

Penerapan Metode Algoritma C4.5 Dalam Klasifikasi Diagnosa Penyakit Umum Menggunakan WEKA

Rachman Komarudin¹, Puspa Vicenna Barulah Yudha², Yana Iqbal Maulana³, Nurul Afni⁴, Agus Salim⁵, Irmawati Carolina⁶

Universitas Nusa Mandiri^{1,2}, Universitas Bina Sarana Informatika^{3,4,5,6}

rachman.rck@nusamandiri.ac.id¹, puspavicenna95@gmail.com², yana.yim@bsi.ac.id³,
nurul.nrf@bsi.ac.id⁴, agus.salim@bsi.ac.id⁵, irmawati.imc@bsi.ac.id⁶,

Abstrak - Kurangnya pengetahuan terhadap penyakit seperti DBD, Tifus dan ISPA yang disebabkan oleh virus perlu adanya tindakan atau penanganan secara cepat. Algoritma C4.5 merupakan klasifikasi yang mendapatkan hasil pohon keputusan yang mudah diinterpretasikan. Penelitian ini bertujuan untuk mengetahui gejala-gejala penyakit yang disebabkan oleh virus seperti penyakit DBD (Demam Berdarah Dengue), Tifus dan ISPA (Infeksi Saluran Pernafasan Atas) menggunakan 4 parameter yaitu jenis kelamin, usia, keluhan dan suhu tubuh. Penelitian adalah mengklasifikasikan penyakit DBD, Tifus dan ISPA. Metode penelitian yang digunakan adalah data kuantitatif. Penelitian ini datanya didapat dari data sekunder. Metode yang digunakan yaitu Algoritma C4.5. Software yang dipakai untuk melihat hasil penelitian yang dibahas yaitu: software Weka 3.9. Hasil pembahasan pada software weka 3.9, tingkat akurasi data sebesar 64.2857 % atau 27 data dan dari tingkat error sekitar 35.7134% atau 15 data. Dari data 42 jumlah kasus pasien puskesmas puger kota bondowoso yang terkena penyakit seperti DBD, Tifus dan ISPA sehingga dapat disimpulkan bahwa yang di implementasikan kedalam software weka 3.9 dapat membantu puskesmas puger dalam diagnosa penyakit umum.

Kata Kunci – Algoritma C4.5, Data Mining, Penyakit Umum, Klasifikasi).

Abstract - Lack of knowledge about diseases such as dengue fever, typhoid and ARI caused by viruses requires immediate action or treatment. The C4.5 algorithm is a classification that gets decision tree results that are easy to interpret. This study aims to determine the symptoms of diseases caused by viruses such as DHF (Dengue Hemorrhagic Fever), Typhus and ARI (Upper Respiratory Tract Infection) using 4 parameters, namely gender, age, complaints and body temperature. The research is to classify the diseases of DHF, Typhus and ARI. The research method used is quantitative data. This research data obtained from secondary data. The method used is the C4.5 algorithm. The software used to view the research results discussed are: Weka 3.9 software. The results of the discussion on the Weka 3.9 software, the data accuracy rate is 64.2857 % or 27 data and the error rate is around 35.7134% or 15 data. From the data of 42 cases of patients at Puger Public Health Center in Bondowoso City who were affected by diseases such as DHF, Typhus and ISPA, it can be concluded that what is implemented into the Weka 3.9 software can help Puger Health Center in diagnosing common diseases..

Keywords: C4.5 Algorithm, Data Mining, General Diseases, Classification).

I. PENDAHULUAN

Masalah kesehatan merupakan suatu masalah yang sangat kompleks, yang sangat berkaitan dengan masalah-masalah lain di luar kesehatan. Banyak faktor yang mempengaruhi kesehatan, diantaranya yaitu pengetahuan dan sikap masyarakat dalam merespon suatu penyakit. Salah satu masalah kesehatan yang banyak dialami oleh sebagian besar masyarakat di Indonesia terutama wilayah Desa Puger adalah masalah yang menyerang sistem kekebalan tubuh yang disebabkan oleh virus, yaitu DBD (Demam Berdarah Dengue), Tifus, dan ISPA (Infeksi Saluran Pernafasan Atas). Penyakit yang disebabkan oleh virus biasanya dipengaruhi oleh banyak faktor. Diantaranya, faktor lingkungan dan kebiasaan hidup sehari-hari. DBD (Demam Berdarah Dengue) atau dengan nama lain DHF (Dengue Haemorrhagic Fever) merupakan salah satu

dari beberapa penyakit menular yang menjadi masalah kesehatan di dunia terutama negara Indonesia. Sejak tahun 1968 di Surabaya, muncul masalah DBD tercatat 334.685 kasus DBD. Penderita yang meninggal 3.092 orang Selama 9 tahun. DBD biasanya terjadi pada awal musim hujan yang sering terjadi. (Fakhriadi et al., 2015). salmonella typhimurium menjadi penyebab penyakit sistemik akut atau demam tifoid. 16 juta kasus setiap tahunnya diperkirakan Insidensi demam tifoid terjadi. dalam setiap tahun 600.000 orang meninggal karena penyakit ini. penyakit tifus atau demam tifoid merupakan penyakit endemik di Indonesia dan menjadi masalah kesehatan yang serius. Insidensi rata-rata mencapai 650 kasus per 100.000 penduduk di Indonesia, dengan mortalitas rata-rata bervariasi dari 3,1 - 10,4% (Ihsan & Wajhillah, 2015) .

