

KLASIFIKASI PENDERITA PENYAKIT HIPERTIROID MENGUNAKAN ALGORITMA DECISION TREE

Deny Hermansyah ¹⁾, Sinarring Azi Laga ²⁾, Ellen Theresia Sihotang ³⁾, Alya Rizky Natasya ⁴⁾

^{1,2,4)}"Informatika" Universitas Hayam Wuruk Perbanas

³⁾"Manajemen" Universitas Hayam Wuruk Perbanas

Email : deny.hermansyah@hayamwuruk.ac.id ¹⁾, sinarring.laga@perbanas.ac.id ²⁾,
ellen@perbanas.ac.id ³⁾, 202102011013@mhs.hayamwuruk.ac.id ⁴⁾

Abstract

Hyperthyroidism is a condition characterized by the excessive production of thyroid hormones by the thyroid gland, leading to an increased metabolic rate. Diagnosing this condition can be challenging due to its diverse range of symptoms. This variability can be addressed through information technology, particularly data mining, by utilizing disease prediction based on available data. Based on this premise, the aim of this study is to classify patients with hyperthyroidism using the Decision Tree algorithm. The data used in this study consist of primary data from 100 patients with 12 symptom attributes. The results indicate that the resulting classification model demonstrates a high level of accuracy. The precision value is 84.81% for female and 85.71% for male. The recall value is 95.71% for females and 60% for males. These findings suggest that the Decision Tree algorithm is sufficiently effective for classifying hyperthyroidism patients based on observed symptoms.

Keywords: *Data Mining, Decision Tree, Disease Classification, Hyperthyroidism*

Abstrak

*Hipertiroidisme merupakan kondisi dengan produksi hormon tiroid berlebihan oleh kelenjar tiroid, sehingga meningkatkan metabolisme tubuh. Diagnosis penyakit tersebut cukup menantang karena memiliki gejala yang beragam. Keragaman tersebut dapat diatasi melalui teknologi informasi data mining dengan memanfaatkan prediksi penyakit berdasarkan data yang tersedia. Berdasarkan pada hal tersebut maka tujuan dari penelitian ini adalah mengelompokkan pasien penderita *hipertiroidisme* berdasarkan algoritma *Decision Tree*. Data yang digunakan adalah data primer sejumlah 100 pasien dengan 12 atribut gejala. Hasil penelitian mengungkapkan bahwa model klasifikasi yang terbentuk memiliki tingkat akurasi tinggi. Nilai presisinya sebesar 84,81% untuk perempuan dan 85,71% untuk laki-laki. Nilai *recall* sebesar 95,71% untuk perempuan dan 60% untuk laki-laki. Berdasarkan pada hasil tersebut maka algoritma *Decision Tree* cukup efektif untuk digunakan pada proses klasifikasi pasien hipertiroidisme berdasarkan gejala yang diamati.*

Kata kunci: *Data Mining, Decision Tree, Hipertiroidisme, Klasifikasi Penyakit*

1. Pendahuluan

Penyakit tiroid merupakan suatu kondisi yang terjadi pada seseorang akibat adanya gangguan pada kelenjar tiroid, sehingga menyebabkan adanya perubahan baik dalam struktural maupun fungsional. Tiroid berfungsi untuk menghasilkan hormon yang berguna untuk menjaga metabolisme tubuh, serta meningkatkan jumlah oksigen pada sel dan rangsangan jaringan tubuh dalam menghasilkan protein (Agustiani et al., 2020). Pada kenyataannya diagnosis terhadap penyakit tersebut sulit dilakukan, karena gejalanya bermacam-macam, sangat bervariasi, tergantung pada naik dan turunnya hormon tiroid. Ketika tiroid memproduksi hormon berlebih, sel tubuh akan bekerja lebih keras dan metabolisme tubuh menjadi lebih cepat, kondisi ini disebut hipertiroid. Sebaliknya, Ketika tiroid tidak memproduksi hormon yang cukup, sel tubuh akan bekerja lebih lambat, kondisi

ini disebut hipotiroid. Selain pemeriksaan dan penyelidikan tiroid, interpretasi data klinis yang tepat merupakan pelengkap penting dalam diagnosa penyakit tiroid (Ariawan & Widiana, 2023).

Data mining merupakan salah satu metode yang dapat digunakan untuk memprediksi penyakit. Data mining sebagai sekumpulan proses untuk mencari pola atau alur yang selama ini tidak diketahui secara manual dari sebuah kumpulan data yang nantinya akan menghasilkan berupa pengetahuan baru (Laga et al., 2023). Penggunaan klasifikasi merupakan teknik data mining yang mempelajari sekumpulan data sehingga menghasilkan aturan untuk dapat mengklafikasi atau mengenali data baru yang belum pernah dipelajari. Proses pembelajaran memerlukan data input berupa data *training* yang memiliki atribut kelas dan menghasilkan luaran berupa model klasifikasi (Hana, 2020).

