

IMPLEMENTASI DATA MINING UNTUK PREDIKSI PENYAKIT MALARIA TROPIKA DENGAN ALGORITMA C4.5

Nur Ain Banyal ^{*1}, Surianti ²
Universitas Doktor Husni Ingratubun Papua ^{1,2}
Jalan Raya Abepura Kotaraja Jayapura Papua
Sur-el : nur.ain.banyal@gmail.com^{*1}, surianti12p@gmail.com ²

Abstract : Malaria is a chronic and acute disease caused by protozoa of the Plasmodium type. There are 4 main species of plasmodium that cause malaria in humans, namely: Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae and Plasmodium ovale. The most important of these species is Plasmodium falciparum (tropic) because the consequences can be fatal and this species is also the one that causes the most deaths. This study aims to create a prediction model using Data Mining Algorithm C4.5 which produces a decision tree and testing carried out using Rapidminer so that prevention of tropical malaria can be examined early. In this study there are several classification attributes, namely normal, stage 1, stage 2, stage 3 and stage 4. The results of this study will be used as a reference to be able to see whether someone is at risk of tropical malaria or not based on predetermined attributes.

Keywords: Tropical Malaria, Data Mining, C.45 Algorithm, Decision Tree

Abstrak Malaria adalah penyakit kronis dan akut yang disebabkan oleh protozoa dari jenis Plasmodium. Ada 4 spesies yang utama dari jenis plasmodium yang menyebabkan penyakit malaria pada manusia, yaitu: Plasmodium Falciparum, Plasmodium Vivax, Plasmodium Malariae dan Plasmodium Ovale. Yang paling penting dari spesies ini adalah Plasmodium Falciparum (Tropika) karena akibatnya bisa fatal dan spesies ini juga yang paling banyak menyebabkan kematian. Pada penelitian ini bertujuan untuk membuat model prediksi menggunakan Data Mining Algoritma C4.5 yang menghasilkan sebuah pohon keputusan serta pengujian yang dilakukan dengan menggunakan Rapidminer agar pencegahan terhadap penyakit malaria tropika dapat dilakukan pemeriksaan sejak dini. Dalam penelitian ini terdapat beberapa atribut klasifikasi yaitu normal, stadium 1, stadium 2, stadium 3 dan stadium 4. Hasil dari penelitian ini akan dijadikan sebagai acuan untuk dapat melihat apakah seseorang beresiko terkena penyakit malaria tropika atau tidak berdasarkan atribut yang telah ditetapkan.

Kata kunci: Malaria Tropika, Data Mining, Algoritma C.45, Decision Tree

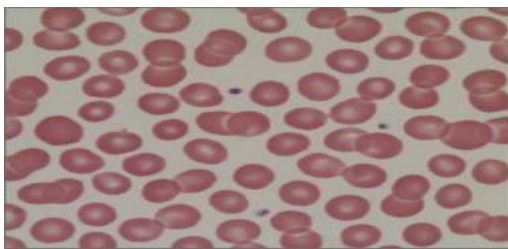
1. PENDAHULUAN

Malaria adalah penyakit kronis dan akut yang disebabkan oleh protozoa dari jenis Plasmodium. Ada 4 spesies yang utama dari jenis plasmodium yang menyebabkan penyakit malaria pada manusia, yaitu: Plasmodium Falciparum, Plasmodium Vivax, Plasmodium Malariae dan Plasmodium Ovale. Yang paling penting dari spesies ini adalah Plasmodium Falciparum (Tropika) karena akibatnya bisa fatal

dan spesies ini juga yang paling banyak menyebabkan kematian. [1]

Diagnosis malaria ditegakkan berdasarkan gejala klinis, pemeriksaan laboratorium, dan pemeriksaan penunjang. Diagnosis pasti dibuat dengan ditemukannya parasit malaria dalam pemeriksaan mikroskopis laboratorium. Pemeriksaan mikroskopik dengan pewarnaan Giemsa sampai saat ini masih merupakan baku emas pemeriksaan malaria. Walaupun demikian, hasil pembacaannya hanya dapat dipercaya jika

dilakukan oleh seorang yang berpengalaman. Kekurangannya adalah subjektivitas pemeriksaan, terutama dalam hal mengdiagnosa infeksi campuran atau infeksi dengan jumlah parasit yang rendah. Pemeriksaan mikroskopis, pemeriksaan ini meliputi pemeriksaan darah yang menurut teknis pembuatannya dibagi menjadi preparat darah (SD) tebal dan preparat darah tipis, untuk menentukan ada tidaknya parasit malaria dalam darah. Melalui pemeriksaan ini dapat dilihat jenis *Plasmodium* dan stadiumnya dalam hal ini *Plasmodium Falciparum*(Tropika) yang diteliti serta kepadatan parasitnya. Kepadatan parasit dapat dilihat melalui dua cara yaitu semi-kuantitatif dan kuantitatif. [1]



