

Penerapan Perhitungan Metode Decision Tree Menggunakan Algoritma Iterative Dichotomiser 3 (ID3) Berbasis Website

Application of Decision Tree Method Calculation Using Website Based Iterative Dichotomiser 3 (ID3) Algorithm

Agus Susanto^{1,1}, Sasmitoh Rahmad Riady^{1,2}, Shita Dwi Ranti^{1,3}, Rila Mandala^{1,4}

President University

¹agus.susanto@student.president.ac.id, ²sasmitohrr@student.president.ac.id, ³shintaranti007@gmail.com,
⁴rilamandala@president.ac.id

Article Info:

Received: 10 – 07 - 2020

Accepted: 16 – 08 - 2020

Available Online: 17 – 08 - 2020

Keywords:

Machine Learning, Decision Tree, Algoritma ID3, Website.

Corresponding Author:

President University, Faculty Of Master Science
Information Technology
phone: (+62) 852-8936-3674
e-mail:
agus.susanto@student.president.ac.id

Abstract: *This study discusses the implementation of the Decision Tree algorithm into a website. Many calculation tools for machine learning methods include Rapidminer, WEKA, etc. These tools can also calculate various other machine learning methods. However, in this tool, many processes are carried out in testing even though only using one method such as the decision tree. Therefore, this study creates a website-based system that focuses on the Decision Tree ID3 method only, The results of the system calculations will be compared with the results of manual calculations. The website-based calculation tool implementation system for the ID3 algorithm is simple and easy to use, only by uploading training and testing data. Then the results of the calculation will be displayed on the website page. Based on the comparison of manual calculations with the system, there is no significant difference from the manual calculations. The system can run online and can access it from anywhere. In addition to displaying the results of calculations, this system also provides a predictive output and rule base from the calculation results of the decision tree algorithm.*

Abstrak: Dipenelitian ini membahas mengenai penerapan perhitungan *Decision Tree ID3* kedalam sebuah website, dimana tool perhitungan metode-metode machine learning yang sudah banyak digunakan diantaranya adalah Rapidminer, WEKA dll yang dapat menghitung berbagai metode machine learning lainnya, namun pada *tool* ini banyak proses dalam melakukan *testing* walaupun hanya menggunakan satu metode seperti halnya *decision tree ID3*. Maka dari itu peneliti membuat sebuah sistem berbasis website yang berfokus pada metode *Decision Tree ID3* saja, dari hasil perhitungan sistem akan dibandingkan dengan hasil perhitungan manual. Sistem penerapan perhitungan algoritma ID3 berbasis website ini sangat sederhana dan mudah digunakan dalam prosesnya, cukup dengan upload data *training* dan *testing* saja maka hasil dari perhitungan akan ditampilkan dihalaman website. Dari hasil perbandingan perhitungan manual dengan sistem, tidak ada perbedaan yang signifikan dengan perhitungan manualnya, dan sistem dapat berjalan dengan online dan di akses dari mana saja, selain menampilkan hasil perhitungan, sistem ini pun memberikan sebuah output hasil prediksi dan *rule base* dari hasil perhitungan algoritma pohon keputusan.

PENDAHULUAN

Metode *Decision Tree* adalah sebuah metode yang memiliki pendekatan fungsi target yang bernilai diskrit, dimana fungsi tersebut membentuk sebuah pohon keputusan, metode *Decision Tree* adalah salah satu bagian dari teknik *machine learning* dengan metode ini pun paling banyak digunakan dalam pembelajaran dasar belajar *machine learning* (Grosan & Abraham, 2011). Metode ini pun sering dijadikan *rule* dalam pembuatan sistem pendukung keputusan atau pun sistem pakar. Seperti mengambil sebuah keputusan dari pakar untuk me-diagnosa mengenai kesehatan kehamilan (Hikmatulloh et al., 2019) dan me-diagnosa untuk penyakit campak menggunakan algoritma ID3 (Srimenganti et al., 2018) dan sistem-sistem pakar lainnya, selain sistem pakar atau pun sistem pendukung keputusan *decision tree* dapat digunakan untuk metode *Data Mining* dalam mengolah data-data yang besar untuk dijadikan sebuah prediksi (Aradea et al., 2011).

