

DIAGNOSA PENYAKIT DISK HERNIA DAN SPONDYLOLISTHESIS MENGGUNAKAN ALGORITMA C5

Ulfi Saidata Aesy⁽¹⁾, Taufaldisatya Wijatama Diwangkara⁽²⁾, Riyanto Tri Kurniawan⁽³⁾

Program Studi Sistem Informasi, Fakultas Teknik dan Teknologi Informasi,
Universitas Jenderal Achmad Yani Yogyakarta
Jl. Siliwangi Km 0.7, Gamping, Sleman, 55293

e-mail : ulfi@unjaya.ac.id⁽¹⁾, t.wijatama.d@gmail.com⁽²⁾, riyantotri14@gmail.com⁽³⁾

Abstract

The vertebral Column or spine is a sequence of bones from the neck to the tail owned by the vertebrates. The function of the spine is to protect the spinal nerves. The spine may experience malfunction if it is caused by abnormalities and diseases such as Disk Hernia and Spondylolisthesis. Based on the problem then it takes a system that can recognize or identify the disease Disk Hernia and Spondylolisthesis that attack the spine. So that it can be diagnosed with early spinal diseases. In building this system using C 5.0 algorithm. This research uses 310 data from the UCI machine Learning, where there are three classification classes of Normal, Hernia disks, and Spondylolisthesis. The results showed that the C 5.0 algorithm was able to identify with accuracy of 79%. Then the resulting decision tree C 5.0 algorithm is maximized by using AdaBoost algorithm, so the accuracy increases to 83%.

Keywords : *Vertebral Column, C 5.0 algorithm, AdaBoost algorithm*

Vertebral Column atau tulang belakang merupakan sebuah rangkaian tulang dari leher ke ekor yang dimiliki oleh vertebrata. Fungsi dari tulang belakang adalah untuk melindungi syaraf tulang belakang. Tulang belakang dapat mengalami kegagalan fungsi jika disebabkan kelainan dan penyakit seperti Disk Hernia dan Spondylolisthesis. Berdasarkan masalah tersebut maka dibutuhkan suatu sistem yang dapat mengenali atau mengidentifikasi penyakit Disk Hernia dan Spondylolisthesis yang menyerang tulang belakang. Sehingga dapat dilakuakn diagnosa penyakit tulang punggung secara dini. Dalam membangun sistem ini menggunakan algoritma C5.0. Penelitian ini menggunakan 310 data dari UCI Mechine Learning, dimana terdapat tiga kelas klasifikasi yaitu Normal, Disk Hernia, dan Spondylolisthesis. Hasil penelitian menunjukkan bahwa algoritma C5.0 mampu melakukan identifikasi dengan akurasi sebesar 79%. Kemudian pohon keputusan yang dihasilkan Algoritma C5.0 dimaksimalkan dengan menggunakan Algoritma AdaBoost, sehingga akurasi meningkat menjadi 83%.

Kata Kunci : *Vertebral Column, Algoritma C5.0, Algoritma AdaBoost*

1. PENDAHULUAN

Vertebral Column atau tulang belakang merupakan sebuah rangkaian tulang dari leher ke ekor yang dimiliki oleh vertebrata (Pallardy, 2019). Fungsi utama tulang belakang adalah melindungi saraf tulang belakang, memberi bentuk tegak pada tubuh, tempat menempelnya otot, tulang pinggang dan panggul, serta fungsi tambahan pada manusia untuk memindahkan bobot tubuh ketika berjalan maupun berdiri (Pallardy, 2019).

Tulang belakang dapat mengalami kegagalan fungsi jika disebabkan kelainan atau penyakit seperti Disk Hernia dan Spondylolisthesis. Disk hernia merupakan suatu permasalahan yang terjadi pada salah satu bantalan karet (disk) antara tulang-tulang penyusun rangkaian tulang belakang (Mayo Clinic, 2019). Disk pada tulang belakang memiliki bentuk seperti jeli donat dengan bagian tengah yang lembut dan bagian luar yang lebih keras. Disk hernia terjadi ketika bagian tengah (jeli yang lebih lembut) terdorong keluar menembus jeli bagian luar. Hal ini dapat mengakibatkan iritasi pada saraf yang berada didekat lokasi disk hernia, rasa sakit, atau mati rasa pada bagian lengan atau tungkai (Mayo Clinic, 2019).