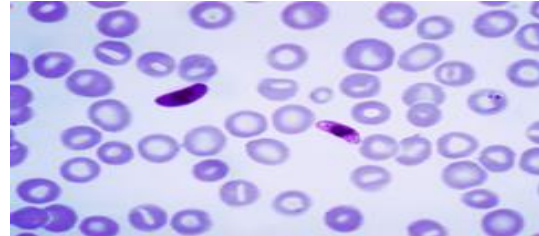
Gambar.1. Sel darah merah normal

Ciri - ciri diagnostik *Plasmodium Falciparum*

1. Sel darah merah tidak membesar.
2. Ring terlihat jelas dan halus dan bisa ditemukan beberapa ring dalam satu sel.
3. Beberapa ring bisa mempunyai 2 bintik-bintik *kromatin*.
4. Adanya bentuk marginal / applique.
5. Biasanya tidak terlihat adanya bentuk *schizont* dalam apus darah tepi, kecuali pada infeksi berat.
6. *Gametosit* mempunyai karakteristik berbentuk pisang. Meskipun begitu, *gametosit* biasanya tidak muncul dalam

darah pada minggu pertama sampai minggu keempat masa infeksi.

7. Bintik-bintik Maurer ditemukan pada stadium ring tua.



Gambar. 2. Plasmodium Falciparum (Tropika)

Data mining merupakan suatu rangkaian atau tindakan untuk menemukan hubungan yang memiliki arti melalui pola dan kecenderungan dalam himpunan besar data yang tersimpan dengan menggunakan metode. “Teknik, metode atau algoritma di dalam data mining sangat beragam. Adapun proses yang umumnya dilakukan oleh data mining meliputi deskripsi, prediksi, estimasi, klasifikasi, *clustering* dan asosiasi. Dalam data mining prediksi dan klasifikasi banyak digunakan untuk menganalisis suatu data yang dapat menggambarkan kelas data atau untuk memprediksi data di masa depan.[2]

Algoritma C4.5 merupakan suatu Algoritma yang dapat digunakan untuk melakukan klasifikasi atau pengelompokan pada dataset. Algoritma C4.5 memiliki ide dasar yaitu pembentukan pohon keputusan (*Decision Tree*). Dimana Pohon keputusan termasuk metode klasifikasi dan prediksi yang terkenal karena sangat mudah dipahami dengan bahasa alami sehingga dapat diinterpretasikan dengan cepat. “Pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki

prioritas tertinggi atau dapat disebut memiliki *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai sumbu atribut klasifikasi”. [2]

Dalam sebuah penelitian sudah banyak percobaan yang telah dilakukan dengan memanfaatkan teknik data mining dalam memprediksi penyakit menggunakan berbagai algoritma seperti *Naive Bayes*, *Decision Tree*, SVM, J48 dan lain-lain. Diantaranya telah dapat hasil penelitian [3] dengan menggunakan penerapan algoritma C4.5 dalam memprediksi gangguan autisme pada anak dapat disimpulkan bahwa metode klasifikasi data mining algoritma C4.5 menghasilkan akurasi sebanyak 72% dengan melakukan pengujian metode *cross validation* pada aplikasi *rapidminer*. Adapun penulis [4] menerapkan data mining untuk klasifikasi prediksi penyakit ISPA (Infeksi Saluran Pernapasan Akut) dengan *Algoritma Decision Tree* (ID3) yang menghasilkan atribut penyakit ISPA yang paling berpengaruh pertama adalah gizi sebesar 240.9 kemudian atribut usia sebesar 10.7, maka atribut yang tidak berpengaruh seseorang terkena penyakit ISPA adalah jenis kelamin.

Penelitian ini melakukan prediksi secara diagnostik apakah pasien menderita penyakit malaria dengan menerapkan algoritma C4.5 dan penulis menggunakan teknik data mining untuk klasifikasi stadium penyakit malaria yang disebabkan oleh *Plasmodium Falciparum* (Malaria Tropika).

