

PREDIKSI DAERAH RAWAN PENYAKIT DEMAM BERDARAH MENGGUNAKAN ALGORITMA C4.5

Ananda Mutia Dewi¹⁾, Putri Tia Novita²⁾, Ghiyalti Novillia³⁾, Lindawati⁴⁾

^{1,3,4}Prodi Informatika Medis, Fakultas Kesehatan Teknologi dan Sains Universitas Bumi Persada

²Prodi Administrasi Rumah Sakit, Fakultas Kesehatan Teknologi dan Sains Universitas Bumi Persada

*Correspondence: husnamaulida88@gmail.com

ABSTRAK:

Penelitian ini mengembangkan sebuah aplikasi berbasis algoritma C4.5 untuk melakukan prediksi dan analisis data secara akurat dan efisien. Tujuan utama dari studi ini adalah menciptakan suatu sistem yang mampu memproses data dengan cepat tanpa mengorbankan keandalan hasil. Algoritma C4.5 diterapkan untuk membangun pohon keputusan (*decision tree*) yang memanfaatkan sejumlah atribut data guna menghasilkan prediksi yang optimal. Metodologi penelitian mencakup beberapa tahapan, yaitu studi literatur untuk mengumpulkan dasar teori, perancangan model sistem melalui diagram alur (*flowchart*) dan diagram UML, implementasi algoritma, serta pengujian dan evaluasi terhadap kinerja sistem. Hasil penelitian menunjukkan bahwa penggunaan algoritma C4.5 dalam mengklasifikasikan data melalui pembentukan pohon keputusan mampu menghasilkan prediksi yang valid dengan tingkat akurasi yang tinggi. Selain itu, pendekatan ini terbukti efisien dalam hal waktu pemrosesan, sehingga cocok untuk diterapkan pada dataset dengan kompleksitas menengah. Evaluasi lebih lanjut mengindikasikan bahwa pemilihan atribut yang tepat turut berperan penting dalam meningkatkan performa klasifikasi. Temuan ini membuka peluang pengembangan lebih lanjut, seperti optimasi algoritma atau integrasi dengan teknik *machine learning* lainnya untuk meningkatkan skalabilitas sistem.

Kata Kunci: Algoritma C4.5, Decision Tree, Klasifikasi Data, Data Mining

ABSTRACT:

This study develops an application based on the C4.5 algorithm to perform accurate and efficient data prediction and analysis. The main objective of this study is to create a system that is able to process data quickly without sacrificing the reliability of the results. The C4.5 algorithm is applied to build a decision tree that utilizes a number of data attributes to produce optimal predictions. The research methodology includes several stages, namely literature studies to collect theoretical basis, system model design through flowcharts and UML diagrams, algorithm implementation, and testing and evaluation of system performance. The results of the study indicate that the use of the C4.5 algorithm in classifying data through the formation of a decision tree is able to produce valid predictions with a high level of accuracy. In addition, this approach is proven to be efficient in terms of processing time, making it suitable for application to datasets with medium complexity. Further evaluation indicates that the selection of the right attributes also plays an important role in improving classification performance. This finding opens up opportunities for further development, such as algorithm optimization or integration with other machine learning techniques to improve system scalability.

Keyword : C4.5 Algorithm, Decision Tree, Data Classification, Data Mining

PENDAHULUAN

Demam berdarah dengue (DBD) termasuk penyakit infeksi yang banyak ditemukan di daerah beriklim tropis dan subtropis. Penyakit ini ditularkan melalui vektor nyamuk *Aedes aegypti* sebagai perantaranya. Gejala klinis yang muncul pada penderita DBD antara lain hiperpireksia (demam tinggi), disertai dengan keluhan mual/muntah, cephalgia (nyeri kepala), nyeri abdomen, dan leukopenia. Hingga saat ini, DBD tetap menjadi tantangan serius dalam bidang kesehatan masyarakat. Dalam hal ini, Dinas Kesehatan memegang peranan penting dalam upaya pengendalian lingkungan untuk pencegahan penyakit.

Dalam aktivitas sehari-hari, manusia senantiasa berhadapan dengan beragam persoalan dari berbagai aspek kehidupan. Permasalahan yang dihadapi memiliki variasi kompleksitas yang beragam, mulai dari isu sederhana dengan sedikit variabel terkait hingga masalah yang sangat kompleks dengan banyak faktor yang perlu dipertimbangkan. Sebagai solusi atas tantangan ini, dikembangkanlah sistem pendukung yang memudahkan manusia dalam proses pengambilan keputusan.