Berbeda dengan disk hernia yang merupakan suatu penyakit atau kelainan yang diakibatkan oleh bantalan karet antar tulang, spondylolisthesis merupakan suatu kondisi dimana satu atau lebih tulang pada rangkaian tulang belakang bergeser ke depan atau belakang dari posisi aslinya (Tebet, 2014). Penyebab terjadinya kondisi ini masih belum jelas dari segi sejarah medis maupun perkembangan medis.

Data mining adalah kegiatan mengekstrak informasi atau pengetahuan penting dari suatu set data berukuran besar dengan menggunakan teknik tertentu (Santosa & Umam, 2018). Adapun tugas-tugas yang dapat dilakukan dengan data mining adalah klastering, klasifikasi, regresi/estimasi, dan asosiasi. Manfaat yang dapat diperoleh dari menggunakan data mining cukup banyak seperti memprediksi harga saham, mencari segmentasi pelanggan, hingga mengklasifikasi penyakit yang diderita pasien.

Algoritma C5.0 merupakan salah satu metode data mining khusus Teknik pohon keputusan yang merupakan penyempurnaan dari algoritma ID3 dan C4.5. Algoritma C5.0 ini mampu melakukan klasifikasi dengan baik. Selain klasifikasi, algoritma ini mampu melakukan penilaian karyawan (Kastawan, dkk, 2018) dengan menghasilkan akurasi sebesar 96,08%. Kemampuan lain yang dimiliki algoritma C5.0 adalah untuk melakukan identifikasi penyakit (Meng, Huang, Rao, Zhang, & Liu, 2013) mencapai akurasi sebesar 97,98%. Kemudian algoritma C5.0 dikombinasikan dengan *boosting method* (Rafe, Farhoud, & Rasoolzadeh, 2014) akurasi meningkat dari 81% menjadi 93%.

Berdasarkan latar belakang tersebut, penelitian ini menggunakan algoritma C5.0 untuk melakukan identifikasi penyakit tulang belakang yaitu Disk Hernia dan Spondylolisthesis. Klasifikasi penyakit Disk Hernia dan Spondylolisthesis menggunakan algoritma C4.5 (Handayani, 2017) menghasilkan akurasi sebesar 89%. Penelitian ini akan menguji performa algoritma C5.0 dalam melakukan identifikasi penyakit Disk Hernia dan Spondylolisthesis.

Penelitian ini bertujuan untuk menerapkan algoritma data mining kedalam bidang medis dalam membantu mendiagnosa penderita penyakit disk hernia dan spondylolisthesis. Manfaat dari penelitian ini adalah mempermudah pihak medis dalam mendiagnosa penyakit disk hernia dan spondylolisthesis.

2. METODE PENELITIAN

Penelitian ini bermula dengan melakukan telaah pada permasalahan yang akan diselesaikan kemudian melakukan analisis dan pemetaan solusi dengan menggunakan sumber daya yang sesuai dan tersedia. Metode penelitian yang digunakan berupa pembuatan aplikasi dan pengujian. Pembuatan aplikasi ditujukan untuk mempermudah diagnosa penyakit tulang belakang dengan menggunakan algoritma C5.0.

2.1 TINJAUAN PUSTAKA

2.1.1 VERTEBRAL COLUMN ATAU TULANG BELAKANG

Penelitian yang dilakukan oleh Handayani (2017), dengan judul "Perbandingan Algoritma C4.5 dan K-Nearest Neighbor pada Klasifikasi Penyakit Disk Hernia dan Spondylolisthesis dalam Kolumna Vertebralis", membahas membandingkan hasil klasifikasi dan kecepatan waktu proses klasifikasi dari penyakit Disk Hernia dan Spondylolisthesis dengan menggunakan Algoritma C4.5 dan K-Nearest Neighbor. Pengujian yang dilakukan dengan menggunakan 310 data pasien dengan kondisi normal 100 data, pasien dengan kondisi disk hernia 60 data, dan pasien dengan kondisi spondylolisthesis 150 data, menghasilkan akurasi 89% menggunakan Algoritma C4.5 dengan kecepatan waktu proses klasifikasi 0,00912297 dan akurasi 83% untuk K-Nearest Neighbor dengan kecepatan waktu proses klasifikasi 0,000212303.