Infeksi Saluran Pernapasan Akut (ISPA) atau yang kita sering sebut ISPA merupakan penyakit yang umum terjadi pada semua kategori umur, terutama pada anakanak. Ada beberapa variasi Tingkat keparahan penyakit ISPA dari faringitis ringan, pneumonia bahkan sampai pada kematian. Pada tingkat morbiditas dan mortalitas Penyakit ISPA sangat berpengaruh besar, karena salah satu penyakit infeksi menular di dunia (Br. Tarigan et al., 2017). Berdasarkan pengambilan data yang sudah ada dan latar belakang tersebut, penulis akan mencoba mengklasifikasikan penyakit Thypoid, Fever, DHF, Infeksi Saluran Penapasan Atas (ISPA) dengan menerapkan metode Algoritma C4.5.

II. METODOLOGI PENELITIAN

1. Metode Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma Decision Tree (Pohon Keputusan). Algoritma ini mempunyai input berupa training samples dan samples. Training sample berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan parameter dalam melakukan klasifikasi data menggunakan field-field data atau sample (Ihsan & Wajhillah, 2015). Adapun Tahapan-tahapan yang digunakan untuk menyelesaikan Algoritma C4.5 dapat dilihat pada Gambar 1:



Sumber : Data Penelitian

Gambar 1. Algoritma Penyelesaian Algoritma C4.5

Dalam langkah pertama untuk mencari nilai Entropy dan Gain menggunakan rumus:

1. Entropy

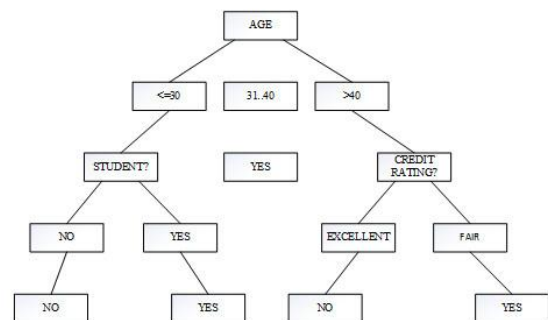
$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \dots (1)$$

2. Gain

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

2. Pohon keputusan

Salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia adalah Pohon keputusan. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple merupakan manfaat utama dari penggunaan pohon keputusan, sehingga dalam pengambilan keputusan lebih menginterpretasikan solusi dari permasalahan. Selain hal tersebut Pohon keputusan dapat digunakan untuk mengeksplorasi data, menemukan hubungan tersembunyi antara jumlah calon variabel input dengan sebuah variabel target (Wulandari, 2017).



Sumber : Data Penelitian

Gambar 2. Model Pohon Keputusan

3. Tahapan Penelitian

a. Melakukan Studi Pustaka

Mempelajari berbagai sumber pustaka yang akan digunakan sebagai kajian teori dalam penelitian ini yang antara lain berasal dari buku dan jurnal penelitian.

b. Menentukan Metode Data Mining

Metode algoritma C4.5 menjadi pilihan penulis, dalam penelitian. Karena algoritma C4.5 dapat melakukan transformasi atau mengubah nilai atribut ke dalam bentuk data yang sesuai agar dapat diproses.

c. Melakukan Tahapan Metode Klasifikasi

Tahapan menggunakan metode klasifikasi menggunakan algoritma C4.5 untuk memperoleh hasil penelitian antara lain:

1). Tahap Pengumpulan Data

Tahap ini peneliti melakukan pengumpulan data yang dilakukan dengan benar. Karena tahap ini sangat berpengaruh terhadap hasil penelitian. Peneliti mendapatkan data dari database UPTD Puskesmas Puger Bondowoso.

2). Tahap Seleksi Data

Tahap ini dilakukan seleksi terhadap data database excel. Karena data yang diperoleh tidak semua digunakan, dipilih sesuai atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan seleksi data sehingga menjadi dataset. Sebagai contoh pada penelitian ini atribut atau variabel yang dipilih adalah jenis kelamin, usia, keluhan, dan suhu

tubuh. Pada tahap seleksi untuk data yang null, data tidak valid, dan data yang ganda akan dihilangkan. Karena data yang kosong ataupun tidak valid akan berpengaruh pada hasil yang diperoleh. Kemudian data akan diolah menggunakan Microsoft Office Excel 2010.

