

Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 (Studi Kasus : Rumah Sakit Umum Daerah (RSUD) Hasanuddin Damrah Manna)

Devi Sartika¹, Yupianti²

^{1,2}Informatika, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
¹Devisartika@unived.ac.id, ²Yupiantiprana@gmail.com

ABSTRAK

Tiroid adalah merupakan salah satu bagian tubuh yang sangat penting bagi manusia, tiroid berbentuk kelenjar dan letaknya di bawah jakun pada leher. Penyakit gondok disebabkan oleh gangguan pada kelenjar tiroid. Berdasarkan data histori penderita yang disebabkan oleh kelenjar tiroid dapat dibuat rekomendasi prediksi penyakit tiroid yang membantu tenaga kesehatan. Klasifikasi merupakan salah satu teknik dari data mining yang dapat digunakan untuk membuat prediksi. Klasifikasi dapat dilakukan dengan decision tree salah satunya dengan algoritma C4.5. Penelitian ini bertujuan membuat klasifikasi data penyakit yang disebabkan oleh gejala kelenjar tiroid dan menerapkannya dalam pembangunan sistem. Implementasi sistem menggunakan bahasa pemrograman Visual Basic 2010 dan metode penelitian yang digunakan pada penelitian ini adalah metode waterfall. metode waterfall mampu melakukan analisa kebutuhan yang digunakan untuk mengetahui dari kelemahan sistem yang lama, kemudian membuat desain dari rancangan tersebut dan dilanjutkan dengan pembuatan rancangan sistem baru. Hasil dari penelitian adalah menghasilkan rule yang dapat membantu tenaga kesehatan di RSUD Hasanuddin Damrah Manna dalam mendiagnosa penyakit yang disebabkan oleh kelenjar tiroid.

Kata kunci : Data Mining, Decision Tree, C4.5, Kelenjar Tiroid

PENDAHULUAN

Tiroid adalah merupakan salah satu bagian tubuh yang sangat penting bagi manusia, tiroid berbentuk kelenjar dan letaknya di bawah jakun pada leher. Tiroid merupakan kelenjar endokrin terbesar dalam tubuh berbentuk kupu-kupu. Fungsi kelenjar tiroid adalah menghasilkan hormon tiroid yang berguna untuk menjaga metabolisme tubuh. Untuk meningkatkan jumlah oksigen pada sel dan rangsangan jaringan tubuh dalam menghasilkan protein. Penyakit gondok disebabkan oleh gangguan pada kelenjar tiroid. Ada dua jenis gangguan tiroid yang dapat muncul yaitu hipertiroid dan hipotiroid. Hipertiroid adalah kondisi dimana kelenjar tersebut bekerja secara berlebihan, sedangkan hipotiroid adalah kebalikannya.

Beberapa teknik data mining telah diterapkan pada dunia kesehatan seperti pada kasus klasifikasi dan data prediktif. Beberapa teknik data mining yang telah digunakan pada bidang kesehatan seperti mendeteksi penyakit hipertensi pada kehamilan dirumah sakit. Untuk menganalisa data dalam jumlah yang besar yang tersimpan pada suatu database, diperlukan teknik data mining.

Data mining adalah suatu proses untuk memperoleh informasi penting yang sangat berguna dari informasi yang sebelumnya tidak diketahui dan pengetahuan yang diperoleh informasi yang

berharga dan dapat dimengerti dari sebuah data (Sartika & Jumadi, 2019). Untuk bisa mendapatkan informasi dari sekumpulan database yang besar diperlukan proses data mining seperti klasifikasi. Klasifikasi atau taksonomi adalah proses menempatkan suatu objek atau konsep kedalam satu set kategori berdasarkan objek atau konsep yang bersangkutan (Andriani, 2013). Metode klasifikasi digunakan untuk membantu dalam memahami pengelompokan data. Klasifikasi sendiri merupakan cabang dari *discovery data mining*.

