

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PENYAKIT *DISK HERNIA* DAN *SPONDYLOLISTHESIS* DALAM KOLUMNA VERTEBRALIS

Irma Handayani

Program Studi Teknologi Informatika, Universitas Teknologi Yogyakarta
Jl. Siliwangi (Ringroad Utara) Jombor , Sleman, Yogyakarta
e-mail: irma.handayani@staff.uty.ac.id

Abstrak

Disk hernia dan Spondylolisthesis merupakan contoh penyakit yang dapat terjadi pada kolumna vertebralis atau tulang belakang. Suatu proses ekstraksi untuk mencari informasi dalam data yang belum diketahui sebelumnya dikenal dengan istilah data mining. Salah satu peranan utama data mining adalah klasifikasi. Klasifikasi banyak digunakan untuk menentukan keputusan sesuai pengetahuan baru yang didapat dari pengolahan data lampau menggunakan algoritma. Penelitian ini melakukan penerapan nilai akurasi algoritma C4.5 pada klasifikasi penyakit disk hernia dan spondylolisthesis serta kecepatan waktu pada proses klasifikasi. Proses klasifikasi dilakukan dengan cara memasukkan data dari sumber utama ke dalam sistem, kemudian melakukan proses perhitungan menggunakan metode algoritma C4.5. Hasil penelitian menunjukkan bahwa akurasi dari classifier C4.5 sebesar 89%. Rata-rata lama waktu yang dibutuhkan untuk melakukan klasifikasi classifier C4.5 0,00912297 detik.

Kata Kunci: Algoritma C4.5, Disk Hernia, Klasifikasi, Kolumna Vertebralis, Spondylolisthesis

1. Pendahuluan

Kolumna vertebralis atau rangkaian tulang belakang adalah sebuah struktur yang lentur dan dibentuk oleh sejumlah tulang yang disebut *vertebra* atau ruas tulang belakang. Di antara tiap dua ruas pada tulang belakang terdapat bantalan tulang rawan. Panjang rangkaian tulang belakang pada orang dewasa mencapai 57 sampai 67 cm. Seluruhnya terdapat 33 ruas tulang, 24 buah diantaranya adalah tulang-tulang terpisah dan 9 ruas sisanya bergabung membentuk tulang [1].

Spinal Cord Injury (SPI) atau cedera tulang belakang bisa menyebabkan gangguan permanen pada fungsi tubuh. Cedera tulang belakang pada umumnya disebabkan karena kecelakaan saat berkendara, jatuh, tindak kekerasan, cedera saat olah raga, dan penyakit seperti kanker, artritis, serta osteoporosis. Berbagai macam trauma pada tulang belakang disebabkan oleh banyak faktor dan menyebabkan kegagalan fungsi. *Disk hernia* terjadi ketika diskus rusak karena beberapa kondisi dengan penyebab paling umum diantaranya adalah jatuh atau kecelakaan, punggung tertarik secara berulang, punggung tertarik secara tiba-tiba saat mengangkat atau memutar punggung dengan kasar, degenerasi diskus akibat penuaan, dan herniasi spontan yang terjadi tanpa didahului cedera tertentu. Ketika diskus rusak, bagian tengah diskus yang kenyal dan lembut akan menonjol keluar melalui titik lemah pada lapisan luar jaringan *fibrosa* sehingga menyebabkan tonjolan menekan saraf di dekatnya [2]. Pada keadaan normal tulang belakang tersusun bertumpuk satu sama lain dalam satu garis lurus. *Spondylolisthesis* adalah suatu kondisi tulang belakang dimana salah satu vertebra bergeser ke depan (*anterolisthesis*) atau ke belakang (*retrolisthesis*) dengan vertebra selanjutnya [3].

Klasifikasi penyakit pada tulang belakang atau *kolumna vertebralis* (*column vertebral*) merupakan salah satu implementasi dari teknik *machine learning* di bidang kedokteran. Penelitian tentang klasifikasi penyakit atau kerusakan tulang dan sendi dari sistem kerangka masih jarang dilakukan karena belum adanya suatu database yang memiliki atribut numeric yang secara kuantitatif mampu menggambarkan penyakit tersebut [4].

Data mining adalah suatu cara yang bertujuan dalam penemuan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam database atau sumber data lain yang

dimanfaatkan untuk menyelesaikan suatu masalah melalui berbagai aturan proses [5].

