

Klasifikasi Diagnosa Pasien Di Klinik Sri Dengan Metode Decision Tree

¹Ihsan Fadlu Rahman, ²Wenny Ayu Shestia, ³Sanniyah Delti Rama, ⁴Salwa Azzahra,
⁵Aulia Agung Dermawan

^{1,2,3,4,5} Program Studi Manajemen Rekayasa,Fakultas Teknologi Industri

^{1,2,3,4,5} Institut Teknologi Batam The Vitka City Complex Jl. Gajah Mada, Tiban, Batam,
KepulauanRiau

e-mail: 2212018@studentiteba.ac.id, 2212019@studentiteba.ac.id,

2212040@studentiteba.ac.id,2212042@studentiteba.ac.id,agung.dermawan29@gmail.com

Abstract

The purpose of this study is to use the C.45 decision tree algorithm to categorize patient diagnoses at Sri Clinic. One of the best data mining methods for creating data-driven decisions is the decision tree. This study's data set comprises a wide range of characteristics. Medical treatment of patients, including their history of health, age, gender, diagnosis, and details regarding whether they are sick or not. Using the C.45 decision tree algorithm, a model was developed to predict. Diagnosis is based on patterns found in patient data. The outcomes demonstrated the Decision Tree model's excellent level of accuracy in diagnosing conditions. Execution it is anticipated that this technique will assist clinics in making diagnoses more quickly and accurately, and enhancing the standard of care provided to patients. The characteristics of medical care that have the biggest impact on diagnosis were also found in the study.

Keywords: *Decision Tree, Data Mining, Patient Diagnosis, Medical Classification, Model Accuracy*

Abstrak

Tujuan dari penelitian ini adalah untuk menggunakan algoritma pohon keputusan C.45 untuk mengkategorikan diagnosis pasien di Klinik Sri. Salah satu metode penambangan data terbaik untuk membuat keputusan berbasis data adalah pohon keputusan. Kumpulan data penelitian ini terdiri dari berbagai karakteristik. Perawatan medis pasien, termasuk riwayat kesehatan, usia, jenis kelamin, diagnosis, dan rincian mengenai apakah mereka sakit atau tidak. Dengan menggunakan algoritma *decision tree* C.45, model dikembangkan untuk memprediksi diagnosa berdasarkan pola-pola yang ditemukan dalam data pasien. Hasilnya menunjukkan tingkat akurasi model *Decision Tree* yang sangat baik dalam mendiagnosis kondisi. Eksekusi Diharapkan bahwa teknik ini akan membantu klinik dalam membuat diagnosis lebih cepat dan akurat, dan meningkatkan standar perawatan yang diberikan kepada pasien. Ciri-ciri perawatan medis yang memiliki dampak terbesar pada diagnosis juga ditemukan dalam penelitian ini.

Kata kunci: Decision Tree, Data Mining, Diagnosa Pasien, Klasifikasi Medis, Akurasi Model

Diterima : Juni 2024

Disetujui : Juni 2024

Dipublikasi : Juni 2024

Pendahuluan

Klinik Sri, salah satu Klinik yang ada di Kota Batam, menyediakan berbagai layanan kesehatan termasuk diagnosa pasien. Dalam proses diagnosa, dokter menganalisis data pasien seperti hasil pemeriksaan fisik, gejala, dan riwayat kesehatan. Klasifikasi data diagnosa pasien sangat penting untuk mengidentifikasi pola dan tren dalam data, mempercepat pengambilan keputusan, serta meningkatkan kualitas layanan dan efisiensi operasional. Di era digital, teknologi informasi berperan penting dalam sektor kesehatan. Klinik Sri menghadapi tantangan dalam pengelolaan dan analisis data pasien yang semakin kompleks. Klasifikasi diagnosa menggunakan metode *Decision Tree* merupakan solusi efektif yang membagi data berdasarkan nilai atribut tertentu, menghasilkan keputusan atau klasifikasi yang mudah dipahami. Metode ini meningkatkan akurasi dan efisiensi diagnosis, membantu tenaga medis mengungkap pola penyakit, dan mendukung pencegahan serta penanganan penyakit secara proaktif.