Algoritma C4.5 adalah salah satu metode klasifikasi dari data mining yang digunakan untuk membangun pohon keputusan (*decision tree*). Algoritma ini mempunyai *input* berupa *training sample* dan *sample*. *Training Sample* berupa data contoh yang akan digunakan untuk membangun sebuah pohon (*tree*) yang telah diuji kebenarannya. Sedangkan *Sample* merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi.

Berdasarkan hal yang telah dijabarkan di atas maka penulis melakukan penelitian yang dituangkan dalam pembuatan aplikasi Klasifikasi Penderita Penyakit Tiroid.

Adapun batasan masalah pada penelitian ini adalah atribut yang digunakan sebanyak 24 (dua puluh empat) gejala.

Article History:

Received: August, 30th 2019; Accepted: January, 12th 2020
ISSN: 2502-5325 (Online) **Terakreditasi Peringkat 3** oleh Kementerian Riset, Teknologi dan Pendidikan Tinggi (ARJUNA), berdasarkan Keputusan Direktur Jenderal Penguatan Riset dan Pengembangan No: 23/E/KPT/2019 tanggal 8 Agustus 2019

Cite this as:

Sartika, D. Y. Y. (2020). Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 (Studi Kasus : Rumah Sakit Umum Daerah (RSUD) Hasanuddin Damrah Manna). *Rekayasa*, 13(1), 71-76. doi: <https://doi.org/10.21107/rekayasa.v13i1.5912>

© 2020 Author(s)

METODE PENELITIAN

Knowledge Discovery In Database (KDD)

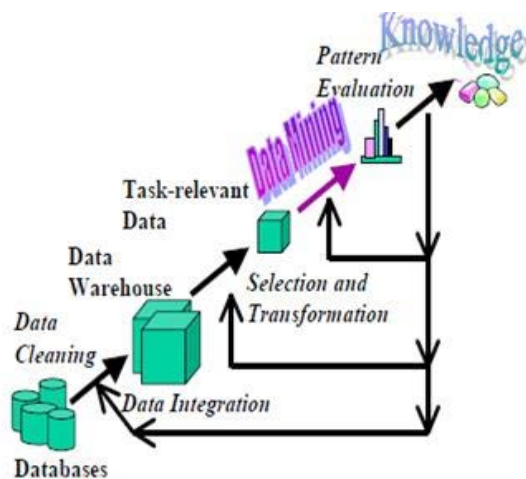
Knowledge Discovery in Database (KDD) merupakan metode untuk memperoleh pengetahuan dari database yang ada. Dalam database terdapat tabel - tabel yang saling berhubungan / berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (knowledge base) untuk keperluan pengambilan keputusan (Yuli, 2017).

Proses KDD secara garis besar dapat dijelaskan sebagai berikut :

1. Data Selection
2. Pre-processing / cleaning
3. Transformation
4. Data Mining
5. Interpretation / Evaluation

Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* (Muzakir & Wulandari, 2016)



Gambar 1. Proses Data Mining

Klasifikasi

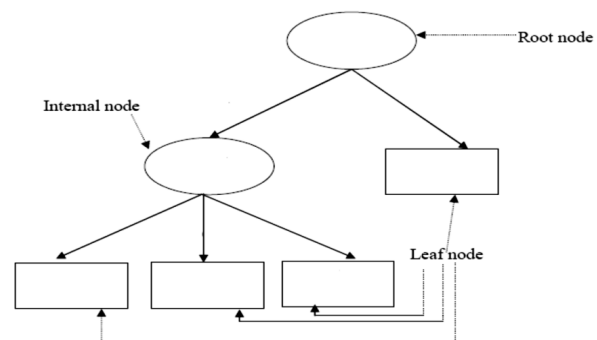
Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasi spesies berdasarkan karakteristik fisik. Selanjutnya dia dikenal sebagai bapak klasifikasi.