Penelitian terdahulu tentang klasifikasi telah dilakukan [6] penelitian yang membandingkan algoritma C4.5 dan algoritma K-NN yang digunakan untuk klasifikasi penyakit diabetes mellitus. Hasil penelitian menunjukkan 76,105% pada algoritma C4.5 dan 79,1436% pada algoritma K-NN. Kemudian [7] telah melakukan komparasi algoritma K-NN, Naïve Bayes, dan C4.5 untuk klasifikasi penyakit kanker payudara. Penelitian ini menghasilkan nilai akurasi 94,70% pada algoritma K-NN, 95,85% pada Naïve Bayes, dan 94,71% pada algoritma C4.5.

Penelitian ini melakukan penerapan metode algoritma C4.5 pada klasifikasi penyakit *disk hernia* dan *spondylolisthesis* dalam kolumna vertebralis. Data disusun dalam dua klasifikasi namun terkait. Pertama mengklasifikasikan pasien sebagai salah satu dari kategori yaitu Normal (100 pasien), *Disk hernia* (60 pasien), dan *Spondylolisthesis* (150 pasien). Kedua, kategori *Disk hernia* dan *Spondylolisthesis* digabung menjadi satu kategori yang diberi label sebagai "abnormal". Dengan demikian, tugas kedua mengklasifikasikan pasien sebagai salah satu dari dua kategori: Normal (100 pasien) atau Abnormal (210 pasien).

2. Metode penelitian

2.1. Analisis Kebutuhan Sistem

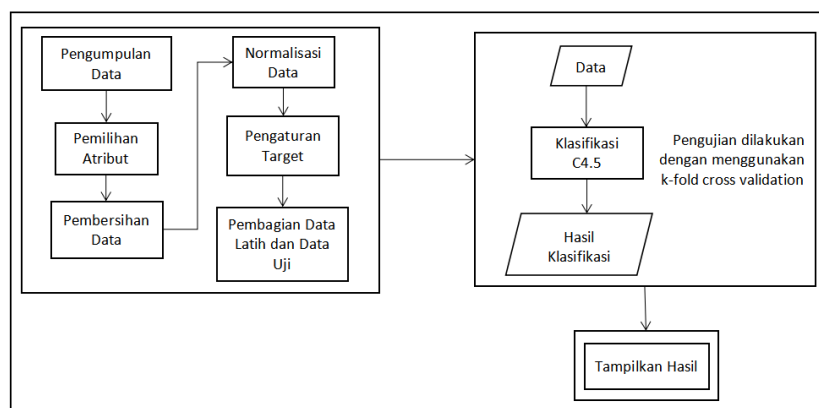
Sistem yang dirancang merupakan sistem yang digunakan untuk mengklasifikasi data pasien yang ada menjadi beberapa kelas, yaitu kelas normal dan kelas abnormal. Untuk kelas abnormal sendiri dibagi menjadi 2 subkelas, yaitu kelas *disk hernia* untuk pasien penderita *disk hernia* dan kelas *spondylolisthesis* untuk pasien penderita *spondylolisthesis*. Pembagian kelas tersebut digunakan berdasarkan nilai dari masing-masing yang dimiliki masing-masing pasien, yaitu *pelvic incidence*, *pelvic tilt*, *lumbar lordosis angle*, *sacral slope*, *pelvic radius*, dan *degree spondylolisthesis*.

2.2. Data Input

Data input merupakan data yang akan digunakan sebagai masukan pada sistem. Data input ini kemudian akan diolah menggunakan metode klasifikasi C.45 untuk menentukan kelas pasien. Data yang digunakan antara lain: *Pelvic incidence* (PI) merupakan sudut antara garis tegak lurus antara pelat sakral dan garis yang menghubungkan titik tengah pelat sakral ke sumbu *bicoxofemoral*; PI merupakan nilai yang spesifik dan konstanta untuk masing-masing pasien; *Pelvic tilt* (PT) merupakan orientasi panggul yang menghubungkan tulang paha dengan bagian tubuh lainnya. PT bisa digerakkan ke depan, belakang, maupun ke arah lainnya; *Lumbar lordosis angle* (LA) merupakan ciri khas tulang belakang manusia dan merupakan acuan postur tubuh seorang pasien itu baik atau buruk; *Sacral slope* (SS) merupakan kemiringan yang terjadi antara *sacral plate* (pelat sakral) dengan bidang horizontal (*horizontal plate*); *Pelvic radius* (PR) merupakan nilai yang mempengaruhi perkembangan dari *lumbar lordosis* besar; dan *Degree spondylolisthesis* (DG) sering juga disebut *grade of spondylolisthesis* merupakan pengukuran yang menyatakan seberapa besar bagian tubuh pasien yang tergelincir ke depan melebihi tubuh bagian bawah.