Keunggulan *Decision Tree* adalah kemampuannya menangani data kompleks dan beragam, menghasilkan model yang mudah diinterpretasikan. Namun, tantangan utama termasuk kualitas data medis dan pemilihan atribut yang relevan. Penelitian ini bertujuan menerapkan metode *Decision Tree* di Klinik Sri untuk meningkatkan kualitas layanan kesehatan melalui analisis data yang lebih canggih dan teknologi informasi. Hasil penelitian diharapkan menjadi dasar pengembangan sistem pendukung keputusan klinis yang lebih efektif dan efisien, serta memberikan wawasan baru dalam pengelolaan dan analisis data pasien. Dengan penelitian ini, Klinik Sri dapat mengoptimalkan penggunaan data pasien untuk meningkatkan akurasi diagnosis dan kualitas pelayanan kesehatan, berdampak positif pada kesehatan pasien dan efisiensi operasional klinik, menjadikannya penyedia layanan kesehatan yang terpercaya dan inovatif.

Metode Penelitian

Pada metode penelitian ini, digunakan metode Klasifikasi Algoritma *Decision Tree C4.5* untuk mengelompokkan atribut atau kelas-kelas pada diagnosa pasien di klinik dan membentuk suatu pohon keputusan. Metode ini membantu dalam mengidentifikasi faktor-faktor penting yang mempengaruhi diagnosis penyakit pada pasien. Adapun kajian teori menurut para ahli sebagai berikut:

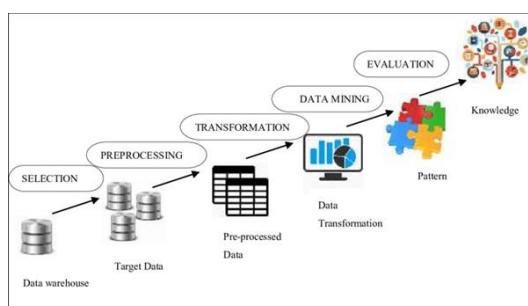
1. Diagnosa Pasien di Klinik

Klinik adalah jenis penyedia layanan yang menawarkan layanan. Menurut Ahman dan Indriani (2007), perusahaan jasa adalah organisasi yang terutama bergerak dalam penyediaan atau penjualan jasa dengan tujuan menghasilkan keuntungan. Dinyatakan secara berbeda, penyedia layanan memasarkan "barang" tidak berwujud. Sementara itu, klinik didefinisikan sebagai fasilitas pelayanan kesehatan perorangan dalam Peraturan Republik Indonesia No. 028/MENKES/PER/I/2011 Tentang Klinik, yang dikeluarkan oleh Menteri Kesehatan. yang, di bawah arahan spesialis medis dan diselenggarakan oleh berbagai petugas kesehatan, menawarkan layanan medis primer dan/atau khusus.

Klinik Sri adalah salah satu klinik ternama di Kota Batam yang menyediakan berbagai layanan kesehatan, termasuk diagnosa pasien. Dalam proses diagnosa, dokter Klinik Sri perlu menganalisis data pasien, seperti hasil pemeriksaan fisik, gejala, dan riwayat kesehatan, untuk menentukan diagnosa yang tepat. Klasifikasi data diagnosa pasien merupakan langkah penting dalam proses diagnosa, karena dapat membantu dokter dalam mengidentifikasi pola dan tren dalam data, serta mempercepat proses pengambilan keputusan. Klasifikasi data juga dapat membantu Klinik Sri dalam meningkatkan kualitas layanan kesehatan dan efisiensi operasional. Dalam memberikan pelayanan kesehatan pasien, klinik membutuhkan unit rekam medis yang bertanggung jawab terhadap pengelolaan data pasien menjadi informasi yang berguna bagi pengambilan keputusan (Budi, 2011 dalam Tominanto, 2015;33).

2. Data Mining

Penambangan data adalah proses mengidentifikasi dan mengekstraksi informasi yang relevan serta pengetahuan terkait dari berbagai database besar menggunakan metode statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. (Turban dan lainnya, 2005). Penambangan data, juga dikenal sebagai penemuan pengetahuan, adalah prosedur yang melibatkan penemuan tren dalam data yang perlu diproses. Hasil penting proses ini adalah informasi. Ada banyak istilah tambahan dengan definisi. *Knowledge Mining with Database* (KDD), proses ekstraksi informasi, analisis informasi/analisis pola, wawasan bisnis, arkeologi data, dan pengeringan data adalah istilah yang identik dengan data mining (Larose, 2005).



