalah himpunan (dataset) kasus

- k adalah banyaknya partisi S
- p_j adalah probabilitas yang didapatkan dari Sum(Y_a) dibagi Total Kasus[8].

E. *Information Gain*

Setelah mendapatkan nilai entropy, maka dapat diukur efektivitas suatu atribut dalam mengklasifikasikan data yang disebut sebagai information gain. Secara matematis, *information gain* dari suatu atribut A, dituliskan sebagai berikut

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i)$$

information gain

Persamaan 2 Information Gain

dimana :

S = ruang (data) *sample* yang digunakan untuk *training*.

A = atribut.

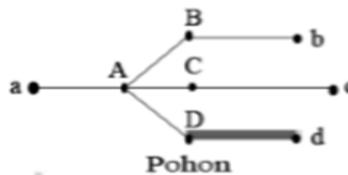
|S_i| = jumlah *sample* untuk nilai V.

|S| = jumlah seluruh *sample* data.

Entropi(S_i) = entropy untuk *sample-sample* yang memiliki nilai i[9]

F. Pohon Keputusan

Pohon (*tree*) merupakan graf dimana dua simpul memiliki paling banyak satu lintasan yang menghubungkannya. Pohon seringkali memiliki akar. karena setiap simpul pada pohon hanya memiliki satu lintasan akses dari setiap simpul lainnya, maka tidak mungkin bagi sebuah lintasan untuk membentuk simpul (*loop*) atau siklus (*cycle*) yang secara berkesinambungan melalui serangkaian simpul[10]

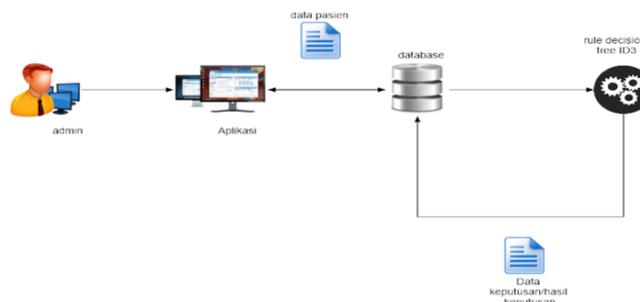


Gambar 1 pohon keputusan

3. Analisis dan Pembahasan

A. ARSITEKTUR SISTEM

Arsitektur sistem pada aplikasi Prediksi penyakit campak dapat dilihat pada Gambar 2 berikut.



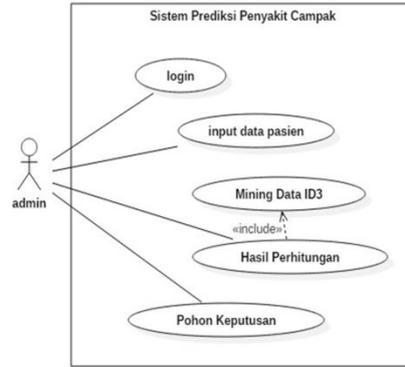
Gambar 2 Arsitektur Sistem

Pada Gambar 2 Arsitektur sistem merupakan rancangan sistem secara umum (garis besar) yang mendefinisikan komponen-komponen yang lebih terstruktur yang dibuat

menggunakan diagram atau *flowchart*. Dalam gambar tersebut menampilkan bagaimana sistem berjalan mulai dari user yang menjalankan aplikasi yang kemudian aplikasi tersebut memunculkan fitur-fitur yang ada di dalamnya dengan mengakses data dari penyimpanan data.

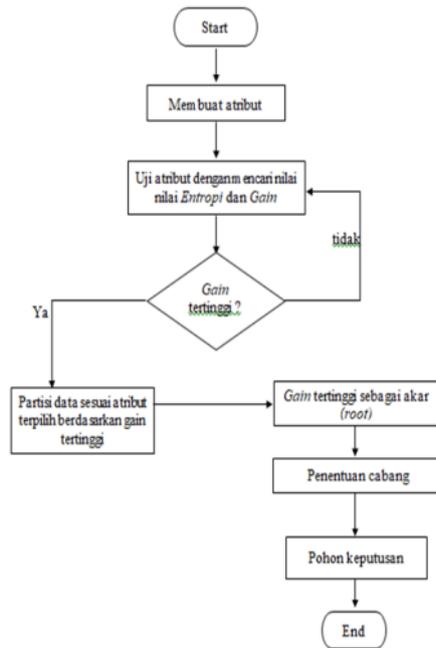
B. USE CASE DIAGRAM

Use Case Diagram merupakan pemodelan untuk menggambarkan interaksi yang dilakukan aktor dengan sistem yang akan dibuat. *Use Case Diagram* pada aplikasi kombinasi menu ini dapat dilihat pada gambar