2.2. Deskripsi Sistem

Sistem dirancang untuk dapat melakukan klasifikasi terhadap data *Kolumna Vertebralis* dengan menggunakan algoritma C.45. Penyakit kemudian dibagi menjadi 3 kelas, yaitu kelas *disk hernia*, *spondylolisthesis*, dan normal. Proses yang diterapkan pada sistem dibagi menjadi 3 tahap, antara lain tahap *pre-processing*, *design classifier*, dan *post-processing*. Alur sistem dipaparkan pada Gambar 1.



Gambar 1. Alur Sistem

Tahap *pre-processing* merupakan tahapan yang dimulai dari proses pengumpulan data. Data yang dikumpulkan kemudian dikelompokkan berdasarkan pengaruh terhadap kelas masing-masing. Setelah itu data dinormalkan dimana data dimasukkan ke dalam kelas yang sesuai. Setelah dinormalkan, data kemudian dibagi menjadi 2 bagian, yaitu data latih dan data uji dengan pembagian 70% data latih dan 30% data uji secara acak.

Setelah tahap *pre-processing* selesai, data kemudian dimasukkan kedalam *classifier* sebagai pengetahuan. *Classifier* kemudian belajar dari data yang telah dimasukkan dan dievaluasi. Jika dari atribut yang ditetapkan ada yang belum dilatih, maka proses latihan sistem akan diulang dengan struktur dan fungsi yang berbeda.

Tahap ketiga adalah tahap *post-processing* dimana hasil klasifikasi ditampilkan kedalam bentuk yang lebih mudah untuk dipahami. Sistem akan menampilkan apakah pasien tersebut normal atau terkena penyakit kolumna vertebralis dan jenis penyakit yang diderita apakah itu *disk hernia* atau *spondylolisthesis*.

2.3. Klasifikasi

Klasifikasi adalah proses menemukan sekumpulan model/ fungsi yang menjelaskan dan membedakan data kedalam kelas-kelas tertentu, dengan tujuan menggunakan model tersebut dalam menentukan kelas dari suatu objek yang belum diketahui kelasnya. Ada dua proses klasifikasi yang terdiri dari: proses *learning/ training* adalah melakukan pembangunan model menggunakan data *training*, dimana tiap-tiap *record* pada data latih dianalisis berdasarkan nilai-nilai atributnya dengan menggunakan algoritma C4.5 untuk mendapatkan model. Kemudian proses *testing* yaitu melakukan tes terhadap data testing menggunakan model yang telah diperoleh dari proses *training* [8].

2.4. Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma yang terdapat dalam klasifikasi data mining dan merupakan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Secara umum algoritma C4.5 untuk membangun pohon keputusan yakni memilih atribut sebagai akar, membuat cabang untuk tiap-tiap nilai, membagi kasus dalam cabang, kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [9].

Komponen-komponen yang menyusun algoritma C4.5 dalam bentuk pohon keputusan yaitu:

a. Entropy

Entropy merupakan distribusi probabilitas dalam teori informasi dan diadopsi ke dalam algoritma C4.5 untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan (*data set*). Sebagai ilustrasi semakin tinggi tingkat entropy dari sebuah data set maka semakin homogen distribusi kelas pada data set tersebut. Perhitungan entropy ditunjukkan pada persamaan (1).