Salah satu solusi yang efektif adalah penerapan pohon keputusan (decision tree). Metode ini merupakan representasi sistematis yang memungkinkan analisis berbagai faktor terkait suatu masalah secara komprehensif. Melalui pendekatan ini, dapat diidentifikasi hubungan kausal antar variabel serta ditentukan solusi optimal dengan mempertimbangkan seluruh parameter yang relevan. Pohon keputusan juga mampu melakukan analisis risiko dan menghitung nilai informasi dari setiap alternatif solusi.

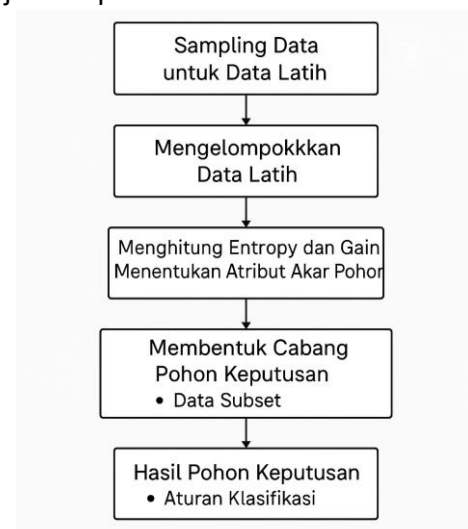
Sebagai alat bantu pengambilan keputusan (decision support system), konsep pohon keputusan telah berkembang pesat sejak penerapan teori

graf dalam struktur pohon. Fleksibilitas dan efektivitas metode ini menjadikannya banyak diaplikasikan dalam berbagai sistem pengambilan keputusan di berbagai bidang.

METODE

Penelitian ini merupakan studi komputasional yang mengintegrasikan pendekatan kuantitatif dan eksplanatoris dalam bidang machine learning. Fokus utama penelitian adalah mengembangkan model prediksi risiko Demam Berdarah Dengue (DBD) dengan memanfaatkan algoritma C4.5 sebagai dasar pembuatan pohon keputusan. Pendekatan kuantitatif digunakan untuk menganalisis data klinis dan epidemiologis secara statistik, sementara aspek eksplanatoris membantu menjelaskan hubungan kausal antara berbagai faktor risiko dengan kejadian DBD.

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma decision tree C4.5 untuk mengklasifikasikan status kerawanan wilayah terhadap demam berdarah dengue (DBD). Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian Dataset yang digunakan berasal dari data agregat 36 kelurahan yang mencakup variabel jumlah penderita DBD,

jumlah penduduk, rata-rata kepadatan penduduk per kilometer, dan status rawan atau tidak rawan terhadap DBD.

Tahapan awal dalam analisis adalah melakukan sampling data untuk dijadikan sebagai data latih. Langkah selanjutnya adalah menghitung nilai entropy dan information gain untuk setiap atribut guna menentukan atribut mana yang paling informatif sebagai akar pohon keputusan. Setelah atribut akar dipilih, dilakukan pembentukan cabang pohon keputusan berdasarkan subdivisi dari nilai atribut tersebut. Proses ini terus dilanjutkan hingga setiap simpul pohon hanya mengandung satu jenis keputusan (rawan atau tidak rawan), sehingga tidak diperlukan perhitungan lanjutan.

Hasil akhir dari pohon keputusan menghasilkan tujuh aturan klasifikasi, seperti: “Jika jumlah penderita 7–14 maka tidak rawan”, dan “Jika jumlah penderita 15–22, rata-rata penduduk 11–16, dan jumlah penduduk 25.001–34.000 maka rawan”. Aturan-aturan ini dapat digunakan untuk mendukung pengambilan keputusan dalam pemetaan wilayah rawan DBD dan perencanaan intervensi kesehatan masyarakat.

HASIL DAN PEMBAHASAN

A. HASIL

Penelitian ini bertujuan untuk mengklasifikasikan tingkat kerawanan wilayah terhadap penyakit demam berdarah dengue (DBD) dengan menerapkan algoritma decision tree C4.5. Peneliti menggunakan data dari 36 kelurahan yang mencakup atribut jumlah penderita DBD, jumlah penduduk, rata-rata kepadatan penduduk per kilometer, dan status kerawanan wilayah (rawan atau tidak rawan).