Didalam metode *Decision Tree* terdapat beberapa algoritma yang sering digunakan diantaranya Algoritma ID3, Algoritma C.45 dan lain-lain, seperti yang akan dibahas di makalah ini yaitu mengenai algoritma ID3, algoritma tersebut adalah sebuah pembelajaran pohon keputusan sederhana yang dikembangkan oleh Ross Quinlan pada tahun 1997 (Quinlan, 1986). Konsep dasar dari algoritma ID3 ini adalah membangun sebuah pohon keputusan yang bercabang atau struktur hirarki, sehingga sistem ini cocok untuk menggambarkan suatu persoalan dan mencari solusi dari persoalan tersebut (Aji et al., 2018). ID3 pun memiliki tiga jenis node, yaitu *root node* sebuah simpul yang yang paling atas yang tidak memiliki inputan dan tidak mempunyai satu *ouput* namun memiliki lebih dari satu *output*, yang kedua adalah *internal node* yang merupakan sebuah node percabangan yang terdapat satu inputan dan mempunyai *ouput* lebih dari dua, dan yang terakhir adalah *leaf node* atau bisa disebut juga terminal node yang merupakan node akhir, hanya memiliki satu inputan dan tidak memiliki output (Kirana et al., 2019).

Pembelajaran *Machine Learning* terdapat beberapa formula yang harus dihitung seperti metode *DT*, *Naïve Bayes*, *SVM* dan lainnya memiliki formula yang harus di hitung baik secara manual ataupun menggunakan tool untuk mendapatkan hasil dari perhitungannya. Diantara tool tersebut yang sering digunakan ialah Rapidminer, WEKA, python ataupun tool yang dibuat sendiri yang bertujuan untuk mempermudah dalam perhitungan manual. Pada metode *Decision Tree* sendiri memiliki ciri khas dalam formula untuk mencari entropy dan gain dalam menentukan sebuah root dan membangun pohon rule keputusan untuk kondisi dari hasil perhitungan klasifikasi data training (Safavian & Landgrebe, 1991). Dalam metode *decision tree* ID3 mempunyai pendekatan perolehan informasi umumnya digunakan untuk menentukan properti yang cocok untuk setiap node pohon keputusan yang dihasilkan. Dengan demikian kita dapat memilih atribut dengan perolehan informasi tertinggi (pengurangan entropi pada tingkat maksimum) sebagai atribut uji dari simpul saat ini. Dengan cara ini, informasi yang diperlukan untuk mengklasifikasikan subset sampel pelatihan yang diperoleh dan kemudian pada partisi akan menjadi yang terkecil. Artinya, penggunaan properti ini untuk mempartisi set sampel yang terkandung dalam node saat ini akan membuat tingkat campuran dari berbagai jenis untuk semua subset sampel yang dihasilkan berkurang menjadi minimum. Oleh karena itu, penggunaan pendekatan teori informasi seperti itu akan secara efektif mengurangi jumlah klasifikasi objek yang diperlukan (Chen et al., 2009). Dari tool yang telah tersedia yang biasa digunakan untuk menghitung berbagai metode yang ada, namun dipenelitian ini peneliti membuat sebuah tool dimana me-focuskan terhadap perhitungan manual yang ada di algoritma ID3 kedalam sebuah website.

METODE PENELITIAN

Terdapat beberapa metode dalam penelitian ini diantaranya adalah, menghitung beberapa data training dan data testing yang diambil dari beberapa literatur yang nantinya akan dijadikan bahan perbandingan dari output sistem yang telah dibuat.

2.1. Formula Algoritma ID3

Berikut ini adalah formula untuk menentukan *Root* berdasarkan pada nilai *Gain* yang tertinggi dari atribut-atribut yang ada, untuk mendapatkan nilai *Gain* harus ditentukan terlebih dahulu nilai *Entropy*.

Rumus *Entropy*:

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

S = Himpunan Kasus

S = Jumlah Partisi S

S = Proporsi dari S_i Terhadap S

Rumus *Gain*:

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

S = Himpunan Kasus

S = Atribut

S = Jumlah Partisi Atribut A

S_i = Jumlah Kasus pada Partisi ke- i

S = Jumlah Kasus dalam S

2.2. Perhitungan Manual Data Training

Berikut ini adalah perhitungan data training yang diambil dari beberapa literatur

a. Perhitungan data training Play Golf

Berikut ini adalah sample data training yang berskala kecil yang mempunyai 4 atribut, 1 kelas, dan 14 data *value*(Peng et al., 2009).