Gambar 3 Use Case Diagram

C. ANALISIS KINERJA MANUAL ALGORITMA DECISION TREE (ID3)



Gambar 4 Activity Diagram

Berdasarkan Flowchart Gambar 4 dapat diketahui alur algoritma ID3 yang digunakan pada persiapan awal ditentukan atribut yang digunakan kemudian melakukan uji atribut dengan mencari nilai Gain tertinggi berdasarkan perhitungan Entropi dari masing-masing atribut. Apabila ditemukan gain tertinggi maka gain tersebut akan menjadi akar (root) awal, selanjutnya dilakukan penentuan cabang dengan cara yang sama dengan melihat gain tertinggi dari tiap hasil partisi.

A. Tahapan-Tahapan Algoritma Decision Tree (ID3)

1. Menentukan Detail Atribut pada Setiap Atribut

Algoritma ID3 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Dasar algoritma ID3 adalah pembentukan pohon keputusan. Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya. Berikut adalah uraian langkah-langkah algoritma ID3 untuk menyelesaikan menghasilkan penentuan keputusan

2. Membuat Kasus Baru

Kasus baru adalah tahap awal untuk menyelesaikan suatu masalah, dengan memasukan inputan dan kriteria. Dalam kolom akhir berisi perkiraan hasil yang akan diperoleh nanti, keputusan Ya (Positif) dan Tidak (Negatif).

Tabel 1 Activity Diagram

Nama	Umur	Jk	panas	rash	Sesak nafas	Pilek	Igm
Kabir	Muda	L	Tinggi	Ya	Pernah	Sakit	Positif
Bagas	Tua	L	Sedang	Tidak	Tidak Pernah	Sakit	Negatif
Siti	Muda	P	Tinggi	Ya	Pernah	Tidak Sakit	Positif
Rani	Tua	P	Tinggi	Ya	Pernah	Tidak Sakit	Positif
Edi	Tua	L	Sedang	Tidak	Pernah	Sakit	Negatif

1. Pembuatan node

Pembuatan tabel ini untuk menentukan nilai terbesar dari setiap kelas dan dijadikan akar dalam setiap keputusan, dengan cara menghitung jumlah kasus sesuai dengan jumlah kasus yang sudah ditentukan. Kemudian setelah itu, menghitung dan mencari nilai *Entropi* setiap atribut dan nilai *Gain* untuk setiap kelas. Sehingga dapat ditemukan nilai terbesar dan menjadi akar (*root*) pohon. Dapat dilihat pada Tabel 2

Tabel 2 Perhitungan Entropy Status Igm Campak

<i>Hasil</i>	<i>Jumlah Kasus</i>	<i>Pj*log2 Pj</i>
<i>Igm Positif</i>	86	0,33
<i>Igm Negatif</i>	32	0,50
	<i>Entropy (Total)</i>	0,83

$$\text{Entropi (Total)} = (-86/118 * \log_2(86/118)) + (-(32/118) * \log_2(32/118)) = 0,83$$