Pada peneliti sebelumnya, Sukmawati dan Pujiyanta (2014) telah dilakukan penelitian tentang penyakit tulang belakang dengan judul "Deteksi Penyakit Tulang Menggunakan Jaringan Syaraf Tiruan dengan Menggunakan Metode Backpropagation". Penelitian ini membahas bagaimana membangun suatu program Jaringan Syaraf Tiruan dengan menggunakan metode

Backpropagation untuk mendiagnosa penyakit tulang. Data yang digunakan adalah 42 gejala dengan 10 penyakit. Hasil pengujian menunjukkan akurasi sebesar 90%.

2.1.2 ALGORITMA C5.0

Algoritma C5.0 adalah salah satu metode data mining khususnya pada teknik pohon keputusan. Algoritma ini adalah algoritma penyempurnaan dari algoritma pohon keputusan sebelumnya, yaitu algoritma ID3 dan C4.5.

Wijaya, dkk (2018) melakukan penelitian klasifikasi pendapatan masyarakat dengan menggunakan algoritma C5.0. Dengan menggunakan algoritma ini, peneliti mampu menemukan hubungan tersembunyi antara variabel input dan variabel target, di mana hubungan ini adalah faktor terpenting dalam menentukan pendapatan masyarakat. Pengujian yang dilakukan dengan menggunakan aplikasi WEKA, diperoleh pohon keputusan dengan hasil yang maksimal.

Algoritma C5.0 ini juga pernah digunakan oleh Kastawan, dkk (2018) untuk melakukan penilaian kinerja Pegawai Negeri Sipil yang menjadi aturan yang dapat digunakan sebagai masukan pengambilan keputusan. Penggunaan data 284 pegawai ini diperoleh tingkat akurasi sebesar 96,08%. Hasil akurasi yang tinggi ini mampu dikembangkan menjadi ekinerja keputusan yang baik sebagai rekomendasi kenaikan pangkat dan tunjangan kinerja.

Selain penggunaan algoritma C4.5, kNN, dan JST untuk identifikasi penyakit, algoritma lain yang bisa digunakan untuk mendeteksi penyakit adalah algoritma C5.0 (Haryanto, Shaufiah, & Nhita, 2013; Rafe, Farhoud, & Rasoolzadeh, 2014; Hassoon, Kouhi, Zomorodi-Moghadam, & Abdar, 2017; Meng, Huang, Rao, Zhang, & Liu, 2013). Penggunaan algoritma C5.0 untuk prediksi penyakit diabetes dan typhus mampu menghasilkan akurasi lebih dari 80% (Haryanto, Shaufiah, & Nhita, 2013), memiliki akurasi yang lebih baik dari ANN dan Logistic Regression dalam memprediksi penyakit diabetes (Meng, Huang, Rao, Zhang, & Liu, 2013), mampu mencapai akurasi 97,98% dengan sensitivity sebesar 99,51% ketika dikombinasikan dengan boosting method (Rafe, Farhoud, & Rasoolzadeh, 2014), dan mampu meningkatkan akurasi prediksi penyakit hati dari 81% menjadi 93% ketika dikombinasikan dengan algoritma genetika (Hassoon, Kouhi, Zomorodi-Moghadam, & Abdar, 2017).