Table 1. Data Training Play Golf

Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
sunny	Mild	High	FALSE	Yes
sunny	Cool	Normal	FALSE	Yes
sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
sunny	Mild	High	TRUE	No

Berikut ini adalah tabel kemungkinan atribut yang bermain *Play Golf*

Table 2. Tabel Kemungkinan Play Golf

Atribut	Nilai Kemungkinan
Outlook	Sunny, Overcast, Rainy
Temperature	Hot, Cool, Mild
Humidity	High, Normal
Windy	True, False

Mencari entropy tabel frekuensi dalam satu atribut

$$\begin{aligned}
 Entropy(Play\ Golf) &= Entropy(5,9) \\
 &= Entropy\left(\frac{5}{14}, \frac{9}{14}\right) \\
 &= Entropy(0,36, 0,64) \\
 &= -(0,36 \log_2 0,36) - (0,64 \log_2 0,64) = 0,94
 \end{aligned}$$

Berikutnya mencari informasi *Gain* untuk setiap atribut

$$Entropy(S, Windy) = Entropy(S) - \left(\frac{9}{14}\right) Entropy(S_{yes}) - \left(\frac{5}{14}\right) Entropy(S_{no}) = 0,048$$

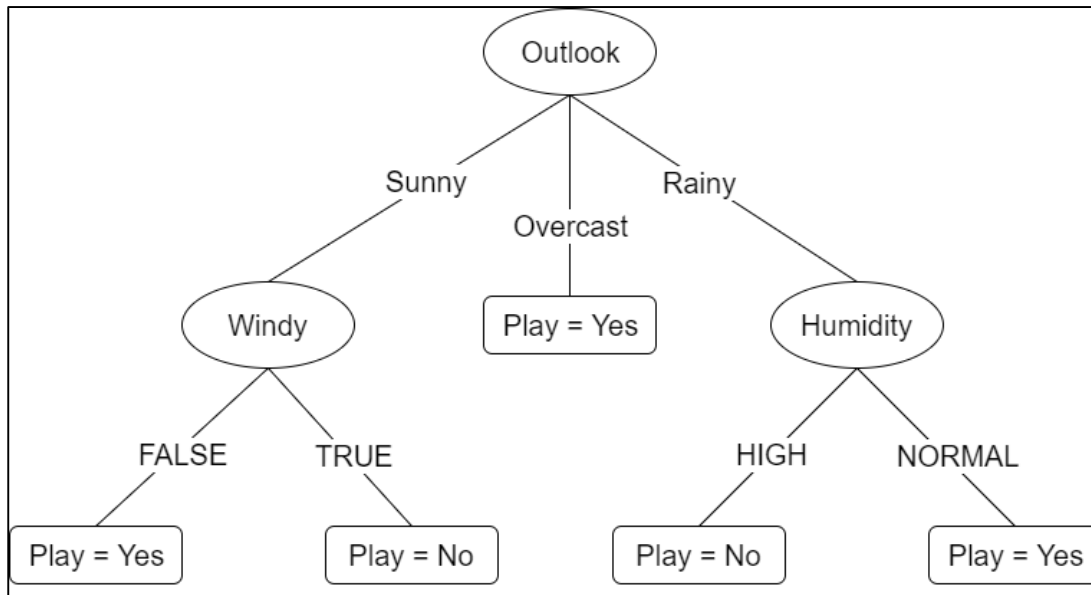
$$Entropy(S, Humidity) = 0,152$$

$$Entropy(S, Temperature) = 0,029$$

$$Entropy(S, Outlook) = \mathbf{0,246}$$

Pilih atribut dengan perolehan nilai maximum untuk di jadikan sebuah *Root*. Dari perhitungan yang sudah dilakukan maka nilai atribut untuk root adalah "Outlook".

Berikut adalah gambar model yang telah dihitung dan untuk *leaf node* dilakukan kembali perhitungan seperti di atas.