Proses awal dimulai dengan penyusunan dataset latih dari tabel data kelurahan yang kemudian dikategorikan

berdasarkan nilai-nilai atributnya. Atribut “jumlah penderita” dibagi menjadi tiga rentang: 7–14, 15–22, dan 23–34. Atribut lain seperti “jumlah penduduk” dan “rata-rata kepadatan” juga dikelompokkan ke dalam tiga rentang nilai. Selanjutnya, dilakukan perhitungan nilai entropy dan gain dari setiap atribut untuk menentukan atribut mana yang paling relevan dalam membagi data secara informatif. Hasil perhitungan menunjukkan bahwa atribut “jumlah penderita” memiliki nilai gain tertinggi (0,808949), sehingga digunakan sebagai akar pohon keputusan (root node).

Setelah atribut akar ditetapkan, dilakukan pembentukan cabang pohon berdasarkan nilai atribut tersebut. Untuk data yang belum dapat diklasifikasikan sepenuhnya, dilakukan iterasi lanjutan dengan menggunakan subset data untuk menghitung gain atribut lainnya. Dalam kasus kedua, atribut “rata-rata penduduk” memberikan gain tertinggi dan digunakan sebagai node berikutnya. Hal yang sama dilakukan pada kasus ketiga hingga semua cabang dapat mengklasifikasikan status kerawanan secara jelas.

Hasil akhir dari proses ini adalah sebuah pohon keputusan yang menghasilkan tujuh aturan klasifikasi. Aturan tersebut, misalnya, menetapkan bahwa wilayah dengan jumlah penderita antara 7–14 tergolong tidak rawan, sedangkan jika jumlah penderita 23–34 maka tergolong rawan. Kombinasi kompleks seperti jumlah penderita 15–22 dengan kepadatan penduduk tertentu dan jumlah penduduk tertentu juga menghasilkan klasifikasi rawan atau tidak rawan tergantung pada nilai atributnya.

Pohon keputusan yang dihasilkan dapat dijadikan alat bantu dalam pengambilan keputusan strategis, seperti penentuan wilayah prioritas intervensi DBD. Dengan demikian, metode C4.5 terbukti efektif dalam mengolah data

spasial dan demografis untuk klasifikasi status kesehatan masyarakat.

B. PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa algoritma decision tree C4.5 dapat digunakan secara efektif untuk mengklasifikasikan wilayah berdasarkan tingkat kerawanan terhadap penyakit demam berdarah dengue (DBD). Proses klasifikasi dilakukan dengan menggunakan atribut-atribut yang relevan, yaitu jumlah penderita DBD, jumlah penduduk, rata-rata kepadatan penduduk per kilometer, dan status rawan/tidak rawan.

Dalam tahap awal, perhitungan entropy dan gain dilakukan untuk setiap atribut guna menentukan mana yang paling informatif untuk menjadi node akar pada pohon keputusan. Nilai entropy mencerminkan ketidakpastian atau impuritas dalam data, sedangkan gain menunjukkan seberapa besar pengurangan ketidakpastian yang dihasilkan oleh suatu atribut. Atribut "jumlah penderita" memperoleh nilai gain tertinggi sebesar 0,808949, sehingga ditetapkan sebagai akar pohon keputusan. Hal ini logis mengingat jumlah penderita secara langsung merepresentasikan intensitas kejadian DBD di suatu wilayah.

Proses pembentukan pohon keputusan tidak berhenti sampai penetapan node akar saja. Ketika sebuah cabang masih mengandung data dengan status klasifikasi yang bercampur (baik rawan maupun tidak rawan), dilakukan proses lanjutan menggunakan subset data untuk menentukan atribut terbaik berikutnya. Pada kasus data dengan jumlah penderita 15–22, atribut "rata-rata kepadatan penduduk" menjadi pembeda yang paling signifikan, dengan nilai gain sebesar 0,458106. Hal ini menunjukkan bahwa selain jumlah penderita, kepadatan penduduk juga menjadi faktor penting dalam menentukan kerawanan terhadap DBD.

Pada tahap lanjutan, untuk kasus dengan atribut "rata-rata penduduk" pada rentang 11–16 yang belum memberikan

klasifikasi tunggal, digunakan kembali atribut lainnya seperti "jumlah penduduk" dan "status kelurahan". Pada akhirnya, melalui proses iteratif ini, diperoleh struktur pohon keputusan yang utuh dan menghasilkan aturan klasifikasi yang eksplisit.