Dalam klasifikasi terdapat target variabel kategori. Metode-metode atau model-model yang telah dikembangkan oleh periset untuk menyelesaikan kasus klasifikasi antara lain :

1. Pohon Keputusan
2. Pengklasifikasian *bayes*
3. Jaringan saraf tiruan
4. Analisa statistik
5. Algoritma genetik
6. Rought set
7. Pengklasifikasi *K-Nearest Neighbour*
8. Metode berbasis aturan
9. Memori *based reasoning*
10. *Support Vector Machine (SVM)*

Decision Tree (Pohon Keputusan)

Sebuah *decision tree* adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Pada *decision tree* setiap simpul daun menandai label kelas. Simpul yang bukan simpul akhir terdiri dari akar dan simpul internal yang terdiri dari kondisi tes atribut pada sebagian *record* yang mempunyai karakteristik yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk segi empat. (Muzakir & Wulandari, 2016). Berikut adalah struktur *decision tree* seperti yang ditunjukkan pada gambar berikut ini



Gambar 2. Struktur Decision Tree

Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma yang digunakan dalam *decision tree*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) yang dikembangkan oleh J.Ross Quinlan (Muzakir & Wulandari, 2016).

Algoritma data mining C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Dasar algoritma adalah pembentukan pohon keputusan (*decision tree*). Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya.

Secara umum, algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut :

1. Hitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu.
2. Pilih atribut sebagai *Node*
3. Buat cabang untuk tiap-tiap anggota dari *Node*
4. Periksa nilai *entropy* dari anggota *node* yang ada bernilai Nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai *entropy* anggota *node* adalah nol, maka proses pun berhenti
5. Jika ada anggota *node* yang memiliki *entropy* lebih besar dari nol, ulangi lagi proses dari awal dengan *node* sebagai syarat sampai semua anggota dari *node* bernilai nol.

Node adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *gain* suatu atribut digunakan rumus sebagai berikut :

$$Gain = (S, A) = Entropy(S) - \sum_{i=1}^n \frac{A_i}{S} * Entropy(A_i)$$

Keterangan :

- S : Kasus
- A : Atribut
- n : Jumlah partisi atribut A
- A_i : Jumlah kasus pada partisi ke-i
- S : Jumlah kasus

Sedangkan untuk menghitung nilai *entropy* dapat dilihat pada persamaan berikut :

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

Keterangan :

- S : Himpunan kasus
- n : Jumlah partisi S
- P_i : Proporsi dari S_i ke S

Proses Klasifikasi dengan C4.5

Pada tahap ini akan dibuat pengetahuan, yang akan dijelaskan pada tabel penyakit yang disebabkan oleh kelenjar tiroid, tabel keputusan, *rule-rule*, dan pohon keputusan, yang akan dijabarkan secara rinci pada tabel 1 dan tabel 2.

Tabel 1 Tabel Penyakit

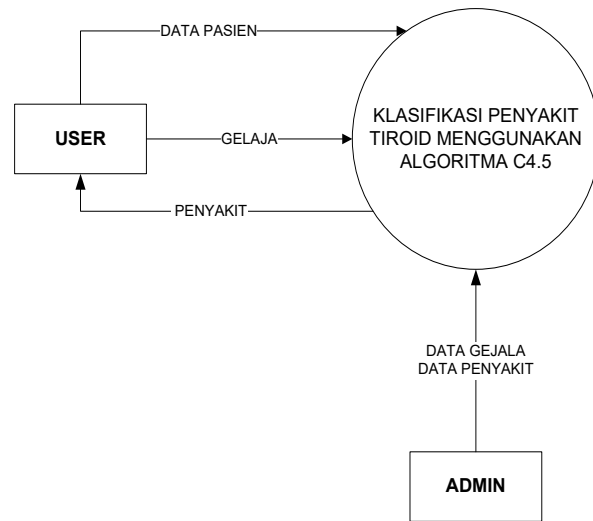
Kode Penyakit	Nama Penyakit
P01	Gondok
P02	Hipotroid
P03	Hipertroid