2. METODOLOGI PENELITIAN

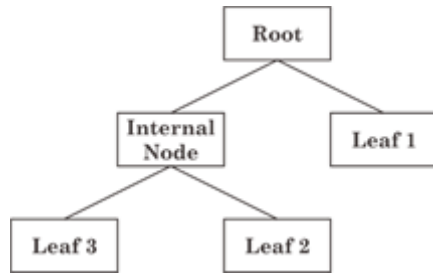
2.1. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data [5]. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster sehingga objek - objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Data mining merupakan bagian dari proses penemuan pengetahuan dari basis *data Knowledge Discovery in* [6].

2.2. Pohon Keputusan (*Decision Tree*)

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Dimana metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang dapat mempresentasikan aturan-aturan [7]. Aturan tersebut dapat dengan mudah dipahami ataupun dijelaskan dengan ulang karena memiliki bahasa alami yang dapat dimengerti. Pohon keputusan juga berguna untuk menjelajahi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. “Pohon keputusan memiliki fleksibilitas (mudah digunakan) yang membuatnya sangat menarik, terutama karena menyajikan keuntungan dari proses pengubahan konsep menjadi gambar yang telah disajikan

sangat mudah dimengerti dimana cabang-cabang dari pohon merangkum klasifikasi” [8].



Gambar 3. Contoh Struktur Pohon Keputusan

2.3. Algoritma C4.5

Algoritma C4.5 merupakan metode yang menjadi pilihan pertama dan sering digunakan dalam pengembangan Data Mining karena kecepatannya dalam proses klasifikasi atau segmentasi dan bersifat prediktif. Algoritma C4.5 salah satu metode untuk membuat Decision Tree berdasarkan training data yang telah disediakan [9]. “Kelebihan algoritma C4.5 adalah metode pohon keputusan yang dapat menghindari munculnya permasalahan dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi keputusan yang dihasilkan”.[10][11][12]

Untuk menghitung nilai *entropy* digunakan rumus 1.

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

Setelah menghitung *entropy* setiap kasus selanjutnya menghitung nilai *gain* untuk pemisah objek dengan rumus 2.

$$Gain(S, A) =$$

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

3. HASIL DAN PEMBAHASAN

3.1. Basis Pengetahuan

Pada basis pengetahuan. Data sampel yang digunakan sebanyak 75 data latih dan 75 data uji, akan dijelaskan tabel dataset malaria tropika, tabel uji malaria tropika dan pohon keputusan.

Row No.	Status	Nama Malaria	Hasil Deteksi
1	Yes	Normal	Normal
2	Yes	Normal	Normal
3	No	Normal	Normal
4	No	Normal	Normal
5	No	Normal	Normal
6	No	Normal	Normal
7	Yes	Normal	Normal
8	Yes	Normal	Normal
9	Yes	Normal	Normal
10	Yes	Normal	Normal
11	Yes	Normal	Normal
12	Yes	Normal	Normal
13	Yes	Normal	Normal
14	Yes	Normal	Normal
15	Yes	Normal	Normal

Gambar 4. Hasil Data Set Malaria Tropika

Pada hasil dataset malaria tropika dapat dilihat pada gambar 4. diatas. Pada kolom berwarna hijau merupakan label. Fungsi dari label adalah sebagai pengarah dalam menentukan stadium atau hasil deteksi pada malaria tropika yaitu: Normal, Stadium 1, Stadium 2, Stadium 3 dan Stadium 4.

3.2. Decision Tree

Pohon Keputusan ini merupakan pengetahuan dari proses belajar Algoritma Decision Tree berdasarkan dataset malaria tropika dengan analisis sebagai berikut:

1. Jika hasil deteksi malaria tropika dinyatakan normal apabila nilainya > 6500 dan ≤ 2500.

2. Jika hasil deteksi malaria tropika dinyatakan stadium 1 apabila nilainya > 17500.
3. Jika hasil deteksi malaria tropika dinyatakan stadium 2 apabila nilainya > 36500 dan ≤ 36500.
4. Untuk hasil deteksi malaria tropika stadium 3 dan 4 disesuaikan dengan kondisi dan tempat tinggal pasien. Karena pada tingkatan stadium ini sangat berbahaya.

Hasil dari Algoritma *Decision Tree* dapat dilihat pada gambar 4.