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

Dimana S adalah himpunan kasus, n adalah jumlah partisi S , dan P_i adalah proposisi dari S_i terhadap S .

b. Information Gain

Setelah membagi *data set* berdasarkan sebuah atribut ke dalam subset yang lebih kecil, entropy dari data tersebut akan berubah. Perubahan entropy ini dapat digunakan untuk menentukan bagus tidaknya pembagian data yang telah dilakukan. Perubahan entropy ini disebut dengan *information gain* dalam algoritma C4.5. *Information gain* ini diukur dengan mengitung selisih antara entropy data set sebelum dan sesudah pembagian (*splitting*) dilakukan. Pembagian yang terbaik akan menghasilkan entropy subset yang paling kecil, dengan demikian berdampak pada *information gain* yang terbesar. Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada dan dapat ditunjukkan dengan persamaan (2).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Dimana S adalah himpunan kasus, A adalah atribut yang dihitung, n adalah jumlah partisi pada atribut A , $|S_i|$ adalah jumlah kasus pada partisi ke i , dan $|S|$ adalah jumlah kasus dalam S .

2.4.1. Normalisasi data dan pembagian kelas atribut

Normalisasi data dilakukan agar terjadi keseimbangan data pada setiap atribut yang digunakan. Pembagian kelas atribut yang digunakan untuk klasifikasi menggunakan algoritma C4.5 dibagi menjadi *low*, *normal*, dan *high* [10]. Selain pembagian kelas atribut, data yang digunakan untuk perhitungan ditunjukkan pada Tabel 1.

Tabel 1. Data Latih Kolumna Vertebralis

No.	PI	PT	LA	SS	PR	DS	Diagnosa	Jenis Penyakit
1	53.94	9.31	43.1	44.64	124.4	25.08	AB	SL
2	84.97	33.02	60.86	51.95	125.66	74.33	AB	SL
3	89.01	26.08	69.02	62.94	111.48	6.06	NO	NO
4	85.35	15.84	71.67	69.51	124.42	76.02	AB	SL
5	45.37	10.76	29.04	34.61	117.27	-10.68	AB	DH
6	50.75	20.24	37	30.52	122.34	2.29	NO	NO
7	77.11	30.47	69.48	46.64	112.15	70.76	AB	SL
8	77.41	29.4	63.23	48.01	118.45	93.56	AB	SL
9	74.38	32.05	78.77	42.32	143.56	56.13	AB	SL
10	50.91	6.68	30.9	44.24	118.15	-1.06	NO	NO

Kemudian data hasil normalisasi ditunjukkan pada Tabel 2.

Tabel 2. Normalisasi Data

No.	PI	PT	LA	SS	PR	DS	Diagnosa	Jenis Penyakit
1	Normal	Normal	High	Normal	Normal	High	AB	SL
2	High	High	High	High	High	High	AB	SL
3	High	High	High	High	Low	High	NO	NO
4	High	Normal	High	High	Normal	High	AB	SL
5	Normal	Normal	Normal	Normal	Normal	Low	AB	DH
6	Normal	Normal	High	Normal	Normal	Normal	NO	NO
7	Normal	High	High	High	Low	High	AB	SL
8	Normal	High	High	High	Normal	High	AB	SL
9	Normal	High	High	Normal	High	High	AB	SL
10	Normal	Normal	Normal	Normal	Normal	Normal	NO	NO

3. Hasil dan Analisis

3.1. Proses Penyusunan Data Penyakit

Data-data yang digunakan sebagai data latih dan data uji adalah data mengenai penyakit *disk hernia* dan *spondylolisthesis*. Terdapat sebanyak 310 data latih yang terdiri dari 150 data kasus penyakit *spondylolisthesis*, 60 data kasus penyakit *disk hernia*, dan 100 data kasus normal.

3.2. Proses Klasifikasi C4.5

Proses klasifikasi C4.5 dilakukan dengan menelusuri posisi data uji berdasarkan pohon yang telah dibangun berdasarkan data latih sebelumnya. Proses pembentukan pohon dilakukan dengan tahapan sebagai berikut:

a. Pilih atribut sebagai akar

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Nilai *gain* masing-masing atribut dihitung dengan menggunakan persamaan (2).

b. Buat cabang untuk tiap-tiap nilai

Setelah mendapatkan root pada tahapan sebelumnya yaitu GS, maka langkah selanjutnya adalah membuat cabang dari pohon keputusan untuk mencari solusi atau mencari node selanjutnya. GS mempunyai 3 atribut, yaitu *low*, *normal*, dan *high*.

c. Bagi kasus dalam cabang

Proses pembagian kasus dalam cabang ini dilakukan sama seperti proses awal, tetapi nilai GS digunakan sebagai root sehingga nilai entropi yang digunakan adalah nilai entropi GS.

d. Ulangi proses yang sama pada semua cabang

Proses perhitungan *gain* akan terus dilakukan hingga pohon menemukan solusi atau hingga ditemukan *leaf*, dimana perhitungan tidak bisa dilanjutkan lagi.