Gambar 1. Model Decision Tree Play Golf

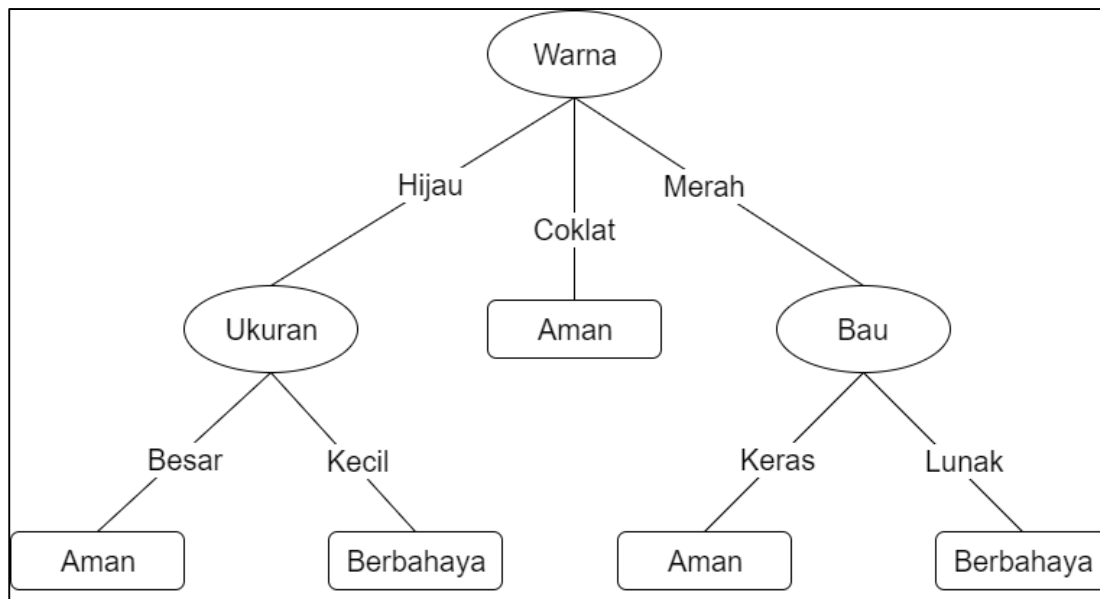
b. Perhitungan Data Training Kelas Buah

Berikut ini adalah sample data training yang berskala kecil mengenai data kelas buah untuk memprediksi tingkat kelas berbahaya atau aman yang memiliki 4 atribut dan 1 kelas lalu memiliki 16 data value.

Table 3. Data Training Kelas Buah

Kulit Buah	Warna	Ukuran	Bau	Kelas
Kasar	Coklat	Besar	Keras	Aman
Kasar	Hijau	Besar	Keras	Aman
Halus	Merah	Besar	Lunak	Berbahaya
Kasar	Hijau	Besar	Lunak	Aman
Kasar	Merah	Kecil	Keras	Aman
Halus	Merah	Kecil	Keras	Aman
Halus	Coklat	Kecil	Keras	Aman
Kasar	Hijau	Kecil	Lunak	Berbahaya
Halus	Hijau	Kecil	Keras	Berbahaya
Kasar	Merah	Besar	Keras	Aman
Halus	Coklat	Besar	Lunak	Aman
Halus	Hijau	Kecil	Keras	Berbahaya
Kasar	Merah	Kecil	Lunak	Aman
Halus	Merah	Besar	Keras	Berbahaya
Halus	Merah	Kecil	Keras	Aman
Kasar	Hijau	Kecil	Keras	Berbahaya

Berikut ini adalah gambar model dari decision tree kelas buah, untuk perhitungannya sesuai dengan point A



Gambar 2. Model Decision Tree kelas buah

PEMBAHASAN

Dibagian ini adalah pengetestan data menggunakan website dalam mencari perhiungan entropy, gain dan leaf node termasuk menentukan prediksi dari data testing.