Gambar 5. Hasil Pohon Keputusan

3.3. Apply Model

Pada kolom dataset pengujian terdapat 2 label yang berwarna hijau dimana ini merupakan perbandingan antara dataset dengan data kalifikasi pada dataset uji dengan hasil analisis sebagai berikut :

1. Jika pada dataset hasilnya yes dan hasil klasifikasinya yes, maka dapat dikatakan hasil klasifikasinya benar.

2. Jika pada *dataset* hasilnya no dan hasil klasifikasinya no, maka dapat dikatakan hasil klasifikasinya benar.
3. Jika pada dataset hasilnya no dan hasil klasifikasinya yes, maka dapat dikatakan hasil klasifikasi pada pohon keputusan tersebut salah. Terdapat pada no.63 dan no.65, dapat dilihat pada gambar 7.

Row No.	Status	prediction(Status)	confidence(Yes)	confidence(No)	No	Nama Malaria	Hasil Deteksi
1	Yes	Yes	1	0	1	Normal	Normal
2	Yes	Yes	1	0	2	Normal	Normal
3	No	No	0	1	3	Normal	Normal
4	No	No	0	1	4	Normal	Normal
5	No	No	0	1	5	Normal	Normal
6	No	No	0	1	6	Normal	Normal
7	Yes	Yes	1	0	7	Normal	Normal
8	Yes	Yes	1	0	8	Normal	Normal
9	Yes	Yes	1	0	9	Normal	Normal
10	Yes	Yes	1	0	10	Normal	Normal
11	Yes	Yes	1	0	11	Normal	Normal
12	Yes	Yes	1	0	12	Normal	Normal
13	Yes	Yes	1	0	13	Normal	Normal
14	Yes	Yes	1	0	14	Normal	Normal

Gambar 6. Hasil dataset pengujian malaria tropika menggunakan *apply model*

Row No.	Status	prediction(Status)	confidence(Yes)	confidence(No)	No	Nama Malaria	Hasil Deteksi
62	Yes	Yes	0.600	0.400	62	Tropika Stadium 4	Tropika Stadium 4
63	No	Yes	0.600	0.400	63	Tropika Stadium 4	Tropika Stadium 4
64	Yes	Yes	0.600	0.400	64	Tropika Stadium 4	Tropika Stadium 4
65	No	Yes	0.600	0.400	65	Tropika Stadium 4	Tropika Stadium 4
66	Yes	Yes	1	0	66	Tropika Stadium 4	Tropika Stadium 4
67	Yes	Yes	1	0	67	Tropika Stadium 4	Tropika Stadium 4
68	Yes	Yes	1	0	68	Tropika Stadium 4	Tropika Stadium 4
69	Yes	Yes	1	0	69	Tropika Stadium 4	Tropika Stadium 4
70	Yes	Yes	1	0	70	Tropika Stadium 4	Tropika Stadium 4
71	Yes	Yes	1	0	71	Tropika Stadium 4	Tropika Stadium 4
72	Yes	Yes	1	0	72	Tropika Stadium 4	Tropika Stadium 4
73	Yes	Yes	1	0	73	Tropika Stadium 4	Tropika Stadium 4
74	Yes	Yes	1	0	74	Tropika Stadium 4	Tropika Stadium 4
75	Yes	Yes	1	0	75	Tropika Stadium 4	Tropika Stadium 4

Gambar 7. Hasil dataset pengujian malaria tropika menggunakan *apply model*

Pengujian terhadap hasil perhitungan menggunakan *software RapidMiner* dilakukan melalui beberapa proses hingga diperoleh hasil pemrosesan data menggunakan *software rapidminer* didapat hasil akurasi penerapan *algoritma C4.5* yaitu 82,67%. Untuk melihat nilai akurasi dapat dilihat pada gambar 8.

	true Yes	true No	class precision
pred. Yes	47	10	82.46%
pred. No	3	15	83.33%
class recall	94.00%	60.00%	

Gambar 8. Nilai Akurasi *Algoritma C4.5*

```

PerformanceVector:
accuracy: 82.67%
ConfusionMatrix:
True:  Yes  No
Yes:   47  10
No:    3   15
classification_error: 17.33%
ConfusionMatrix:
True:  Yes  No
Yes:   47  10
No:    3   15
    
```

Gambar 9. Nilai *Performance Vector* *Algoritma C4.5*

Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 82,67%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%. Dimana Jumlah benar prediksi *yes* sebanyak 47 data, salah prediksi *no* sebanyak 10 data, jumlah benar prediksi *no* sebanyak 3 data sedangkan salah prediksi *no* sebanyak 15 data. Prediksi *yes* dari seluruh data menghasilkan 82,67% seseorang terdeteksi penyakit malaria tropika.