3). Tahap Pengolahan Data

Setelah semua data yang diperlukan telah dipilih, maka tahap penelitian selanjutnya yaitu pengolahan data. Proses mengubah nilai atribut data ke dalam bentuk data yang sesuai atau transformasi data akan dilakukan akan dilakukan pada tahap ini agar dapat diproses menggunakan algoritma C4.5. Dengan menggunakan rumus 1 :

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

Dan rumus yang ke 2 :

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

4). Tahap Pengujian Data

Pada tahap pengujian hasil akan dilakukan pengujian data baik secara manual dengan menggunakan algoritma C4.5 dan menggunakan software WEKA (Waikato Environment for Knowledge Analysis).

5). Kesimpulan Penelitian

Kesimpulan merupakan langkah terakhir dari sebuah penelitian yang diambil berdasarkan analisa data dan diperiksa sesuai maksud dan tujuan peneliti sehingga menghasilkan jawaban terhadap rumusan masalah serta memberikan informasi mengenai solusi masalah yang bermanfaat untuk klasifikasi penyakit di UPT Puskesmas Pujer Bondowoso.

III. HASIL DAN PEMBAHASAN

Dari data sampel yang sudah di dapatkan sebanyak 42 sampel keseluruhan penyakit maka akan di hitung jumlah kasus dari setiap data dengan hasil rekapitulasi pada Tabel 1 Hasil Rekapitulasi.

Menghitung jumlah kasus, yaitu jumlah kasus untuk keputusan "Typhoid Fever", jumlah kasus untuk keputusan "Demam berdarah (DHF)", jumlah kasus untuk keputusan "Infeksi Saluran Pernafasan Atas (ISPA)", dan perhitungan nilai entropy dari semua kasus yang dibagi berdasarkan kriteria jenis kelamin dengan atribut laki-laki dan perempuan, kriteria usia dengan atribut anak, remaja, dan dewasa, kriteria keluhan memiliki atribut sakit kepala, batuk berdarah, dan Bb turun, dan pada kriteria suhu memiliki atribut tinggi, sedang, dan rendah. Menggunakan rumus Entropy dan Gain.

A. Nilai Entrophy

a. Entrophy Total

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= \left(-\frac{20}{42} * \log_2 \left(\frac{20}{42} \right) \right) + \left(-\frac{11}{42} * \log_2 \left(\frac{11}{42} \right) \right) + \\ &\quad \left(-\frac{11}{42} * \log_2 \left(\frac{11}{42} \right) \right) \\ &= 0.509709 + 0.506231 + 0.506231 \\ &= 1.522173 \end{aligned}$$

b. Entrophy Jenis Kelamin

1. Laki-laki

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= \left(-\frac{12}{24} * \log_2 \left(\frac{12}{24} \right) \right) + \left(-\frac{6}{24} * \log_2 \left(\frac{6}{24} \right) \right) + \\ &\quad \left(-\frac{6}{24} * \log_2 \left(\frac{6}{24} \right) \right) \end{aligned}$$

$$\begin{aligned} &= 0.5 + 0.5 + 0.5 \\ &= 1.5 \end{aligned}$$

2. Perempuan

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= \left(-\frac{8}{18} * \log_2 \left(\frac{8}{18} \right) \right) + \left(-\frac{5}{18} * \log_2 \left(\frac{5}{18} \right) \right) + \\ &\quad \left(-\frac{5}{18} * \log_2 \left(\frac{5}{18} \right) \right) \\ &= 0.519966 + 0.513332 + 0.513332 \\ &= 1.546632 \end{aligned}$$

c. Entrophy Usia

1. Anak

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6} \right) \right) + \left(-\frac{3}{6} * \log_2 \left(\frac{3}{6} \right) \right) + \\ &\quad \left(-\frac{1}{6} * \log_2 \left(\frac{1}{6} \right) \right) \\ &= 0.528320 + 0.5 + 0.430827 \\ &= 1.459148 \end{aligned}$$

2. Remaja

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= \left(-\frac{5}{10} * \log_2 \left(\frac{5}{10} \right) \right) + \left(-\frac{3}{10} * \log_2 \left(\frac{3}{10} \right) \right) + \\ &\quad \left(-\frac{2}{10} * \log_2 \left(\frac{2}{10} \right) \right) \\ &= 0.5 + 0.521089 + 0.464385 \\ &= 1.485475 \end{aligned}$$