3.1. Testing kasus Play Golf

Berikut ini adalah hasil data yang akan ditesting menggunakan sistem yang berada di website, kemudian akan dibandingkan dengan perhitungan manual.

a. Data testing

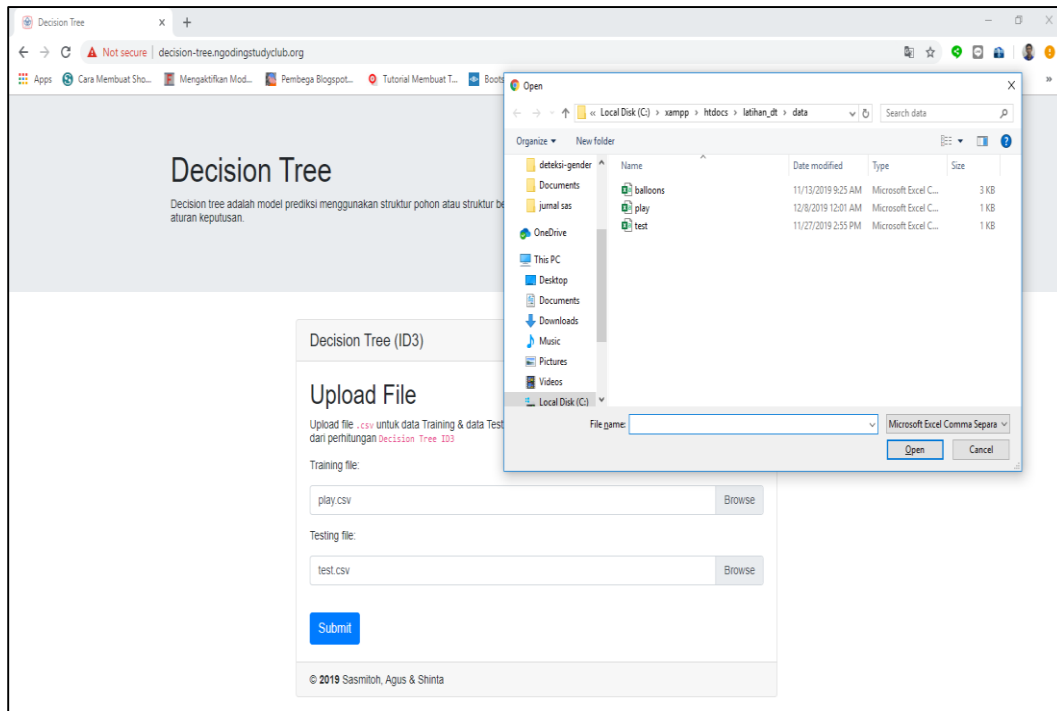
Berikut ini adalah tabel data testing untuk mencari prediksi yang bermain golf.

Table 4 Data Testing Play Golf

Outlook	Temperature	Humidity	Windy	Play Golf
Overcast	Mild	High	FALSE	?

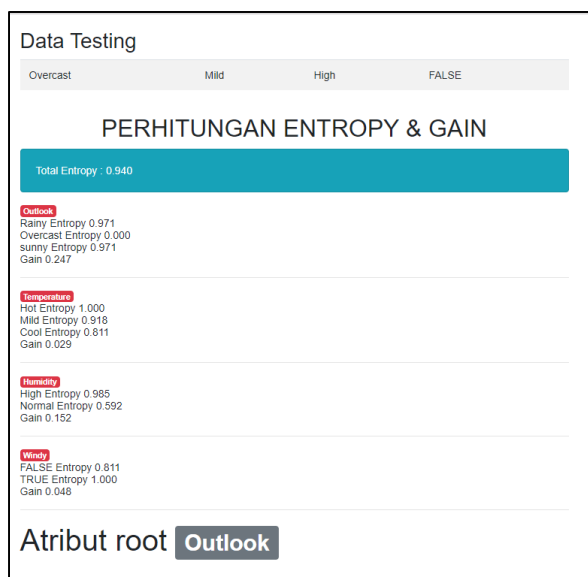
b. Proses Testing menggunakan website

Didalam sitem yang telah dibuat data harus berupa file csv dan di upload untuk pencarian nilai entropy, gain, dan root. Berikut adalah gambar didalam proses upload file.