Gambar 1. Data Mining

3. Decision Tree

Salah satu metode model prediktif untuk mengklasifikasikan dan memprediksi kegiatan adalah pohon keputusan. (2003, Dunham). Mengubah tabel informasi menjadi model pohon adalah proses pohon keputusan. A aturan akan dihasilkan dan disederhanakan oleh model pohon (Basuki & Syarif, 2003).

Dalam C4.5, yang merupakan algoritma Larose, membuat pohon keputusan melibatkan beberapa langkah:

1. Persiapan latihan yang sudah disiapkan. Informasi pelatihan biasanya berasal dari data sebelumnya, yang merupakan informasi masa lalu yang sudah diatur ke dalam kategori tertentu.
2. Tentukan bilangan akar pohon. Setelah menentukan nilai keuntungan dari masing-masing karakteristik, nilai keuntungan tertinggi yang akan digunakan untuk mengekstrak akar dari atribut yang dipilih.yang pertama menghitung nilai entropia, atau akar pertama dari properti.
3. Formulir untuk Entropy Value dijelaskan:

S= Himpunan Kasus

A= Fitur

n= Jumlah partisi S

pi= Proporsi dari S terhadap S

$$\text{Entropy}(S) = \Sigma - pi * \log_2 pi$$

4. Preprocessing Data

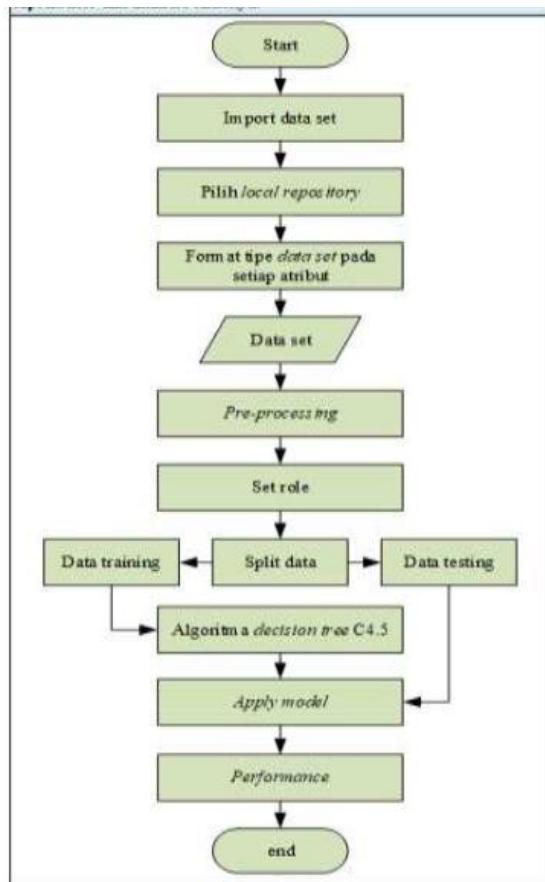
Data yang belum diproses disebut sebagai data mentah, dan perlu diperlakukan agar dapat dimanfaatkan dalam suatu proses. Data nyata atau mentah sering kali memiliki ketidakakuratan atau mengandung angka yang berbeda dari apa yang diantisipasi (Kumar & Chadha, 2012). Data yang tidak konsisten atau kurang di area tertentu, sehingga mengakibatkan ketidakakuratan. Adanya atribut informasi yang tidak tersedia atau devaluasi data tertentu (atribut) dapat mengakibatkan data tidak lengkap karena penghapusan data yang dianggap tidak perlu. Data dianggap tidak konsisten karena, selama titik pengumpulan data, suatu instrumen mungkin telah rusak oleh kesalahan manusia atau komputer juga, mungkin ada konvensi penamaan yang tidak konsisten antara dua set data yang identik (Han & Kamber, 2006).

5. Klasifikasi

Menurut Bafadal (2009:51) adalah tindakan memilih dan mengatur buku-buku perpustakaan dan bahan-bahan lainnya sesuai dengan serangkaian kriteria yang diberikan dan menyusunnya di satu lokasi. Proses mengklasifikasikan hal-hal atau konsep secara konseptual menjadi kelas hierarki, subklasi, dan subklase sesuai dengan kesamaan dan perbedaan mereka yang luas disebut klasifikasi. Secara umum, klasifikasi juga dianggap sebagai pengaturan yang teratur dari pengetahuan universal. Pekerjaan yang menilai suatu objek data agar masuk kedalam kelas tertentu dari sejumlah kelas yang sudah ada. Klasifikasi dibagi menjadi dua pekerjaan utama yaitu membangun model sebagai prototype yang disimpan sebagai memori dan menggunakan model untuk melakukan pengklasifikasian prediksi pada objek data lain agar diketahui terdapat dikelas mana objek data yang disimpan (Putri, Suparti, & Rahmawati, 2014).