3. Dewasa

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{13}{26} \cdot \log_2 \left(\frac{13}{26}\right)\right) + \left(-\frac{5}{26} \cdot \log_2 \left(\frac{5}{26}\right)\right) + \\
 &\quad \left(-\frac{8}{26} \cdot \log_2 \left(\frac{8}{26}\right)\right) \\
 &= 0.5 + 0.457406 + 0.523212 \\
 &= 1.480618
 \end{aligned}$$

d. *Entropy* Keluhan

1. Sakit Kepala

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{6}{15} \cdot \log_2 \left(\frac{6}{15}\right)\right) + \left(-\frac{4}{15} \cdot \log_2 \left(\frac{4}{15}\right)\right) + \\
 &\quad \left(-\frac{5}{15} \cdot \log_2 \left(\frac{5}{15}\right)\right) \\
 &= 0.528771 + 0.508504 + 0.528320 \\
 &= 1.565596
 \end{aligned}$$

2. Batuk Berdahak

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{5}{12} \cdot \log_2 \left(\frac{5}{12}\right)\right) + \left(-\frac{3}{12} \cdot \log_2 \left(\frac{3}{12}\right)\right) + \\
 &\quad \left(-\frac{4}{12} \cdot \log_2 \left(\frac{4}{12}\right)\right) \\
 &= 0.526264 + 0.5 + 0.528320 \\
 &= 1.554585
 \end{aligned}$$

3. Bb Turun

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{7}{15} \cdot \log_2 \left(\frac{7}{15}\right)\right) + \left(-\frac{3}{15} \cdot \log_2 \left(\frac{3}{15}\right)\right) + \\
 &\quad \left(-\frac{5}{15} \cdot \log_2 \left(\frac{5}{15}\right)\right) \\
 &= 0.513116 + 0.464385 + 0.528320 \\
 &= 1.505823
 \end{aligned}$$

e. *Entropy* Suhu

1. Tinggi

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{8}{14} \cdot \log_2 \left(\frac{8}{14}\right)\right) + \left(-\frac{2}{14} \cdot \log_2 \left(\frac{2}{14}\right)\right) + \\
 &\quad \left(-\frac{4}{14} \cdot \log_2 \left(\frac{4}{14}\right)\right) \\
 &= 0.461345 + 0.401050 + 0.516387 \\
 &= 1.378783
 \end{aligned}$$

2. Sedang

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{7}{18} \cdot \log_2 \left(\frac{7}{18}\right)\right) + \left(-\frac{6}{18} \cdot \log_2 \left(\frac{6}{18}\right)\right) + \\
 &\quad \left(-\frac{5}{18} \cdot \log_2 \left(\frac{5}{18}\right)\right) \\
 &= 0.529888 + 0.528320 + 0.513332 \\
 &= 1.571542
 \end{aligned}$$

3. Rendah

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\
 &= \left(-\frac{5}{10} \cdot \log_2 \left(\frac{5}{10}\right)\right) + \left(-\frac{3}{10} \cdot \log_2 \left(\frac{3}{10}\right)\right) + \\
 &\quad \left(-\frac{2}{10} \cdot \log_2 \left(\frac{2}{10}\right)\right) \\
 &= 0.5 + 0.521089 + 0.464385 \\
 &= 1.485475
 \end{aligned}$$

B. Nilai *Gain*1. *Gain* Jenis Kelamin

$$\begin{aligned}
 Gain(\text{total}, \text{jk}) &= 1.522173 - \sum_{i=0}^n \left(\left(\frac{24}{42} \right) * 1.5 \right) + \\
 &\quad \left(\left(\frac{18}{42} \right) * 1.546632 \right) \\
 &= 1.327872
 \end{aligned}$$

2. *Gain* Usia

$$\begin{aligned}
 Gain(\text{total}, \text{usia}) &= 1.522173 - \sum_{i=0}^n \left(\left(\frac{6}{42} \right) * 1.459148 \right) + \\
 &\quad \left(\left(\frac{10}{42} \right) * 1.485475 \right) + \left(\left(\frac{26}{42} \right) * 1.480618 \right) \\
 &= 2.583981
 \end{aligned}$$

3. *Gain* Keluhan

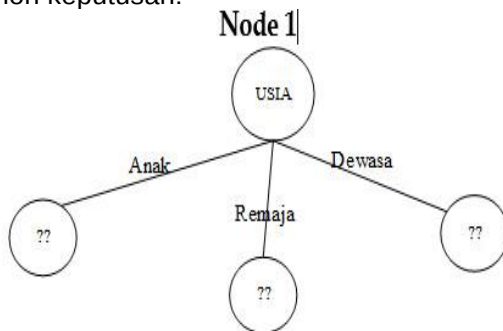
$$\begin{aligned}
 Gain(\text{total}, \text{keluhan}) &= 1.522173 - \sum_{i=0}^n \left(\left(\frac{15}{42} \right) * 1.565596 \right) + \\
 &\quad \left(\left(\frac{12}{42} \right) * 1.554585 \right) + \left(\left(\frac{15}{42} \right) * 1.505823 \right) \\
 &= 1.944993
 \end{aligned}$$