2. Menentukan Gain

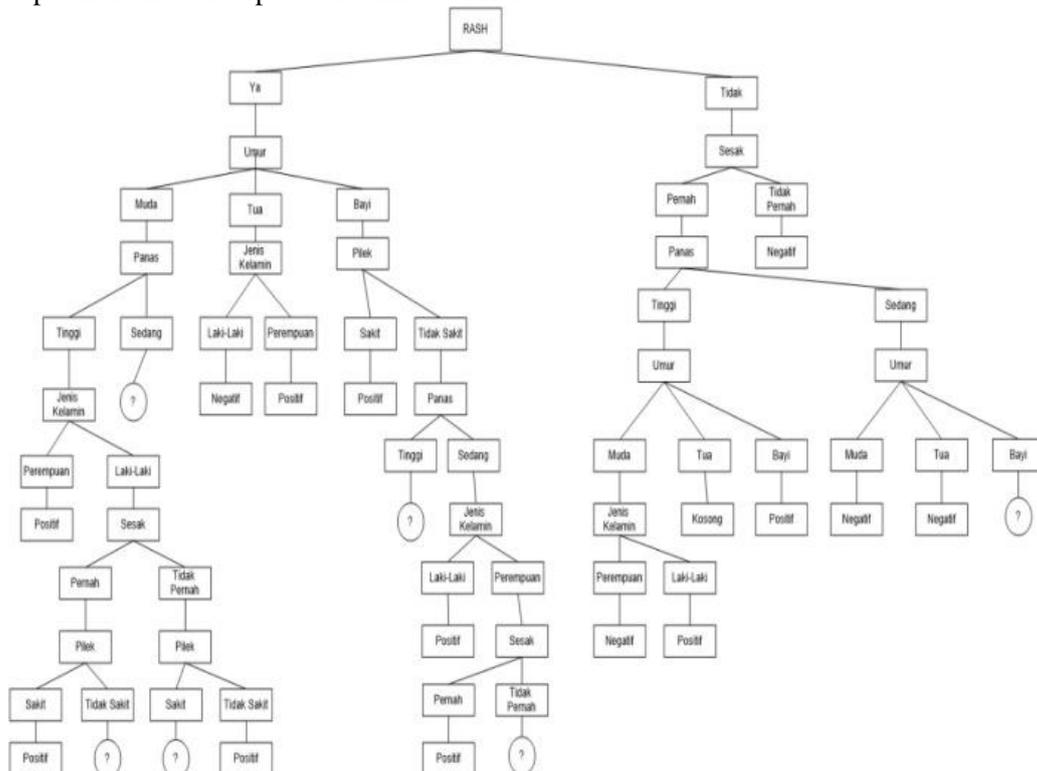
$$\text{Gain (Total Umur)} = 0,8432 - \left(\left(\frac{46}{118} \times 0,950 \right) + \left(\frac{6}{118} \times 0,918 \right) + \left(\frac{66}{118} \times 0,65 \right) \right) = 0,63$$

Tabel 3 Perhitungan Gain

Node	Atribut	value	Jumlah Kasus	Campak	Tidak Campak	Entropy	Gain
1	Umur						0.063
		Muda	46	29	17	0.9503	
		Tua	6	2	4	0.918	
		Bayi	66	55	11	0.65	
	JK						0.0064
		L	61	42	19	0.894	
		P	57	44	13	0.774	
	Panas						0.0594
		Tinggi	68	42	11	0.638	
		Sedang	50	44	21	0.981	
	Rash						0.255
		Ya	93	81	12	0.554	
		Tidak	25	5	20	0.721	
	Sesak						
		Pernah	53	43	10	0.698	0.022
		Tidak Pernah	65	43	22	0.923	
	Pilek						0.009
		Sakit	43	34	9	0.740	
		Tidak Sakit	75	52	23	0.889	

3. Membuat pohon keputusan

Pada Gambar 5 Dapat dilihat dan diuraikan pohon keputusan dengan hasil nilai Gain tertinggi yang akan dijadikan akar (root) sehingga dapat membuat pohon keputusan yang dapat menemukan keputusan akhir sebuah kasus.



Gambar 5 pohon keputusan

Pohon Keputusan dengan hasil nilai *Gain* tertinggi merupakan **Rash**, dengan jumlah nilai *gain* **0.255**.

5. Implementasi dan Pengujian

A. Implementasi

Setelah dilakukan perancangan flowchart, data flow diagram dan sketsa aplikasi, dilakukanlah pembangunan aplikasi dan implementasi *coding*. Pada Gambar 6 merupakan tampilan proses perhitungan aplikasi prediksi penyakit campak

ITERASI	ATT/INF GAIN MAX	ATRIBUT	NILAI ATRIBUT	TOTAL KASUS	JUMLAH CAMPAK	JUMLAH TIDAK CAMPAK	ENTROPY	GAIN
1	umur	Total	Total	34	17	17	1	
2	umur	umur	muda	15	4	11	0.8368	0.232
3	umur	umur	tua	6	2	4	0.9183	0.232
4	umur	umur	bayi	13	11	2	0.6194	0.232
5	umur	jk	L	16	8	11	0.9819	0.0229
6	umur	jk	P	15	9	6	0.971	0.0229
7	umur	panas	tinggi	21	14	7	0.9183	0.1348
8	umur	panas	sedang	13	3	10	0.7793	0.1348
9	umur	rash	ya	25	17	11	0.9666	0.204
10	umur	rash	tidak	6	0	6	0	0.204
11	umur	sesak	pernah	20	13	7	0.9341	0.0951
12	umur	sesak	tidak pernah	14	4	10	0.8831	0.0951
13	umur	pilek	sakit	16	6	10	0.9544	0.0405

Gambar 6 Proses perhitungan

Antarmuka perhitungan Id3 menampilkan hasil perhitungan sesuai dengan data campak.