Tujuh aturan klasifikasi yang dihasilkan dari pohon keputusan akhir mencerminkan pola-pola tertentu yang menghubungkan variabel demografis dengan risiko DBD. Sebagai contoh, wilayah dengan jumlah penderita dalam rentang 7–14 diklasifikasikan sebagai tidak rawan tanpa memperhatikan atribut lainnya. Sebaliknya, wilayah dengan penderita antara 23–34 secara langsung dikategorikan rawan. Namun untuk jumlah penderita yang berada di tengah (15–22), diperlukan kombinasi atribut tambahan seperti kepadatan dan jumlah penduduk untuk mencapai keputusan yang akurat.

Kondisi ini memperlihatkan bahwa kasus-kasus DBD tidak hanya dipengaruhi oleh jumlah kasus saja, tetapi juga oleh faktor lingkungan demografis yang menjadi konteks persebarannya. Semakin tinggi kepadatan penduduk, semakin besar pula kemungkinan penyebaran penyakit, terutama untuk penyakit yang ditularkan melalui vektor seperti nyamuk *Aedes aegypti*. Oleh karena itu, atribut kepadatan penduduk menjadi relevan sebagai penentu dalam proses klasifikasi.

Selain itu, hasil klasifikasi ini juga menunjukkan bahwa klasifikasi "potensi rawan" dan "endemis rawan" memiliki kecenderungan kuat untuk diklasifikasikan sebagai wilayah rawan. Sementara itu, status "sporadis" sebagian besar dikategorikan sebagai tidak rawan, yang secara konseptual sesuai dengan kenyataan bahwa sporadis menunjukkan penyebaran yang tidak merata atau bersifat insidental.

Secara keseluruhan, metode C4.5 tidak hanya berhasil mengidentifikasi atribut-atribut kunci dalam klasifikasi kerawanan DBD, tetapi juga memberikan representasi visual dalam bentuk pohon keputusan yang dapat diinterpretasikan dengan mudah oleh pembuat kebijakan atau petugas lapangan. Model ini bersifat

transparan dan interpretable, berbeda dengan model-model black-box seperti neural network yang membutuhkan pendekatan pasca-analisis untuk interpretasi.

Dari hasil ini, dapat disimpulkan bahwa model klasifikasi C4.5 sangat potensial untuk diaplikasikan dalam perencanaan intervensi kesehatan masyarakat, seperti penentuan prioritas wilayah fogging, kampanye PSN (Pemberantasan Sarang Nyamuk), dan edukasi masyarakat. Dengan menggunakan model ini, alokasi sumber daya dapat dilakukan secara lebih tepat sasaran dan berbasis data.

KESIMPULAN

Hasil dari pengembangan sistem berbasis algoritma C4.5 menunjukkan bahwa metode ini efektif dalam membentuk pohon keputusan yang dapat dimanfaatkan sebagai sistem pendukung keputusan bagi Dinas Kesehatan Kabupaten Tangerang. Dengan menggunakan data yang tersedia, model ini mampu memprediksi status kerawanan wilayah terhadap demam berdarah dengue secara akurat. Keunggulan dari pendekatan ini terletak pada kemampuannya mengklasifikasikan wilayah berdasarkan sejumlah atribut penting, sehingga dapat membantu dalam identifikasi awal daerah berisiko tinggi serta mempercepat proses pengambilan keputusan.

DAFTAR PUSTAKA

Abdillah, S. (n.d.). Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pemalang.

Prasetyo, Data Mining - Konsep dan Aplikasi Menggunakan MATLAB (p.2). Gresik: ANDI Yogyakarta.

Prasetyo, E. (2012). Data Mining - Konsep dan Aplikasi Menggunakan MATLAB. Gresik: Andi.

Rusda Wajhilla, Ihsan. (2015). Penerapan Algoritma C4.5 Terhadap Diagnosa Penyakit Demam Tifoid Berbasis Mobile. Swabumi Vol Iii No. 1, 24-30.

Tan P, Steinbach M, Kumar V. (2006). Introduction to Data Mining. In E.

Wibowo, I. A. (2017). Implementasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Demam Berdarah Di Kota Tangerang Selatan Menggunakan Algoritma K-Means. Jakarta.

Widoyono. (2008). Demam Berdarah. 60-63.