4. *Gain* Suhu

$$\begin{aligned}
 Gain(\text{total}, \text{suhu}) &= 1.522173 - \sum_{i=0}^n \left(\left(\frac{14}{42} \right) * 1.378783 \right) + \\
 &\quad \left(\left(\frac{18}{42} \right) * 1.571542 \right) + \left(\left(\frac{10}{42} \right) * 1.485475 \right) \\
 &= 2.089781
 \end{aligned}$$

Setelah nilai *gain* didapat dari hasil perhitungan masing-masing nilai *entropy* per kriteria, Hasil perhitungan dapat dilihat pada tabel 2 Hasil Rekapitulasi *Entropy* dan *Gain*.

Pada tabel 2 Hasil Rekapitulasi Entropy dan Gain menunjukkan bahwa kriteria “Usia” memiliki nilai gain tertinggi. Sehingga untuk kriteria usia akan menjadi node pertama pada pohon keputusan. Gambar 3 menunjukkan pohon keputusan:

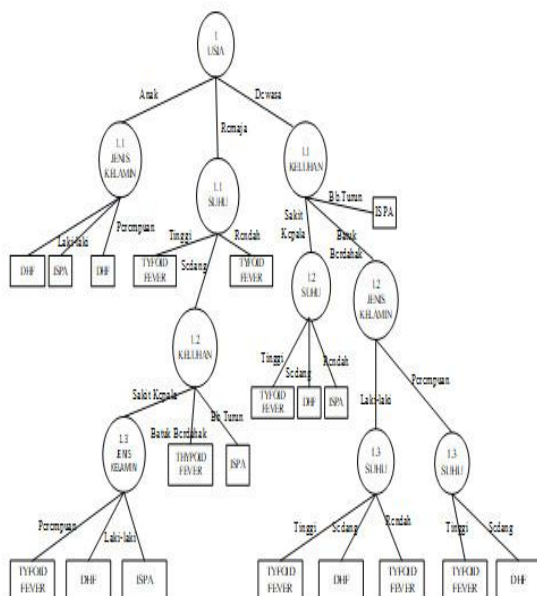


Sumber : Data Penelitian
Gambar 3. Pohon Keputusan

Setelah mendapatkan hasil pohon keputusan seperti diatas kita tahu yg menjadi akar pertama dari pohon tersebut. Dan untuk mengetahui hasil keseluruhan dari pohon tersebut kita harus melakukan perhitungan ulang untuk mendapatkan hasil node berikutnya.

Dengan cara yang sama kita ulangi perhitungan nilai entropy dan gain dari setiap atribut yang didapatkan seperti pada atribut Anak, Remaja, Dewasa. Kita harus mencari nilai Gain yang tertinggi sehingga bisa dihitung nilainya.

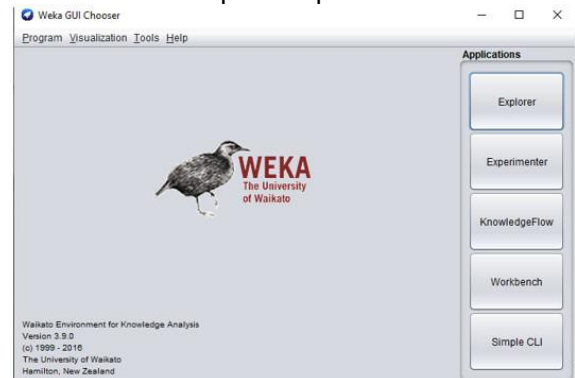
Setelah melakukan perhitungan keseluruhan maka akan didapat hasil pohon keputusan keseluruhan seperti ditunjukkan pada Gambar 4:



Sumber : Data Penelitian
Gambar 4. Pohon Keputusan

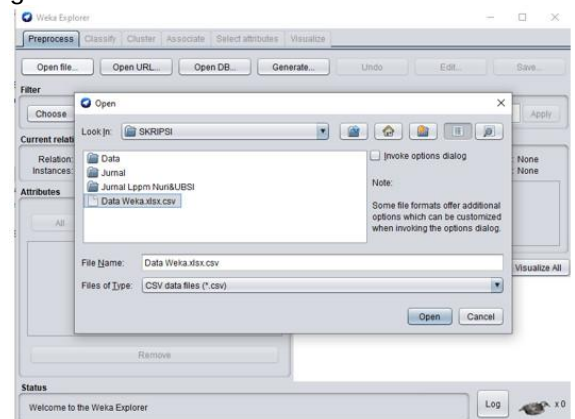
Proses Data Menggunakan Aplikasi Weka 3.9

Pertama-tama klik aplikasi weka 3.9 maka akan muncul tampilan seperti :