No	Nama	Umur	Jenis Kelamin	Panas	Rash	Sesak Napas	Pilek	Prediksi	Aksi
1	Tes 2	Tua	L	Sedang	Tidak	Tidak Pernah	Tidak Sakit	negatif	Hapus Edit
2	INDA SRIMENGANTI	Bayi	L	Tinggi	Ya	Pernah	Sakit	positif	Hapus Edit
3	jnsdknd	Tua	L	Sedang	Tidak	Tidak Pernah	Tidak Sakit	negatif	Hapus Edit
4	sift	Muda	P	Tinggi	Ya	Pernah	Tidak Sakit	positif	Hapus Edit
5	rani	Muda	P	Tinggi	Tidak	Pernah	Tidak Sakit	negatif	Hapus Edit
6	staf	Bayi	L	Sedang	Ya	Tidak Pernah	Sakit	positif	Hapus Edit

Gambar 7 Tampilan hasil testing

B. Pengujian

Pada tahap uji coba ini, peneliti mendapatkan seratus data dari dinas kesehatan provinsi jawa barat berdasarkan atribut yang dibutuhkan. Peneliti menggunakan seratus data campak. Pada pemilihan data *training*, peneliti memilih data dari bulan januari- maret dengan data 118 data yang mengalami penyakit campak dan untuk data testing 118 data

Tabel 4 Perhitungan akurasi

Data Training	118
Data Testing	118
Jumlah tidak sama datanya	12

$$\begin{aligned} \text{Akurasi} &= \frac{\text{jumlah data} - \text{selisih kebenaran data}}{\text{jumlah data}} \times 100\% \\ &= 89,83 \%. \end{aligned}$$

5. Kesimpulan

- a. Sistem yang dibangun mampu memprediksi penyakit campak berdasarkan penyakit yang diderita oleh pasien (Positif dan Negatif).
- b. Hasil pengujian akurasi aplikasi ini dari 118 data dengan menggunakan Algoritma ID3 yaitu 89,83%, dari hasil ini bahwa algoritma ID3 mampu membantu dalam memprediksi penyakit campak.

Daftar Pustaka

- [1] A. E. Pramadhani and T. Setiadi, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (Infeksi Saluran Pernapasan Akut) DENGAN ALGORITMA DECISION TREE (ID3)," *J. Sarj. Tek. Inform. e-ISSN 2338-5197*, vol. 2, no. 1, pp. 831–839, 2014.
- [2] C. pengertian Penyakit, *No Title*, Yudoyono. 2005.
- [3] L. Fakta and P. Menular, "Campak," pp. 1–3, 2008.
- [4] Santosa, B. Teknik Pemanfaatan Data untuk Keperluan Bisnis. Graha Ilmu: Yogyakarta. 2007
- [5] Wahyudin, "Metode Iterative Dichotomizer 3 (ID3) untuk Penyeleksian Penerimaan Mahasiswa Baru," *J. Pendidik. Teknol. Inf. Dan Komun.*, vol. 2, no. 2, pp. 5–15, 2009.
- [6] David, Mcg. Tutorial: The ID3 Decision Tree Algorithm. Monash University Faculty of Information Technology. 2004
- [7] D. Setiawati, I. Taufik, Jumadi, and W. Z. Budiawan, "Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile," *J. Online Inform.*, vol. 1, no. 1, pp. 24–27, 2016.
- [8] A. K. Nugroho and D. Iskandar, "Algoritma Iterative Dichotomizer 3 (ID3) Pengambilan Keputusan Decision Making Using Iterative Dichotomizer 3 Algorithm," vol. 3, pp. 43–47.
- [9] A. E. Tyasti, D. Ispriyanti, and A. Hoyyi, "1 , 2 , 3 1," *J. Gaussian*, vol. 4, no. 2, pp. 237–246, 2015.
- [10] D. Lavanya and K. U. Rani, "Performance Evaluation of Decision Tree Classifiers on Medical Datasets," *Int. J. Comput. Appl.*, vol. 26, no. 4, pp. 1–4, 2011.