Algoritma ini dimulai dengan menentukan *root node* dengan menghitung informasi *gain* tertinggi dari atribut. Rumus menghitung ukuran atribut adalah (Kastawan, dkk, 2018):

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

Info (D) atau *entropy*, merupakan informasi untuk mengklasifikasikan label kelas dari *tuple* acak yang ada di D. Variabel pi adalah peluang yang nilainya bukan nol dari dari *tuple* acak yang ada di D. Dalam menghitung *entropy*, terdapat fungsi log basis 2 untuk mengkodekan kedalam bit. Hasil *entropy* kemudian dipartisi oleh A:

$$Info_A(D) = \sum_{j=1}^Y \frac{|D_j|}{D} \times Info(D_j) \quad (2)$$

Information gain pada atribut A dapat dihitung menggunakan rumus:

$$GAIN(A) = Info(D) - Info(D_j) \quad (3)$$

Hasil dari menghitung GAIN (A) dapat digunakan untuk menghitung banyaknya cabang dari atribut A. *Information Gain* tertinggi dari atribut A akan dipilih sebagai atribut node N.

2.1.3 AdaBoost

Algoritma Adaboost (*Adaptive Boosting*) adalah algoritma untuk meningkatkan klasifikasi dengan menggabungkan pengklasifikasian kuat dengan pengklasifikasian lemah (Mulyati, dkk, 2017). Algoritma ini dirumuskan oleh Yoav Freund dan Robert Schapire. Adapun rumus AdaBoost sebagai berikut:

$$F(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (4)$$

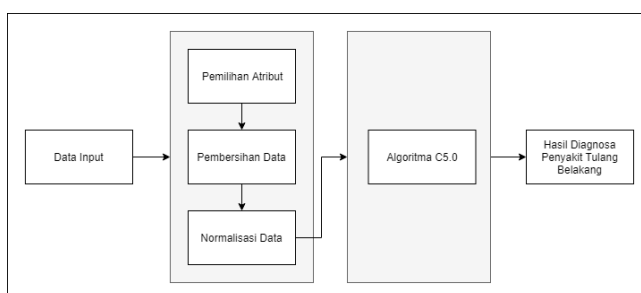
Untuk menghitung pengklasifikasian kuat maka $h_t(x)$ sebagai pengklasifikasian lemah dikalikan dengan α_t yang merupakan *learning rate*.

2.2 DESAIN SISTEM

Secara garis besar, penelitian ini memiliki 2 proses utama, yaitu *preprocessing* data dan pembuatan pohon keputusan menggunakan algoritma C5.0 yang ditunjukkan pada gambar 1. Penelitian dimulai dari pengumpulan data, dimana data yang digunakan pada penelitian ini tidak dibuat melainkan data sekunder yang diperoleh dari repository machine learning milik UCI Irvine (Barreto, Neto, & Filho, 2011). Data yang diperoleh dari UCI Irvine tentang penyakit Disk Hernia dan Spondylolisthesis sebanyak 310 data yang terdiri dari kategori Normal (100 pasien), Disk Hernia (60 pasien), dan Spondylolisthesis (150 pasien). Data yang diolah dari penyakit ini adalah *pelvic incidence* (pi), *pelvic tilt* (pt), *lumbar lordosis angle* (la), *sacral slope* (ss), *pelvic radius* (pr), dan *degree spondylolisthesis* (gs).

Tahap selanjutnya adalah melakukan pemilihan terhadap data yang telah disiapkan. Data yang telah dipilih atributnya tersebut kemudian dibersihkan untuk mencari data yang sesuai untuk melakukan klasifikasi. Data yang sudah dibersihkan kemudian dinormalisasi agar bisa diolah. Data kemudian dibagi kedalam dua kategori, yaitu data latih untuk melatih sistem untuk mempelajari serta membentuk pohon keputusan dan data uji untuk menguji akurasi atau kemampuan sistem dalam melakukan diagnosa secara benar. Dari 310 data yang diperoleh, 248 data sebagai data latih dan 62 data sebagai data uji.

Setelah data selesai dibagi, maka dari data latih dibangun model untuk mengidentifikasi penyakit Disk Hernia dan Spondylolisthesis menggunakan algoritma C5.0. Model dari sistem ini akan menentukan node dengan menentukan *information gain* terbesar dari atribut yang ada.