Flowchart

Flowchart menunjukkan tahapan-tahapan dalam suatu penelitian atau proyek. Berikut:



Gambar 2. Flowchart

Pengolahan Data

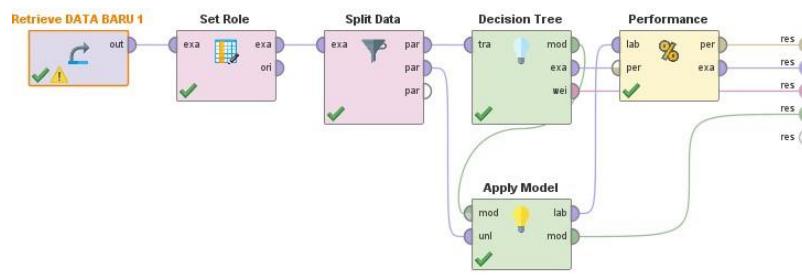
Kumpulan berkas diagnosis pasien dari Sri Clinics di Kota Batam berfungsi sebagai sumber data penelitian. Karakteristik yang digunakan algoritma adalah Diagnosis penyakit pasien (diagnosis), jenis kelamin, usia, dan informasi terkait lainnya membentuk diagnostik pasien ini. Untuk informasi tambahan pengelompokan table dibawah ini menunjukkan karakteristik diagnosis pasien.

Tabel 1. Data Set

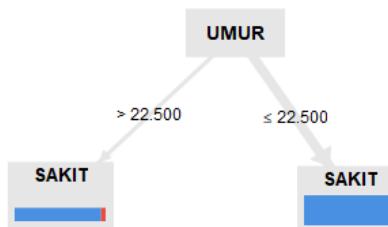
NO	NAMA PASIEN	UMUR	DIAGNOSA	KETERANGAN	JENIS KELAMIN
1	KLORA	14	PUSING, LAMBUNG	SAKIT	P
2	DESNIZAL	28	DIARE, PUSING	SAKIT	L
3	ZAHRA	3	LAMBUNG	SAKIT	P
4	TOPSAN	28	GASTRITIS/MULAS	SAKIT	L
5	SOFIA	18	PUSING, KEMBUNG	SAKIT	P
6	ABDAN	9	LUKA TERTUSUK	SAKIT	L
7	ISKANDAR	6	NYERI LAMBUNG	SAKIT	L
8	RITA	50	LAMBUNG	SAKIT	P
9	SUCIEM	45	DEMAM, PUSING	SAKIT	P
10	ALUIN	10	BATUK SESAK	SAKIT	P
11	KEISYA	7	DEMAM, BATUK	SAKIT	P
12	ARIF	4	DEMAM, MUNTAH	SAKIT	P
13	FAJAR	8	DEMAM, BAPIL	SAKIT	L
14	DAVID	8	BATUK, PILEK	SAKIT	L
15	HANIM	2	DEMAM, PILEK	SAKIT	L
16	JAWAHIR	33	KOLESTROL	SAKIT	L
17	ALFATIH	4	DEMAM	SAKIT	L
18	BETRAND	7	MUNTAH, KEMBUNG	SAKIT	L
19	MINCE	24	SUNAT DEWASA	TIDAK SAKIT	L
20	YOLANDA	24	DEMAM, FLU, RADANG	SAKIT	P
... 100					

Hasil dan Pembahasan

Pada proses pengolahan data pada RapidMiner dilakukan beberapa proses yaitu terlampir pada gambar berikut:



Gambar 3. RapidMiner



Gambar 4. Pohon Keputusan

Dalam proses pengolahan data pada RapidMiner ini, beberapa parameter diubah untuk menyesuaikan dengan kebutuhan analisis. Salah satu perubahan penting dilakukan pada operator "Set Role". Pada operator ini, atribut nama yang dipilih diubah menjadi "KET sakit atau tidak sakit diagnosa pasien". Selain itu, peran target diubah menjadi "label". Perubahan ini dilakukan agar penggambaran Decision Tree lebih fokus, dengan menargetkan diagnosa pasien sebagai hasil yang ingin diprediksi. Dengan demikian, model Decision Tree akan lebih spesifik dalam memprediksi apakah seorang pasien sakit atau tidak berdasarkan data yang tersedia.