3.3. Proses Klasifikasi

Proses klasifikasi merupakan proses dimana data masukkan akan diklasifikasi oleh sistem kedalam kelas yang sesuai berdasarkan pohon yang telah dibentuk sebelumnya. Data uji yang telah dinormalisasi dimasukkan kedalam sistem untuk diklasifikasi. Data yang telah diinputkan kemudian akan melakukan pencarian atau peruntukan kecabang mana pada pohon sehingga sampai pada solusi.

3.4. Pengujian Sistem

Pengujian dilakukan dengan memasukkan data uji satu persatu kedalam sistem dan kemudian dicatat hasil klasifikasi dan *running time* yang dibutuhkan sistem untuk melakukan klasifikasi. Perhitungan akurasi dilakukan dengan menggunakan persamaan (5). Berdasarkan hasil pengujian terhadap *classifier* C4.5 diperoleh tingkat akurasi sistem sebesar 89% dan detail hasil pengujian diperlihatkan pada Tabel 8.

Tabel 3. Pengujian *Classifier* C4.5

No	Jenis penyakit	Jumlah data	Akurasi	Rata-rata <i>Running time</i> (detik)
1	<i>Spondylolisthesis</i>	150	96%	0,00912297
2	Normal	100	84%	
3	Disk Hernia	60	86%	
Total		310	89%	

4. Kesimpulan

Algoritma C4.5 dapat digunakan untuk proses klasifikasi penyakit *Disk hernia* dan *Spondylolisthesis* dalam Kolumna Vertebralis dengan nilai akurasi yang didapat adalah 89% dari total jumlah data uji sebanyak 310 data. Kemudian rata-rata *running time* 0,00912297 detik. Masih diperlukan variable input tambahan yang mendukung agar tingkat akurasinya lebih tinggi.

References

- [1] E. C. Pearce, *Anatomi dan Fisiologi untuk Paramedis*. PT. Gramedia Pustaka Utama, Jakarta, 2012.
- [2] J. Jordan, K. Konstantinou, and J.O'Dowd, "Herniated Lumbar Disc: injection interventions for sciatica," *Clin. Evid.*, vol. 9, no. February, pp. 34-44, 2016.
- [3] J. C. Eck, "Spondylolisthesis," pp. 12–15, 2012.[4] A. R. Da Rocha Neto and G. D. A. Barreto, "On the application of ensembles of classifiers to the diagnosis of pathologies of the vertebral column: A comparative analysis," *IEEE Lat. Am. Trans.*, vol. 7, no. 4, pp. 487–496, 2009.
- [5] I. H. Witten, E. Frank, and M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*, 2011.
- [6] G. Karyono, "Analisis teknik data mining 'algoritma c4.5 dan k-nearest neighbor' untuk mendiagnosa penyakit diabetes mellitus," *Semin. Nas. Teknol. Informasi, Blnis, dan Desain*, pp. 77–82, 2016.
- [7] Kurniawan, M, F., and Ivandari, 2017, Komparasi Algoritma Data Mining Untuk Klasifikasi Penyakit Kanker Payudara.
- [8] G. Abdillah *et al.*, "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means," *Semin. Nas. Teknol. Informasi dan Komunikasi*, pp. 18–19, 2016.
- [9] S. Lorena, W. Zarman, and I. Hamidah, "Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik," *Pros. Semin. Nas. Apl. Sains dan Teknol.*, no. November, pp. 263–272, 2014.
- [10] M. A. Tebet, "Current concepts on the sagittal balance and classification of spondylolysis and spondylolisthesis," *Rev. Bras. Ortop. (English Ed.)*, vol. 49, no. 1, pp. 3–12, 2014.



Irma Handayani lahir di Yogyakarta, 23 Agustus 1989. Memperoleh gelar Sarjana dari STMIK AKAKOM, Yogyakarta Indonesia pada 2014, Magister dari Universitas Gadjah Mada, Indonesia pada 2018. Saat ini, bekerja di Teknik Informatika, Universitas Teknologi Yogyakarta sebagai Dosen. Bidang penelitian saat ini adalah Data Mining.

Alamat Email: irma.handayani@staff.uty.ac.id