Faktor persentase pembagian jumlah data *training* dan data *testing* untuk klasifikasi penderita penyakit hipertiroid dapat mempengaruhi kinerja model klasifikasi. Penggunaan seluruh fitur dalam proses klasifikasi juga dapat mempengaruhi kinerja model klasifikasi, baik meningkatkan atau menurunkan performa klasifikasi dalam hal akurasi dan juga waktu komputasi yang diperlukan, sehingga pendekatan yang dapat dilakukan yaitu dengan memilih persentase data *training* dan data uji yang tepat serta melakukan reduksi dimensi (Azmi et al., 2023).

Berdasarkan pada penjelasan tersebut maka penelitian ini menggunakan atribut sebanyak 12 gejala dengan data *training* maupun data *testing* melalui penerapan algoritma *decision tree*, sehingga tujuannya adalah mengklasifikasi para pasien penderita penyakit hipertiroid melalui algoritma *decision tree* secara akurat.

2. Metode Penelitian

Penelitian ini melibatkan beberapa tahapan yang dimulai dari pengumpulan data awal hingga penerapan algoritma untuk memastikan kecocokan antara hasil akhir dengan berbagai perhitungan. Hasil akhir dari penelitian ini adalah analisis terhadap faktor-faktor yang mempengaruhi jenis penyakit. Tahapan yang dilakukan pada penelitian ini tercantum pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian adalah sebagai berikut:

a. Pengumpulan Data Awal

Dataset yang digunakan dalam penelitian ini merupakan dataset primer yang berisi data penderita penyakit tiroid, terdiri atas 100 data pasien dan 12 atribut. Atribut-atribut tersebut meliputi, usia, jenis kelamin, sesak napas, palpitasi, mudah lelah, senang hawa dingin, keringat berlebihan, gugup, nafsu makan bertambah, nafsu makan berkurang, berat badan naik, dan berat badan turun.

b. Pengolahan Data Awal

Beberapa tahapan pada pengolahan data yaitu:

- 1) Validasi data (*data validation*), digunakan untuk menemukan dan menghilangkan data yang tidak lengkap atau *missing value*, *duplicate*, maupun data yang tidak valid.
- 2) Integrasi dan transformasi data (*data integration and transformation*) dilakukan untuk meningkatkan akurasi dan efisiensi algoritma. Integrasi data adalah proses menggabungkan data dari berbagai sumber yang berbeda menjadi satu dataset yang kohesif sedangkan transformasi data sebagai proses mengubah atau memodifikasi data ke dalam format yang sesuai atau lebih berguna untuk analisis
- 3) Reduksi ukuran data dan diskritisasi (*data size reduction and dicritization*), bertujuan untuk memperoleh dataset dengan jumlah atribut dan *record* yang lebih sedikit namun tetap informatif.

c. Metode Yang Diusulkan

Pada penelitian ini menggunakan algoritma *decision tree* untuk melakukan pengujian model dalam memprediksi penyakit *hipertiroid*. Tahap untuk membuat pohon keputusan secara garis besar yaitu tahap pertama adalah membuat akar pohon, kemudian membagi data berdasarkan atribut yang sesuai untuk membentuk daun. Selanjutnya, dilakukan pemangkasan pohon (*tree pruning*) untuk mengurangi ukuran pohon dan mengurangi tingkat kesalahan pada prediksi kasus baru. Pemangkasan ini dapat dilakukan dengan pendekatan *pre-pruning* atau *post-pruning*. Tahap terakhir adalah menyusun aturan keputusan dari pohon yang terbentuk dengan menelusuri jalur dari akar hingga ke daun.

3. Hasil dan Pembahasan

Deskripsi data pada penelitian ini yang diukur sebagai variabel, tercantum pada Tabel 1.

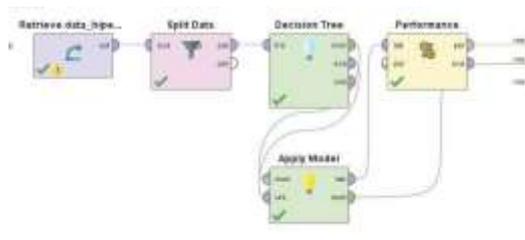
Tabel 1. *Variabel Data Penelitian*

No.	Atribut	Value
1	Umur	1) 20-35 2) 36-45 3)46-55 4) > 55
2	Jenis kelamin	1) Pria 2) Wanita
3	Sesak nafas	1) Ya 2) Tidak
4	Palpitasi	1) Ya 2) Tidak
5	Mudah lelah	1) Ya 2) Tidak
6	Senang hawa dingin	1) Ya 2) Tidak
7	Keringat berlebihan	1) Ya 2) Tidak
8	Gugup	1) Ya 2) Tidak
9	Nafsu makan bertambah	1) Ya 2) Tidak
10	Nafsu makan berkurang	1) Ya 2) Tidak
11	Berat badan naik	1) Ya 2) Tidak
12	Berat badan turun.	1) Ya 2) Tidak

