

IMPLEMENTASI ALGORITMA C.45 DALAM PREDIKSI PENYAKIT KANKER

Warid Yunus ¹, Abd Rahmat Karim Haba ^{2*}

^{1,2*} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo, Kota
Gorontalo, Provinsi Sulawesi Utara, Indonesia.

Email: warid.dsn@gmail.com ¹, rkarimhaba@gmail.com ^{2*}

Histori Artikel:

Dikirim 20 November 2022; *Diterima dalam bentuk revisi* 26 Desember 2022; *Diterima* 3 Januari 2023; *Diterbitkan* 10 Januari 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Pendeteksi kanker serviks belum di jalankan secara stabil sebab pasien datang ke RSUD AlieSaboe di mana kondisi pasien sudah memasuki fase lanjut stadium. Sehingga hasil pengobatan sangat rendah, maka akan lebih baik pengobatan dilakukan dalam stadium lebih dini. Agar angka kematian penderita di harapkan menurun. Faktor yang mempengaruhi pasien terlambat datang dengan alasan malu, takut apabila mengetahui hasil deteksi penyakitnya. Tujuan penelitian ini adalah untuk memperoleh Model Algoritma C.45 dalam memprediksi penderita penyakit kanker serviks. Hasil yang dicapai penelitian ini adalah memperoleh hasil prediksi dengan menggunakan algoritma C.45 dan memiliki keakuratan sebesar 70.37% yang diukur dengan menggunakan Confusion Matrix pada tools Rapid Miner.

Kata Kunci: : Kanker Serviks; Prediksi; Algoritma C45.

Abstract

Cervical cancer detection has not been carried out stably because the patient came to the AlieSaboe Hospital where the patient's condition had entered an advanced stage. So that the results of treatment are very low, it would be better if the treatment is carried out at an earlier stage. So that the death rate of patients is expected to decrease. Factors that influence patients arriving late with reasons of embarrassment, fear of knowing the results of the detection of their disease. The purpose of this study was to obtain the C.45 Algorithm Model in predicting cervical cancer patients. The results achieved by this study are to obtain prediction results using the C.45 algorithm and have an accuracy of 70.37% as measured using the Confusion Matrix on Rapid Miner tools.

Keyword: Cervical Cancer; Prediction; Algorithm C45.

1. Pendahuluan

Penderita kanker serviks di Gorontalo tergolong cukup tinggi. Berdasarkan data dari Dinas Kesehatan Provinsi Gorontalo mencatat bahwa pada 2017 saja ada 98 wanita usia 30-50 tahun positif iVa. Jumlah ini bisa saja bertambah jika cakupan Iva tes yang dilakukan Dinas Kesehatan provinsi, kabupaten/Kota dapat berjalan optimal. Pasalnya, di tahun 2017 sasaran perempuan Gorontalo sebanyak 173.604 orang, progres cakupan Iva tes hanya mencapai 5 persen (2.572 perempuan Gorontalo) dari target 30 persen (52.082 perempuan) usia 30-50 tahun. Hal inilah yang membuat pemerintah Provinsi Gorontalo kesulitan untuk mendeteksi dini kanker mulut rahim tersebut. Pada tahun 2018 jumlah perempuan yang memeriksakan diri hanya bertambah sedikit. Sampai dengan bulan Juni 2018, Dinas Kesehatan Provinsi Gorontalo baru berhasil memeriksa 1087 perempuan dan ditemukan 31 perempuan positif Iva.

Berdasarkan tes wawancara di bagian promosi kesehatan di RSUD Aloe Saboe pendeteksi dini kanker serviks belum di jalankan secara stabil sebab pasien datang ke RSUD di mana kondisi pasien sudah memasuki fase lanjut stadium. Sehingga hasil pengobatan sangat rendah, maka akan lebih baik pengobatan dilakukan dalam stadium lebih dini. Agar angka kematian penderita di harapkan menurun. Faktor yang mempengaruhi pasien terlambat datang dengan alasan malu, takut apabila mengetahui hasil deteksi penyakitnya. Mereka juga takut tidak mampu membayar biaya pengobatannya. Kondisi seperti ini perlu segera mendapatkan penanganan. Dari sinilah peran proses data mining berdasarkan data sebelumnya akan diolah dengan cara prediksi dini dan menerapkan metode C45 diinginkan untuk mempercepat kesadaran dan pemahaman warga, agar diri perlu diperiksa lebih dini terhadap kanker serviks.

Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1]. *Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* [2]. Prediksi merupakan sebuah proses dimana memperkirakan tentang sesuatu yang paling mungkin akan terjadi di masa yang akan datang [3]. Algoritma adalah upaya untuk memecahkan masalah dengan menggunakan urutan operasi yang logis dan sistematis untuk menghasilkan keluaran tertentu. Selain itu, juga mempunyai karakteristik yang berbeda untuk setiap algoritmanya misalnya pada algoritma data mining. Terdapat lima kategori algoritma data mining diantaranya klasifikasi, estimasi, prediksi, *clustering*, dan asosiasi [4].

Penelitian ini menggunakan salah satu metode Data Mining yaitu Algoritma C4.5. Algoritma C4.5 merupakan Algoritma yang digunakan dalam *decision tree*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) [5]. Konsep dari Algoritma C4.5 adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*). C4.5 adalah algoritma yang cocok untuk masalah [6]. Hasil dari proses klasifikasi pada algoritma c45 yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru [7].

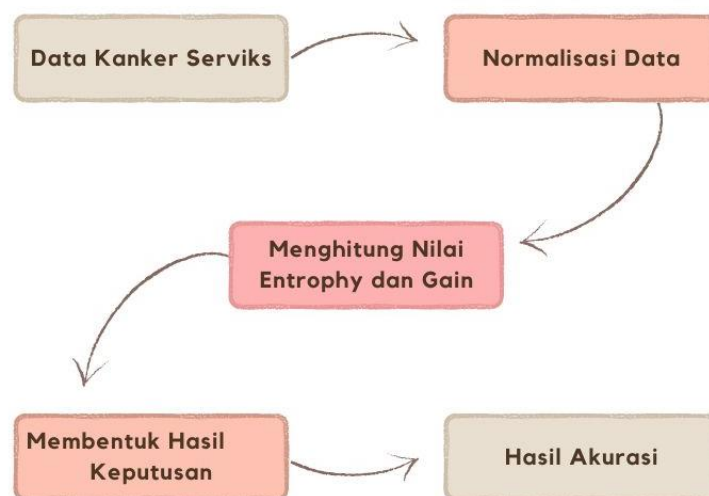
Pada penelitian oleh Ade Fatma Ayu Rahman. Penelitian ini bertujuan untuk mengetahui tingkat akurasi Algoritma C45 dalam prediksi kelulusan Mahasiswa, Penelitian ini menunjukkan bahwa akurasi yang diperoleh dalam melakukan prediksi kelulusan mahasiswa sebesar 88,74% [8]. Pada penelitian oleh T. Adi Kurniawan, hasil yang diperoleh dari penelitian ini dapat digunakan untuk memprediksi daerah rawan atau tidak rawan penyakit demam berdarah di kabupaten tangerang, Hasil tingkat keberhasilan prediksi daerah rawan secara keseluruhan yang telah diukur adalah sebesar 91% valid. Yaitu untuk daerah Pasir Gadung, Talaga Dan Suka Nagara [9]. Penelitian dilakukan oleh Sanni Ucha Putri, pada penelitian ini dengan menggunakan Algoritma C4.5 dapat membantu peneliti dalam prediksi penyakit diabetes berdasarkan variabel-variabel yang diperlukan yaitu usia, tekanan darah, denyut nadi, berat badan dan kadar gula darah. Pada penelitian ini menghasilkan jumlah prediksi positif sebanyak 36 data dan jumlah prediksi negatif sebanyak 2 Prediksi positif dari seluruh data menghasilkan 90,00% [10].

Dengan diterapkannya algoritma C4.5 dalam proses prediksi penyakit kanker serviks maka dapat digunakan untuk menganalisa atribut yang paling berpengaruh dalam prediksi penyakit kanker

serviks dan dikeluarkan dalam bentuk hasil perhitungan *information Gain* dari algoritma C4.5. Dengan algoritma C4.5 akan diurutkan atribut yang paling tinggi nilainya. Adapun attribut yang akan digunakan yaitu Merokok, kurangnya konsumsi buah dan sayur, obesitas, penggunaan kontrasepsi minum, melahirkan usia muda, jumlah melahirkan, dan usia. Rumusan Masalah pada penelitian ini yaitu bagaimana hasil pemodelan algoritma C.45 untuk prediksi jumlah penderita kanker serviks. Tujuannya adalah untuk memperoleh Model Algoritma C.45 yang paling baik kinerjanya dalam memprediksi jumlah penderita penyakit kanker serviks.