Gambar 1. Desain Sistem

2.3 TAHAPAN PENELITIAN

Tahapan pertama dalam penelitian ini adalah dengan melakukan pengumpulan data dan analisis data, dimana data diambil dari database UCI Irvine machine learning Repository, kemudian dilakukan pra proses data yaitu dengan melakukan pemilihan atribut dan pembersihan data. Setelah itu, data akan dinormalisasikan. Hasil dari normalisasi data ini, data akan dibagi sebagai data latih dan data uji.

Tahap selanjutnya adalah implementasi sistem yang dilakukan adalah dengan membuat pohon keputusan dan membuat sistem diagnosa. Tahap akhir dalam penelitian ini adalah melakukan pengujian dan evaluasi sistem.

3. HASIL DAN PEMBAHASAN

3.1 DESKRIPSI DATA

Data yang digunakan dalam penelitian ini adalah data penyakit disk hernia dan spondylolisthesis. Data yang diperoleh sebanyak 310 data yang terdiri dari 150 pasien spondylolisthesis, 60 pasien

disk hernia, dan 100 pasien normal. Atribut yang digunakan dalam penelitian ini adalah pi,pt,la,ss,pr,dan gs.

3.2 PEMBAHASAN

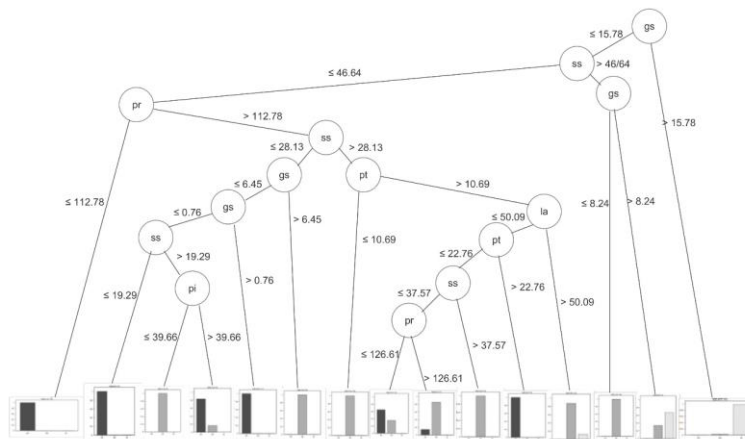
Dalam membuat pohon keputusan, data dibagi menjadi data latih dan data uji yaitu 248 data sebagai data latih dan 62 data sebagai data uji. Dari data latih, dibangun model pohon keputusan. *Information gain* tertinggi akan dijadikan sebagai *root* dalam pohon keputusan. *Information gain* tertinggi adalah atribut gs.

Pohon keputusan pada gambar 2 memiliki error sebesar 5.2%. Klasifikasi dalam pohon dibagi menjadi 3, yaitu DH, NO,dan SL. Adapun hasil klasifikasi ditunjukkan pada tabel 1.

Tabel 1. Hasil klasifikasi

	DH	NO	SL
DH	46	2	0
NO	5	72	4
SL	1	1	117

Dari hasil klasifikasi diperoleh penggunaan atribut gs sebesar 100%, atribut ss 52,82%, atribut pr 44,35%, atribut pt 36,69%, atribut la 12,90%, dan atribut pi 5,24%.



Gambar 2. Pohon Keputusan Algoritma C5.0

Pohon keputusan yang sudah diperoleh kemudian dimaksimalkan dengan menggunakan metode boost yang menghasilkan error 1,2%. Adapun hasil klasifikasi ditunjukkan pada tabel 2.

Tabel 2. Hasil klasifikasi setelah menggunakan metode boost

	DH	NO	SL
DH	48	0	0
NO	0	78	3
SL	0	0	119

Dari hasil klasifikasi diperoleh penggunaan atribut gs sebesar 100%, atribut 96.37%, atribut pr 74,60%, atribut la 65,73%, atribut pi 54,03%%, dan atribut 53,63%.