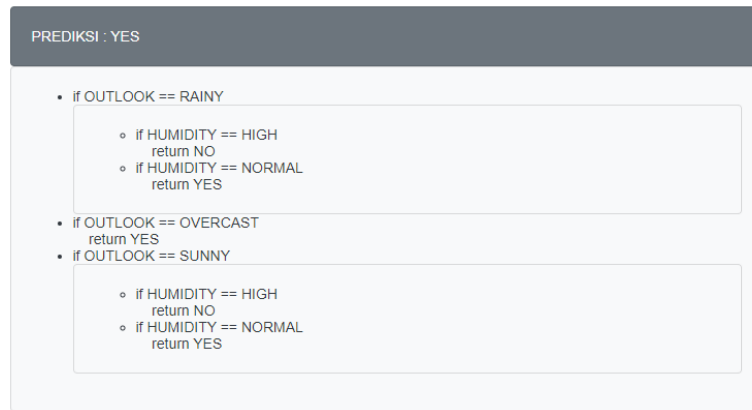
Gambar 3. Upload data training dan data testing

Selanjutnya ditahap ini adalah me-submit data *training* dan data *testing* untuk menghasilkan prediksi. Berikut adalah gambar dari hasil *testing* menggunakan website.



Gambar 4 Hasil Perhitungan

Berikutnya adalah gambar hasil prediksi dan *rule base* sistem



Gambar 5. Prediksi dan Rule Base

3.2. Testing Kasus Kualitas Buah

Berikut ini sama seperti kasus *play golf* dalam mencari hasil data testing, kemudian menggunakan sistem yang berada didalam website kemudian akan dibandingkan dengan perhitungan manual.

a. Data Testing

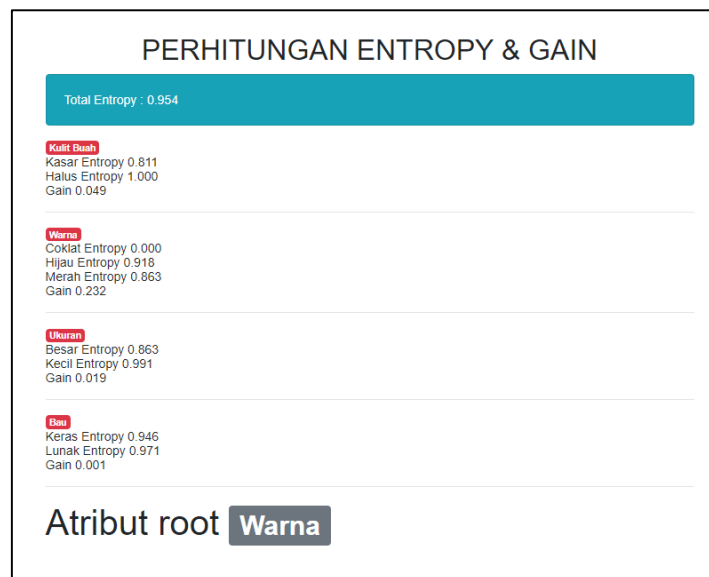
Berikut ini adalah tabel data testing yang dalam mencari prediksi Kualitas buah.

Table 5. Data Testing Kualitas Buah

Kulit Buah	Warna	Ukuran	Bau	Kelas
Kasar	Hijau	Kecil	Keras	?