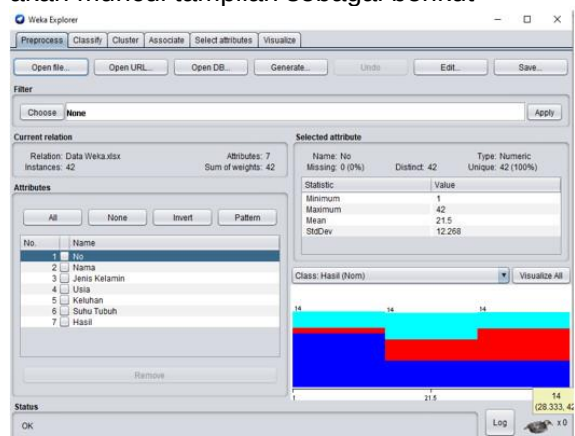


Sumber: software weka 3.9
Gambar 5 Tampilan Interface Pada Weka 3.9

Kemudian klik Explorer pada command button diatas, lalu klik open file yang ada di pojok kiri atas. Cari file yang akan diuji validasinya dengan format file .csv seperti gambar dibawah.

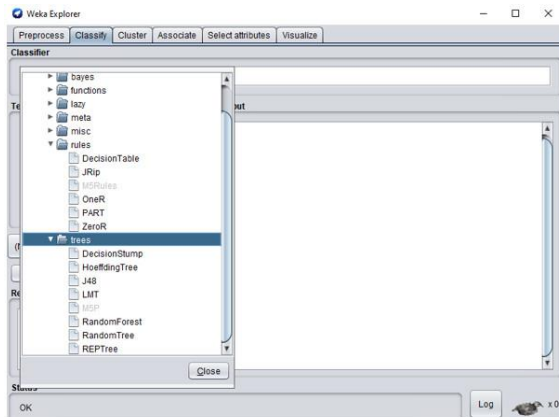


Sumber: software weka 3.9
Gambar 6 Tampilan Insert File
Langkah selanjutnya yaitu klik open maka akan muncul tampilan sebagai berikut



Sumber: software weka 3.9
Gambar 7 Tampilan Proses Pengujian
Jangan lupa pada command button class, pilih class yang sudah ditentukan

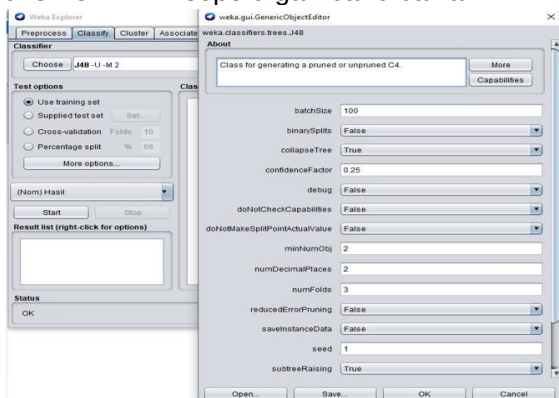
sebelumnya. Kemudian klik button classify yang ada dipojok kiri atas. Lalu klik Choose, maka akan muncul tampilan seperti di bawah ini.



Sumber: software weka 3.9

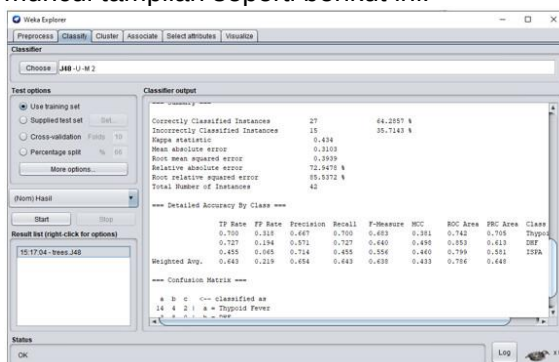
Gambar 8 Tampilan Pemilihan Metode Algoritma C4.5

Pilih trees lalu klik J48 pada aplikasi weka ini metode algoritma C4.5 adalah J48. Langkah berikutnya klik use training set yang artinya semua data yang diinput tadi akan diujikan semuanya. Pada tulisan J48-C 0.25 – M 2 yang ada disamping choose lalu pilih true pada unpruned. Kemudian klik Ok maka tulisan J48-C 0.25 –M 2 berubah menjadi J48 –U –M 2 seperti gambar dibawah ini:



Sumber: software weka 3.9

Gambar 9 Tampilan Pemilihan Unpruned Klik Ok kemudian Klik Start maka akan muncul tampilan seperti berikut ini:

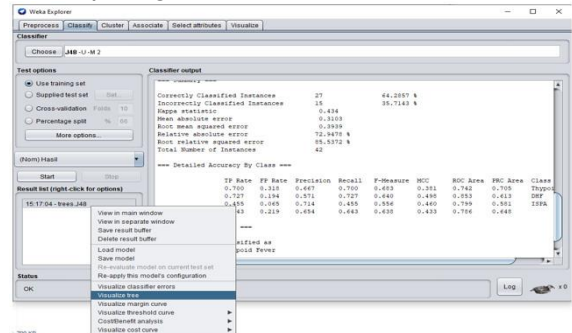


Sumber: software weka 3.9

Gambar 10 Tampilan Hasil Analisa Classify

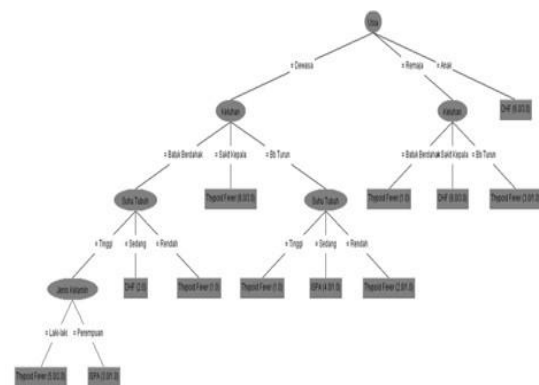
Dimana pada script di samping bahwa tingkat pengklasifikasian yang benar adalah 64,2857% sedangkan tingkat kesalahannya adalah sekitar 35,7143% .

Untuk melihat hasil pohon keputusannya pada kolom results list klik kanan pada trees J48 seperti gambar dibawah ini:



Sumber: software weka 3.9

Gambar 11 Tampilan Proses Visualize Tree Pada visualize tree kemudian akan muncul pohon keputusan dari hasil classifikasi data yang telah dilakukan berikut hasil pohon keputusan



Sumber : Data Penelitian software weka 3.9

Gambar 12. Hasil Pohon Keputusan Aplikasi Weka

Gambar diatas merupakan hasil pengujian menggunakan software Weka 3.9. Dapat dilihat bahwa dari hasil pengujian software kriteria “usia” yang menjadi akar pertama dari pohon keputusan. Pada node akar “usia” yang memiliki hasil diagnosa adalah atribut anak yaitu DHF. Pada atribut remaja kriteria “keluhan” memiliki hasil diagnosa di atribut batuk berdarah dan Bb.Turun diagnosanya typhoid fever dan atribut sakit kepala diagnosanya DHF.

Sedangkan pada atribut dewasa “keluhan” menjadi kriteria. Pada tahap selanjutnya di kriteria “keluhan” atribut sakit kepala sudah memiliki diagnosa typhoid fever, sedangkan atribut Bb.Turun pada kriteria “suhu” atribut tinggi diagnosanya typhoid fever, atribut sedang diagnosanya ISPA, dan atribut rendah diagnosanya typhoid fever.

Kriteria “keluhan” pada atribut batuk berdahak “suhu tubuh” yang memiliki diagnosa pasti yaitu atribut sedang diagnosanya DHF, atribut rendah diagnosanya typhoid fever, dan atribut tinggi memiliki kriteria “jenis kelamin” yang hasil diagnosanya pasti, pada atribut laki-laki

diagnosanya typhoid fever, dan atribut perempuan diagnosanya ISPA.

Dari hasil keseluruhan didapat tingkat akurasi kebenaran sebesar 64.2857% atau 27 data sedangkan tingkat error sekitar 35.7134% atau 15 data.

Tabel 1. Hasil Rekapitulasi

Node	Jumlah kasus (S)	Typoid Fever	Demam berdarah (DHF)	Infeksi Saluran Pernafasan Atas (ISPA)	Entropy	Gain
1	42	20	11	11		
Total						
Jenis Kelamin						
Laki – Laki	24	12	6	6		
Perempuan	18	8	5	5		
Usia						
Anak	6	2	3	1		
Remaja	10	5	3	2		
Dewasa	26	13	5	8		
Keluhan						
Sakit Kepala	15	6	4	5		
Batuk Berdahak	12	5	3	4		
Bb Turun	15	7	3	5		
Suhu						
Tinggi	14	8	2	4		
Sedang	18	7	6	5		
Rendah	10	5	3	2		

sumber : Data Penelitian

Tabel 2. Hasil Rekapitulasi Entropy dan Gain

Node	Jumlah kasus (S)	Typoid Fever	Demam berdarah (DHF)	Infeksi Saluran Pernafasan Atas (ISPA)	Entropy	Gain
1	42	20	11	11	1.522173	
Total						
Jenis Kelamin						1.327872
Laki - Laki	24	12	6	6	1.5	
Perempuan	18	8	5	5	1.546632	
Usia						2.583981
Anak	6	2	3	1	1.459148	
Remaja	10	5	3	2	1.485475	
Dewasa	26	13	5	8	1.480618	
Keluhan						1.944993
Sakit Kepala	15	6	4	5	1.565596	
Batuk Berdahak	12	5	3	4	1.554585	
Bb Turun	15	7	3	5	1.505823	
Suhu						2.089781
Tinggi	14	8	2	4	1.378783	
Sedang	18	7	6	5	1.571542	
Rendah	10	5	3	2	1.485475	

sumber : Data Penelitian

IV. KESIMPULAN

Berdasarkan uraian dan analisi yang telah dipaparkan pada bab sebelumnya dan hasil penelitian dari data sampel yang diambil dengan mengklasifikasikan menggunakan metode Algoritma C4.5 (Decision Tree), penulis menarik kesimpulan sebagai berikut :