2. Metode Penelitian

Berikut diagram alir atau tahapan dalam metode penelitian :



Gambar 1. Diagram Alir Penelitian (Sumber: Penulis)

Berikut tahapan dalam diagram alir pada penelitian ini :

- 1) Data Penelitian
Pada data penelitian ini terdiri dari tahun atau periode 2017 sampai 2021 mengenai penyakit kanker serviks yang diperoleh pada RSUD AlieSaboe
- 2) Normalisasi Data
Data yang diperoleh akan dilakukan normalisasi yang berguna untuk mempermudah dalam pengolahan data sebelum masuk ke pengolahan metode c4.5
- 3) Nilai *Entrophy* dan *Gain*
Entrophy dan *Gain* ini berguna untuk mencari nilai tertinggi yang berada pada metode C45 dari attribut yang digunakan pada penelitian ini berupa: Merokok, Obesitas, Melahirkan Usia Muda, Melahirkan, Usia
- 4) Membentuk Hasil Pohon Keputusan
Dalam pembentukan pohon keputusan terdiri dari memilih attribut sebagai akar keputusan, membuat cabang untuk tiap nilai attribut, bagi kasus dalam cabang, dan Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.
- 5) Hasil Akurasi
Dalam mencari *performance* penelitian ini menggunakan akurasi *Confusion Matrix*.

3. Hasil dan Pembahasan

Hasil dari penelitian ini adalah dengan adanya prediksi, maka dalam mengetahui penyakit kanker serviks dapat dilakukan sehingga dapat memberikan hasil sesuai dengan yang diharapkan. Berikut pembahasan dari penelitian ini :

3.1. Data Penelitian

Data yang digunakan penelitian ini sebanyak 135 *record* data penelitian, berikut data penelitian:

Tabel 1. Data Penelitian

No	Merokok	Usia	Obesitas	Melahirkan Usia Muda	Hasil
1	Tidak	24 Tahun	Ya	Ya	Ya
2	Tidak	30 Tahun	Tidak	Tidak	Ya
3	Ya	31 Tahun	Ya	Tidak	Ya
4	Ya	30 Tahun	Ya	Tidak	Tidak
5	Tidak	32 Tahun	Tidak	Tidak	Tidak
..					
135	Tidak	30 Tahun	Tidak	Tidak	Ya

3.2 Normalisasi Data

Dalam normalisasi data ini peneliti melakukan pengelompokkan data dan attribut yang digunakan. Berikut tampilan dari normalisasi data :

- 1) Merokok : Ya dan Tidak
- 2) Usia : 20-30 Tahun = Muda
31-40 Tahun = Tua
- 3) Obesitas : Ya dan Tidak
- 4) Melahirkan Usia Muda : Ya dan Tidak

Tabel 2. Normalisasi Data

No	Merokok	Usia	Obesitas	Melahirkan Usia Muda	Hasil
1	Tidak	Muda	Ya	Ya	Ya
2	Tidak	Muda	Tidak	Tidak	Ya
3	Ya	Tua	Ya	Tidak	Ya
4	Ya	Muda	Ya	Tidak	Tidak
5	Tidak	Tua	Tidak	Tidak	Tidak
..					
135	Tidak	Muda	Tidak	Tidak	Ya

3.3 Nilai *Entropy* dan Nilai *Gain*

Dalam metode C45 dapat dilakukan berdasarkan metode C45, berikut persamaan dari nilai *entropy* dan nilai *Gain*:

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

Keterangan :

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi S_i terhadap S

Kemudian hitung nilai *Gain* menggunakan rumus

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

Keterangan :

S = Himpunan kasus

A = Fitur

n = Jumlah partisi atribut A

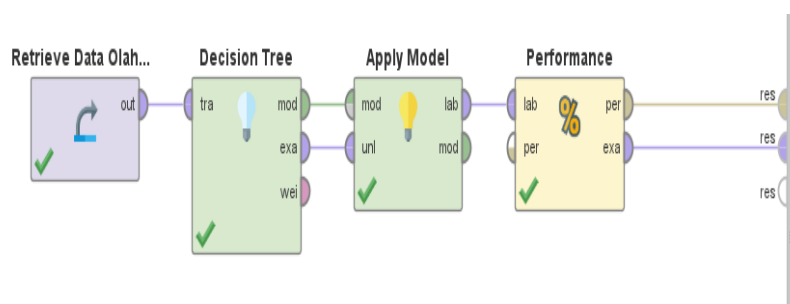
|Si| = Proporsi Si terhadap S

|S| = Jumlah kasus dalam S

3.4 Membentuk Hasil Keputusan

Berikut tahapan dalam membentuk pohon keputusan dalam tahapan ini peneliti menggunakan tools Rapid Miner dalam Membentuk Hasilnya. RapidMiner adalah perangkat lunak sumber terbuka. RapidMiner juga menggunakan berbagai teknik prediktif dan deskriptif untuk berbagi informasi dengan pengguna sehingga mereka dapat membuat keputusan terbaik [7].