Pada akurasi model ini sangat tinggi, yaitu 96.67%, yang menunjukkan bahwa model ini sangat baik dalam memprediksi dengan benar apakah pasien sakit atau tidak sakit. Dimana hasil ini menggambarkan jumlah-jumlah dari tiap label seperti dalam pred. SAKIT memiliki jumlah true SAKIT = 29 dan true TIDAK SAKIT = 1 , dan class precision menunjukkan bahwa dari semua pasien yang di prediksi sakit, 96,67% bener-bener sakit. Kemudian pada tabel pred. TIDAK SAKIT jumlah dari tiap tabel true SAKIT= 0 dan pred TIDAK SAKIT= 0 dan class precision=

0,00%. Tabel terakhir CLASS RECALL dengan 100% untuk kelas "SAKIT". Namun, model ini kurang efektif dalam mendeteksi pasien yang tidak sakit, dengan presisi dan recall 0% untuk kelas "TIDAK SAKIT". Hal ini menunjukkan bahwa model cenderung overfit pada data pasien yang sakit. Analisis lebih lanjut menunjukkan bahwa untuk individu dengan umur di bawah atau sama dengan 22,5 tahun, kemungkinan diagnosis penyakit demam lebih rendah, yang bisa diartikan sebagai kondisi sehat. Sebaliknya, untuk individu dengan umur di atas 22,5 tahun, kemungkinan diagnosis penyakit demam sedikit lebih tinggi, yang bisa diartikan sebagai kondisi tidak sehat. Ini menunjukkan bahwa umur merupakan faktor penting dalam menentukan diagnosis penyakit demam .

Kesimpulan

1. Algoritma Decision Tree C4.5 terbukti sangat efektif dalam membangun model klasifikasi untuk diagnosa pasien di Klinik Sri. Model ini menunjukkan akurasi yang tinggi, mencapai 96.67%, yang mengindikasikan bahwa algoritma ini dapat memprediksi diagnosa dengan tepat pada sebagian besar kasus.
2. Analisis data mengungkap bahwa terdapat beberapa atribut yang memiliki pengaruh signifikan terhadap hasil diagnosa pasien. Atribut-atribut ini dapat digunakan untuk meningkatkan akurasi model di masa depan dan membantu dalam pengambilan keputusan medis.
3. Implementasi model klasifikasi ini di Klinik Sri dapat membantu meningkatkan kualitas layanan kesehatan dengan mempercepat proses diagnosa dan mengurangi kesalahan. Ini juga mendukung efisiensi operasional klinik melalui penggunaan teknologi informasi yang lebih canggih.

Daftar Pustaka

- Turban, E, 2005, Sistem Pendukung Keputusan dan Sistem Cerdas Edisi Bahasa indonesia
Jilid 1. Andi: Yogyakarta.
- Larose, Daniel T. 2005. Menemukan Pengetahuan dalam Data: Pengantar Penambangan
Data John Willey & Sons, Inc.
- Ahman Dan Indriani, 2007. Panduan Klinik Dan Praktek Mandiri Bidan Bandung:
Grafindo Media Pratama.
- Basuki, Ahmad dan Syarif, Iwan, 2003. "Decision Tree". Surabaya: Politeknik
Electronika Negeri
- Dunham, Margareth H., 2003. "Data Mining Introductory and Advanced Topics".
New Jersey: Prentice Hall.

- Han, J. dan M. Kamber. 2006. Data Mining Concepts and Techniques Second Edition. SanFrancisco: Morgan Kaufmann.
- Kumar, V., & Chadha, A. (2012). Mining Association Rules in Student's Assessment Data. International Journal of Computer Science Issues , 9, 211-216.
- M. C., L. C., & D., A. K. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining . International Journal of Computer Science Issues , 257-264
- Bafadal, Ibrahim. 2009. Pengelolaan Perpustakaan Sekolah. Jakarta: Bumi Aksara.
- Putri, R. E., Suparti, & Rahmawati, R. (2014). Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun2012. Jurnal Gaussian, 3, 831–838. Retrieved
- Budi, S. C. (2011). Manajemen Unit Kerja Rekam Medis. Yogyakarta Quantum Sinergis Media.
- Tominanto. 2015. Perancangan Sistem Informasi Pendaftaran Rawat Jalan Klinik. ISSN: 2337-6007.

..