Dari tabel analisa perhitungan tersebut diatas, akan diperhitungkan nilai entropy dan gain dari

masing-masing gejala. Berikut perhitungan nilai entropy dan nilai gain node pertama:

Rancangan Sistem

Diagram Konteks

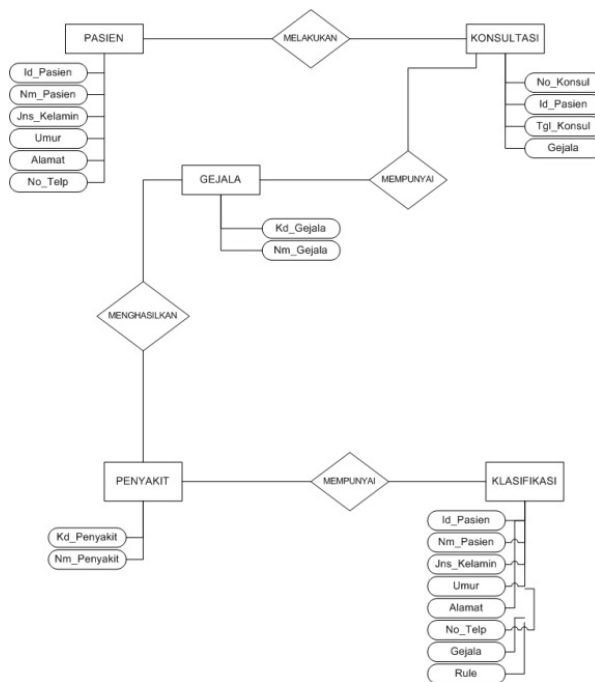


Gambar 3 Diagram Konteks

Table 2 Gejala Penyakit

Kode Gejala	Nama Gejala
G01	Nafsu makan berkurang
G02	Mulut terasa tegang dan nyeri
G03	Sering mual dan muntah
G04	Suhu badan tinggi
G05	Telinga berdengung
G06	Kelelahan
G07	Tidak tahan dingin
G08	Depresi
G09	Rambut dan kulit kering dan kasar
G10	Nafsu makan berkurang
G11	Berat badan naik
G12	Nyeri otot
G13	Wajah membengkak
G14	Suara kasar dan lambat
G15	Sendi kaku
G16	Sembelit (BAB tidak teratur)
G17	Ngantuk yang berlebihan
G18	Konsentrasi menurun
G19	Pembengkakan di pangkal leher
G20	Tidak tahan panas
G21	Gemetaran
G22	Denyut jantung yang cepat
G23	Konsentrasi yang berkurang
G24	Aliran menstrual yang tidak teratur dan sedikit

Entity Relationship Diagram (ERD)



Gambar 4 ERD

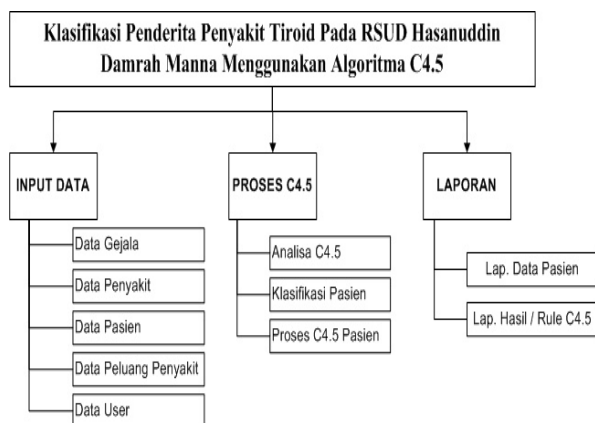
Kode Gejala	Nama Gejala
G-15	Sendi kaku
G-16	Sembelit (BAB tidak teratur)
G-17	Ngantuk yang berlebihan
G-18	Konsentrasi menurun
G-19	Pembengkakan di pangkal leher
G-20	Tidak tahan panas
G-21	Gemetaran
G-22	Denyut jantung yang cepat
G-23	Konsentrasi yang berkurang
G-24	Aliran menstruasi yang tidak teratur dan sedikit