1) Tampilan Pembentukan Model



Gambar 2. Pembentukan Model

Pada gambar diatas menunjukkan merupakan tampilan dalam melakukan prediksi, yang mana tampilan ini terdiri dari pemasukan data penelitian, algoritma C45 (*Decision Tree*), *Apply model* dan *performance*.

2) Hasil Prediksi

Row No.	Hasil	prediction(Hasil)
1	Tidak	Tidak
2	Tidak	Tidak
3	Ya	Ya
4	Tidak	Tidak
5	Ya	Ya
6	Ya	Ya
7	Ya	Ya
8	Ya	Tidak
9	Tidak	Tidak
10	Tidak	Tidak
11	Ya	Tidak
12	Tidak	Tidak
13	Tidak	Tidak

Gambar 3. Hasil Prediksi

3) Hasil Akurasi

☒ Table View ☐ Plot View

accuracy: 70.37%

	true Tidak	true Ya	class precision
pred. Tidak	76	40	65.52%
pred. Ya	0	19	100.00%
class recall	100.00%	32.20%	

Gambar 4. Hasil Akurasi

Pada gambar 4 merupakan hasil akurasi dalam memprediksi kanker serviks yang mana hasil keakuratannya mencapai 70.37%.

4. Kesimpulan

Berdasarkan hasil pada penelitian ini maka dapat disimpulkan bahwa dengan menggunakan algoritma C45 dapat dilakukan prediksi penyakit kanker Serviks dan memiliki keakuratan sebesar 70.37%.

5. Daftar Pustaka

- [1] Pambudi, R.H., 2017. *Penerapan Algoritma C4. 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal* (Doctoral dissertation, Universitas Brawijaya).
- [2] Novianti, B., Rismawan, T. and Bahri, S., 2016. Implementasi Data Mining Dengan Algoritma C4. 5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak). *Coding Jurnal Komputer dan Aplikasi*, 4(3). DOI: <http://dx.doi.org/10.26418/coding.v4i3.17034>.
- [3] Ginting, V.S., Kusrini, K. and Taufiq, E., 2020. Implementasi Algoritma C4. 5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), pp.36-44. DOI: <http://dx.doi.org/10.35585/inspir.v10i1.2535>.
- [4] Lia, E., 2022. *Implementasi Algoritma C4. 5 Untuk Penentuan Produk Terlaris (Studi Kasus: Toko XYZ Palembang)* (Doctoral dissertation, Universitas Katolik Musi Charitas). URL: <http://eprints.ukmc.ac.id/8634/>.
- [5] Prakarsya, A. and Prambayun, A., 2020. Implementasi Data Mining Untuk Prediksi Penyebaran Virus Hiv/Aids Di Bandar Lampung Dengan Teknik Decision Tree. *Jurnal Sistem Informasi Komputer dan Teknologi Informasi (SISKOMTI)*, 3(2), pp.18-26. URL: <https://www.ejournal.lembahdempo.ac.id/index.php/SISKOMTI/article/view/120>.
- [6] Dewi, K. R. ., Mauladi, K. F. ., & Masruroh., 2020. Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 4(3), 109–114. DOI: <https://doi.org/10.29407/inotek.v4i3.72>.

- [7] Suwandi, Endah Gustami, Kurnia Puji Astuti, & Yunita. (2022). Klasifikasi Barang Menggunakan Algoritma C4.5 Dalam Penentuan Prediksi Stok Barang pada PT Aerofood Indonesia. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, 3(2), 12-18. DOI: <https://doi.org/10.31294/reputasi.v3i2.1398>.
- [8] Rahman, A.F.A., Sorikhi, and Wartulas., 2020. Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4. 5 (Studi Kasus Di Universitas Peradaban). *Indonesian Journal of Informatics and Research*, 1(2), pp.70-77.
- [9] IHFANDI, A., 2018. *Implementasi Data Mining Untuk Prediksi Daerah Rawan Penyakit Demam Berdarah Menggunakan Algoritma C4. 5 (Studi Kasus: Dinas Kesehatan Kabupaten Tangerang)* (Doctoral dissertation, Universitas Satya Negara Indonesia). URL: <http://repo.usni.ac.id/68/>.
- [10] A. A. Ronaldi and Nanang Hunafi., 2020. Implementasi Data Mining Untuk Prediksi, *Semin. Nas. Has. Penelit. ...*, vol. 1, no. 1, pp. 250–257.