Setelah pohon keputusan selesai dibangun, maka proses selanjutnya adalah melakukan evaluasi. Evaluasi ini melakukan perbandingan hasil keputusan C5.0 dengan keputusan pada

data uji. Data uji yang digunakan adalah 62 data. Setelah dilakukan evaluasi, terdapat 49 data yang dapat diidentifikasi dengan benar oleh algoritma C5.0. Berdasarkan hasil evaluasi tersebut, pohon keputusan algoritma C5.0 memiliki akurasi sebesar 79%. Kemudian hasil tersebut dikuatkan dengan menggunakan metode boost sehingga menghasilkan 51 data yang dapat diidentifikasi dengan benar dengan akurasi yang diperoleh sebesar 83%.

KESIMPULAN

Dari hasil penelitian menggunakan data penyakit disk hernia dan spondylolisthesis maka dapat disimpulkan dengan menggunakan algoritma C5.0 dapat membentuk pohon keputusan dengan menentukan nilai information gain tertinggi dari atribut yang ada. Berdasarkan hasil evaluasi yang diperoleh, maka dapat disimpulkan bahwa algoritma C5.0 dapat digunakan untuk membantu dalam mendiagnosa penyakit disk hernia dan spondylolisthesis. Hasil akurasi yang diperoleh dari 62 data uji adalah 79%. Setelah menggunakan metode boost akurasi sistem mengalami kenaikan dari 79% menjadi 83%.

DAFTAR PUSTAKA

- Barreto, G. d., Neto, A. R., & Filho, H. A. (2011, 08 09). *Vertebral Column Data Set*. Retrieved from UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/Vertebral+Column>
- Handayani, I. (2017). *Perbandingan Algoritma C4.5 dan K-Nearest Neighbor pada Klasifikasi Penyakit Disk Hernia dan Spondylolisthesis dalam Kolumna Vertebralis*. Yogyakarta: Universitas Gadjah Mada.
- Haryanto, T., Shaufiah, & Nhita, F. (2013). *Prediksi Penyakit Demam Berdarah dan Typhus dengan Algoritma C5*. Telkom University.
- Hassoon, M., Kouhi, M. S., Zomorodi-Moghadam, M., & Abdar, M. (2017). Rule Optimization of Boosted C5.0 Classification using Genetic Algorithm for Liver Disease Prediction. *2017 International Conference on COmputer and Application (ICCA)* (pp. 299-305). Doha, UAE: IEEE.
- Kastawan, P. W., Wiharta, D. M., & Sudarma, I. M. (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro*, 371-376.
- Mayo Clinic. (2019, 04 26). *Herniated Disk*. Retrieved from Mayo Clinic: <https://www.mayoclinic.org/diseases-conditions/herniated-disk/symptoms-causes/syc-20354095>
- Meng, X.-H., Huang, Y.-X., Rao, D.-P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung Journal of Medical Sciences*, 93-99.
- Mulyati, S., Yulianti, & Saifudin, A. (2017). Penerapan Resampling dan AdaBoost untuk Penanganan Masalah Ketidakseimbangan Kelas Berbasis Naïve Bayes pada Prediksi Churn Pelanggan. *Jurnal Informatika Universitas Pamulang*, 190-199.
- Pallardy, R. (2019, 4 26). *Vertebral Column*. Retrieved from britannica.com: <https://www.britannica.com/science/vertebral-column>
- Rafe, V., Farhoud, S. H., & Rasoolzadeh, S. (2014). Breast Cancer Prediction by using C5.0 Algorithm and Boosting Method. *Journal of Medical Imaging and Health Informatics*, 600-604.
- Santosa, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics: Teori dan Implementasi Menggunakan Python & Apache Spark*. Yogyakarta: Penebar Media Pustaka.
- Sukmawati, K., & Pujiyanta, A. (2014). Deteksi Penyakit Tulang Menggunakan Jaringan Syaraf Tiruan dengan Metode Backpropagation. *Jurnal Sarjana Teknik Informatika*, 233-246.
- Tebet, M. A. (2014). Conceitos atuais sobre equilíbrio sagital e classificac,ão da espondilólise e espondilolistese. *Rev. Bras. Ortop.*, 3-12.
- Wijaya, A. C., Hasibuan, N. A., & Ramadhani, P. (2018). Implementasi Algoritma C5.0 dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus: Kelurahan Mesjid Kecamatan Medan Kota). *Majalah Ilmiah INTI*, 192-198.
-