Gambar 6 Form Input Data Gejala

Form Input Data Penyakit

Sub menu *input* data penyakit merupakan *form* untuk melakukan entri data penyakit. Input data penyakit berfungsi sebagai *atap muka (interface)* guna melakukan penambahan data, perbaikan data, pembatalan data, dan penghapusan data penyakit

Struktur Menu



Gambar 5 Struktur Menu

Kode Penyakit	Nama Penyakit
P-01	GONDOK
P-02	HIPOTIROID
P-03	HIPERTIROID

Gambar 7 Form Input Data Penyakit

HASIL DAN PEMBAHASAN

Implementasi aplikasi klasifikasi penyakit tiroid untuk pasien pada RSUD Hasanuddin Damrah Manna berdasarkan gejala yang dialami penderita mempunyai beberapa *interface* (antar muka) antara lain adalah sebagai berikut :

Form Input Data Gejala

Sub menu *input* data gejala merupakan *form* untuk melakukan entri data gejala yang terdapat pada aplikasi ini :

Form Input Data Pasien

Sub menu *input* data pasien merupakan *form* untuk melakukan entri data pasien. Input data pasien berfungsi sebagai *atap muka (interface)* guna melakukan penambahan data, perbaikan data, pembatalan data, dan penghapusan data pasien

Kode Pasien	Nama Pasien	Jenis Kelamin	Umur Pasien	Alamat
S-02	Boby	Laki-Laki	22	Kota I
S-03	Nurmah	Perempuan	35	Dusa
S-04	Nafita Dewi	Perempuan	23	Teluk
S-05	Marhayati	Perempuan	31	Gunur
S-06	Kamadi	Laki-Laki	41	Air Ni
S-07	Jurman Hayadi	Laki-Laki	36	Ulu M
S-08	Rismadi	Laki-Laki	28	Batu I
S-09	Eni Puspta	Perempuan	25	Padar
S-10	Henri Masrah	Perempuan	35	Palak

Gambar 8 Form Input Data Pasien

Form Input Data Peluang Penyakit

Form *input* data peluang penyakit merupakan *interface* yang berfungsi untuk mengetahui hubungan gejala dengan penyakit

Kode Peluang	Kode Penyakit	Nama Penyakit
E-01	P-01	GONDOK
E-02	P-01	GONDOK
E-03	P-02	HIPERTIROID
E-04	P-03	HIPERTIROID
E-05	P-03	HIPERTIROID
E-06	P-01	GONDOK
E-07	P-01	GONDOK
E-08	P-03	HIPERTIROID

Gambar 9 Form Input Data Peluang Penyakit

Pada *interface input* data peluang penyakit terdapat tombol "Detail Gejala". Tombol "Detail Gejala" ini berfungsi untuk menampilkan *interface* detail data penyakit terhadap gejala

Gambar 10. Form Data Detail Penyakit Terhadap Gejala

Form Proses Klasifikasi Penyakit Pasien

Menu proses klasifikasi pasien merupakan *interface* yang berfungsi untuk mengetahui penyakit yang diderita oleh pasien berdasarkan dengan gejala yang dipilih atau dialami oleh pasien

Gambar 11 Form Proses Klasifikasi Penyakit Pasien

Form Proses C4.5 Pasien

Form proses C4.5 pasien merupakan *interface* untuk memproses data gejala yang telah dipilih pada *interface* klasifikasi pasien agar mendapatkan hasil penyakit apa yang diderita oleh pasien.

Gambar 12 Form Proses C4.5 Pasien

Pada gambar *interface* diatas dapat dilihat bahwa semua data pasien yang telah melakukan diagnosa penyakit berhasil direkam oleh sistem yang kemudian akan diproses dengan menggunakan algoritma C4.5 dengan cara meng-klik tombol "Proses" maka akan tampil hasil seperti terlihat seperti gambar berikut ini :