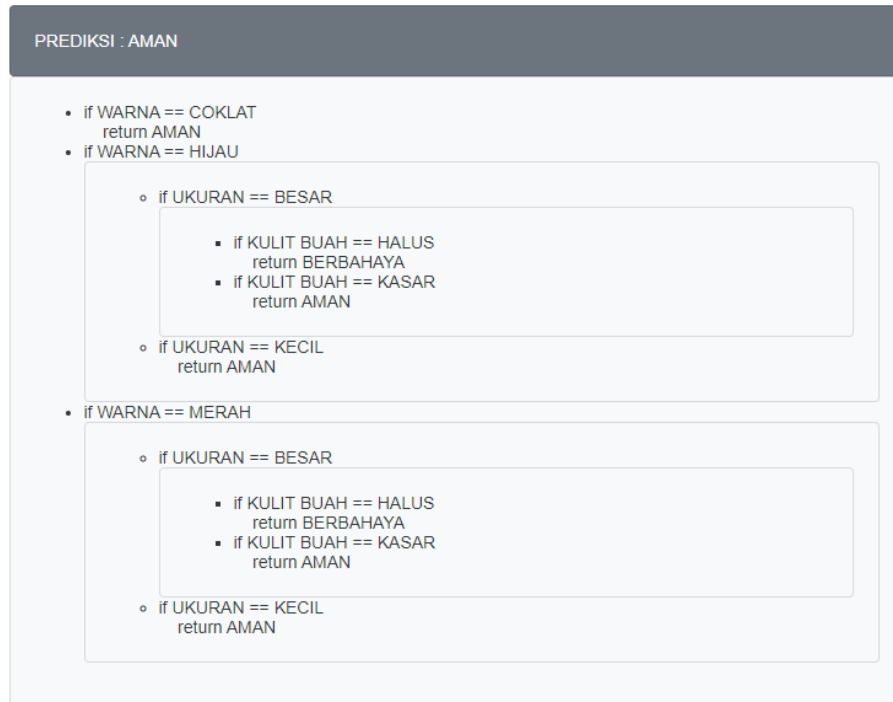
b. Data Testing

Proses testing menggunakan sistem website, pada tahap ini sama seperti kasus *play golf* dalam mencari nilai *entropy* dan *Gain* yang harus di hitung terlebih dahulu. Berikut adalah hasil perhitungan dari sistem untuk entropy dan gain.



Gambar 6. Hasil perhitungan

Selanjutnya adalah gambar hasil prediksi dan *rule base*



Gambar 7. Hasil Prediksi

KESIMPULAN

Dari hasil penelitian yang telah dilakukan pada penerapan perhitungan Decision tree didalam sebuah website, maka hasilnya sama dengan perhitungan manual yang telah dilakukan sebelumnya, sistem ini hanya me-focuskan pada pencarian nilai *entropy* dan *gain*, dalam pencaharian *root* dan *leaf node*, adanya sistem website ini sangat membantu dalam klasifikasi data terhadap metode *Decision Tree ID3*, dan proses pembelajaran inipun sangatlah mudah di mengerti, tidak halnya seperti tool lainnya yang memerlukan banyak proses dalam menentukan sebuah prediksi, tentunya didalam sistem ini masih banyak kekurangannya diantaranya tidak bisa me-cover banyaknya suatu atribut dan banyaknya data dalam menentukan pola.

DAFTAR PUSTAKA

- Aji, E., Kurniawan, P., Hidayat, N., & Wijoyo, S. H. (2018). Implementasi Metode Iterative Dichotomizer Tree (ID3) Untuk Diagnosis Penyakit Pada Tanaman Bawang Merah. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(10), 3391–3396.
- Aradea, S. A., Ariyan Z, & Yuliana A. (2011). Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru. *Jurnal Penelitian Sitrotika*, 7(1), 1693–9670.
- Chen, J., Luo, D. L., & Mu, F. X. (2009). An improved ID3 decision tree algorithm. *Proceedings of 2009 4th International Conference on Computer Science and Education, ICCSE 2009*, 3, 127–130. <https://doi.org/10.1109/iccse.2009.5228509>
- Grosan, C., & Abraham, A. (2011). Machine Learning. In *Intelligent Systems Reference Library*

(Vol. 17). https://doi.org/10.1007/978-3-642-21004-4_10

- Hikmatulloh, H., Rahmawati, A., Wintana, D., & Ambarsari, D. A. (2019). Penerapan Algoritma Iterative Dichotomiser Three (Id3) Dalam Mendiagnosa Kesehatan Kehamilan. *Klik - Kumpulan Jurnal Ilmu Komputer*, 6(2), 116. <https://doi.org/10.20527/klik.v6i2.189>
- Kirana, K. C., Bintang, C., Winata, W., Astuti, I., & Putra, I. R. (2019). *Prediksi rating reksadana berbasis algoritma decision tree pada sistem informasi reksadana*. 29(2), 140–151.
- Peng, W., Chen, J., & Zhou, H. (2009). An Implementation of IDE3 Decision Tree Learning Algorithm. *Project of Comp 9417: Machine Learning*, 1, 1–20.
<http://cis.k.hosei.ac.jp/~rhuang/Miccl/AI-2/L10-src/DecisionTree2.pdf>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
<https://doi.org/10.1007/bf00116251>
- Safavian, S. R., & Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3), 660–674.
<https://doi.org/10.1109/21.97458>
- Srimenganti, I., Taufik, I., & Mulyana, E. (2018). Implementasi Algoritma Decision Tree (ID3) Untuk Penyakit Campak. *Seminar Nasional Teknik Elektro*, 235–242.