Berdasarkan Tabel 1, terdapat 12 dataset gejala dan satu variabel *class* sebagai penentu klasifikasi. Terdapat 100 data dari keseluruhan jumlah data yang digunakan, data dibagi menjadi dua dengan presentase 70% data *training* dan 30% data *testing*. 70% dari jumlah data 100 didapatkan data *training* dengan jumlah 30 dan 20% dari jumlah data 100 didapatkan data *testing* dengan jumlah 70. Berdasarkan data yang diperoleh maka selanjutnya diproses membuat pohon keputusan

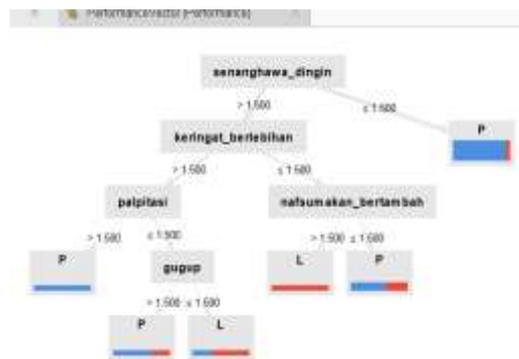
menggunakan algoritma *decision tree* guna menghasilkan alur keputusan.

Tahap awal adalah memilih atribut sebagai akar dengan melakukan perhitungan gain information. Atribut dengan nilai information gain paling tinggi dipakai sebagai akar untuk membangun pohon keputusan. Penerapan dan pengujian dataset dilakukan pada *rapidminer*. Proses klasifikasi penderita *hipertiroidisme* dirancang seperti tercantum pada Gambar 2.



Gambar 2. Proses Klasifikasi

Model pohon keputusan yang diperoleh ditunjukkan pada Gambar 3, dengan pola-pola klasifikasi tertentu sebagaimana terlihat pada gambar berikut.



Gambar 3. Pohon Keputusan

Tahapan lebih lanjut, setelah data diolah melalui *tools rapidminer* diperoleh hasil berupa akurasi *decision tree*. Berikut analisis hasil pengujian dari 70 data *testing* dengan 30 data *training* dengan setiap data memiliki 12 variabel dataset gejala dan satu variabel kelas. Gambar 4, menunjukkan bahwa dengan menggunakan *decision tree*, didasarkan pada 100 data diperoleh 67 orang penderita wanita dan 12 orang penderita laki-laki yang di prediksi mengalami *hipertiroid*.

accuracy 85.00%			
	true P	true L	class precision
pred P	67	12	84.81%
pred L	3	18	85.71%
classical	85.71%	83.33%	

Gambar 4. Hasil Akurasi

Kesimpulan

Penyakit tiroid merupakan kondisi yang disebabkan oleh gangguan pada kelenjar tiroid seseorang. Diagnosis penyakit ini cukup kompleks karena gejalanya sangat beragam dan bervariasi, tergantung

pada perubahan kadar hormon tiroid. Pada penelitian ini, digunakan *dataset* primer yang terdiri dari 100 data pasien dengan 12 atribut, meliputi umur, jenis kelamin, sesak napas, palpitasi, mudah lelah, senang terhadap hawa dingin, keringat berlebihan, gugup, nafsu makan bertambah, nafsu makan berkurang, berat badan naik, dan berat badan turun. Dataset tersebut dimanfaatkan untuk melakukan klasifikasi penderita hipertiroid dengan menerapkan algoritma *decision tree*. Hasil penelitian menunjukkan bahwa klasifikasi menggunakan algoritma *decision tree* menghasilkan akurasi sebesar 85,00%, dengan nilai *precision* sebesar 84,81% untuk pasien perempuan dan 85,71% untuk pasien laki-laki. Nilai *recall* mencapai 95,71% untuk pasien perempuan dan 60,00% untuk pasien laki-laki. Berdasarkan perolehan tersebut maka model pohon keputusan yang dihasilkan membuktikan bahwa algoritma *decision tree* efektif mengklasifikasikan penderita *hipertiroid* berdasarkan atribut-atribut yang tersedia dalam dataset.

Daftar Pustaka

- Agustiani, S., Mustopa, A., Saryoko, A., Gata, W., & Wildah, S. K. (2020). Penerapan Algoritma J48 Untuk Deteksi Penyakit Tiroid. *Paradig.-J. Komput. Dan Inform*, 22(2), 153–160.
- Ariawan, W. E., & Widiana, P. I. M. A. (2023). Sistem Pakar Mendiagnosa Penyakit Tiroid Menggunakan Metode Certainty Factor Berbasis Web. *Jurnal Sutasoma*, 1(2), 104–110.
- Azmi, B. N., Hermawan, A., & Avianto, D. (2023). Analisis Pengaruh komposisi data training dan data testing Pada penggunaan PCA Dan Algoritma decision tree untuk KLASIFIKASI Penderita Penyakit liver. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290.
- Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 32–39.
- Laga, S. A., Hermansyah, D., Rithmaya, C. L., Zainuddin, M., Aji, G. A. I. P., & Mukhlis, I. R. (n.d.). Analyzing Customer Loyalty Levels through Segmentation in Aesthetic Clinics Using K-Means and RFAM. *IJID (International Journal on Informatics for Development)*, 13(2), 473–484.