1. Pada pengujian manual pada node akar 1.1 masing-masing kriteria memiliki hasil diagnosa tetapi pada kriteria suhu dan kriteria keluhan ada node yang belum memiliki diagnosa yang tepat sehingga harus dihitung lagi nilai entropy dan gainnya dan mendapat diagnosa yang tepat. Sama seperti pada node 1.1 dan node 1.2 juga masih ada kriteria yang belum mendapatkan diagnosa yang tepat. Sehingga melakukan perhitungan ulang pada nilai entropy dan gain yang masih memiliki nilai. Dan pada node 1.3 (terakhir) sudah memiliki hasil diagnosa yang pasti di setiap kriteria.

2. Pada pengujian software weka 3.9 dari hasil pengujian software kriteria “usia” yang menjadi akar pertama dari pohon keputusan. Pada node akar “usia” yang memiliki hasil diagnosa adalah atribut anak yaitu DHF. Pada atribut remaja kriteria “keluhan” memiliki hasil diagnosa di atribut batuk berdarah dan Bb.Turun diagnosanya typhoid fever dan atribut sakit kepala diagnosanya DHF. Sedangkan pada atribut dewasa “keluhan” menjadi kriteria. Pada tahap selanjutnya di kriteria “keluhan” atribut sakit kepala sudah memiliki diagnosa typhoid fever, sedangkan atribut Bb.Turun pada kriteria “suhu” atribut tinggi diagnosanya typhoid fever, atribut sedang diagnosanya ISPA, dan atribut rendah diagnosanya typhoid fever. Kriteria “keluhan” pada atribut batuk berdarah “suhu tubuh” yang memiliki diagnosa pasti yaitu atribut sedang diagnosanya DHF, atribut rendah diagnosanya typhoid fever, dan atribut tinggi memiliki kriteria “jenis kelamin” yang hasil diagnosanya pasti, pada atribut laki-laki diagnosanya typhoid fever, dan atribut perempuan diagnosanya ISPA.

3. Dari hasil pengujian manual dan hasil pengujian pada aplikasi software weka 3.9

dengan tingkat akurasi data sebesar 64.2857% atau 27 data, dan tingkat error sekitar 35.7134% atau 15 data pada aplikasi hasi node akar pertama hasilnya sama pada kriteria usia.

4. Dari hasil data 42 sampel yang terbentuk pada pohon keputusan dengan pengujian secara manual dan aplikasi software weka 3.9 dapat di informasikan bahwa kriteria usia adalah atribut yang paling berpengaruh dalam mengklasifikasikan penyakit umum berdasarkan data yang di proses dengan melihat kriteria jenis kelamin, keluhan dan suhu.

Sehingga penulis menarik kesimpulan bahwa adanya validasi (pembuktian) data dari hasil pengujian manual maupun hasil pengujian aplikasi software weka 3.9, dikarenakan hasil pohon keputusannya tidak sama.)

V. REFERENSI

- Br. Tarigan, D. M., Dr. Rini, D. P. M. ., & Puspita, V. (2017). Perancangan Data Mining untuk Klasifikasi Prediksi Penyakit ISPA dengan Algoritma C4.5. *Computer Science and ICT*, 3(1), 179–182. <https://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1711>
- Fakhriadi, R., Yulidasari, F., & Setyaningrum, R. (2015). Faktor risiko penyakit Demam Berdarah Dengue di Wilayah Kerja Puskesmas Guntung Payung Kota Banjarbaru (Tinjauan terhadap faktor manusia, lingkungan, dan keberadaan jentik). *Jurnal Publikasi Kesehatan Masyarakat Indonesia*, 2(1), 8–12. <https://ppjp.ulm.ac.id/journal/index.php/JPKMI/article/view/2703/2353>
- Ihsan, & Wajhillah, R. (2015). Penerapan Algoritma C4.5 Terhadap Diagnosa Penyakit Demam Tifoid Berbasis Mobile. *Jurnal Swabumi AMIK BSI Sukabumi*, III(1), 50–58. <https://ejournal.bsi.ac.id/ejurnal/index.php/swabumi/article/view/1006>
- Wulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta : Gava Media.