4. KESIMPULAN

Implementasi Data Mining Untuk Prediksi Penyakit Malaria Tropika Dengan *Algoritma C4.5* dapat membantu penelitian ini untuk memprediksi seseorang terinfeksi penyakit malaria tropika dengan tingkat stadium tertentu sejak dini. Pada penelitian ini menghasilkan jumlah prediksi *yes* sebanyak 47 data dan jumlah prediksi *no* sebanyak 15. Prediksi *yes* dari seluruh data menghasilkan 82,67%. Berdasarkan pengolahan data menggunakan *Software Rapidminer* didapat nilai akurasi sistem sebesar 82,67%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%.

DAFTAR PUSTAKA

- [1] N. A. Banyal, S. Surianti, and A. R. Dayat, "Klasifikasi *Citra Plasmodium* Penyebab Penyakit Malaria dalam Sel Darah Merah Manusia dengan Menggunakan Metode *multi class support vector machine (svm)*. *ILKOM Jurnal Ilmiah* 8 (2), pp.111-118, 2016.
- [2] Sanni Ucha Putri, Eka Irawan, Fitri Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan *Algoritma C4.5*", *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*. Vol.2, No. 1, pp. 39-46, Januari, 2021.
- [3] B. Sugara, D. Widyatmoko, B. S. Prakoso, And D. M. Saputro, "Penerapan *Algoritma C4.5* Untuk Deteksi Dini Autisme Pada Anak," *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)* 2018, Pp. 87-96.
- [4] A. E. Pramadhani, T. Setiadi, "Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernapasan Akut) Dengan *Algoritma Decision Tree*. *Jurnal Sarjana Teknik Informatika* Vol.2 No.1, pp: 160-168, Februari 2014
- [5] Ade Purwanto, Aji Primajaya, Apriade Voutama, "Penerapan *Algoritma C4.5*

- dalam Prediksi Potensi Tingkat Kasus Pneumonia di Kabupaten Karawang”, *Jurnal Sistem dan Teknologi Informasi Universitas Singaperbangsa Karawang*. Vol. 08, No.4, Oktober 2020.
- [6] S. R. Ningsih, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, And A. Wanto, “Analisis K-Medoids Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi,”. *SENARIS*. pp.721, 2019, Doi: 10.30645/Senaris.V1i0.78.
- [7] Sunanto, Ghazi Falah., “Perapan *Algoritma C4.5* untuk Membuat Model Prediksi Pasien yang Mengidap Penyakit Diabetes,” *Jurnal Teknologi dan Sistem Informasi Univra*. Vol. 7 No. 2, pp. 208-216, Juli 2022.
- [8] A. S. Sunge Et Al., “Prediksi Kompetensi Karyawan Menggunakan *Algoritma C4.5* (Studi Kasus: PT Hankook Tire Indonesia),” *SENTIKA*, 2018, Pp. 23–24.
- [9] I.G.A Mahardika Pratama, L.G Astuti, I.M Widiartha, dkk., “Diagnosis Penyakit Ginjal Kronis dengan *Algoritma C4.5*, *K-Means* dan *BPSO*,” *Jurnal Elektronik Ilmu Komputer Udayana*. Volume 10 No 4, May 2022.
- [10] A. Muzakir And R. A. Wulandari, “Model Data Mining Sebagai Prediksi Penyakit Hipertensi Kehamilan Dengan Teknik Decision Tree,” *Sci. J. Informatics*, Vol. 3, No. 1, Pp. 19–26, 2016, Doi: 10.15294/Sji.V3i1.4610
- [11] Abdul Rohman., “Penerapan *Algoritma C4.5* Berbasis Adaboos untuk Prediksi Penyakit Jantung,” *Majalah Ilmiah Universitas Pandanaran*. Vol 11, No. 26, 2013.
- [12] Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. Klasifikasi Dengan Pohon Keputusan Berbasis *Algoritme C4.5*. *PRISMA, Prosiding Seminar Nasional Matematika 3*, 2020, pp. 64–71.