



## **KLASIFIKASI TREN PENYAKIT PASIEN BEDASARKAN DIANOCSA ICD-10 DI RUMAH SAKIT UMUM CUT MEUTIA MENGGUNAKAN METODE C4.5**

**Zakial Vikki<sup>1)</sup>, Yuswandi<sup>2)</sup>**

<sup>1,2)</sup> Jurusan Informatika, Fakultas Ilmu Komputer dan Multimedia, Universitas Islam Kebangsaan Indonesia Bireuen

*e-mail:* \*) [zakialviki.mkom@gmail.com](mailto:zakialviki.mkom@gmail.com)

### **Abstract**

*[Classification of Patient Disease Trends Based on ICD-10 Diagnosis at Cut Meutia General Hospital Using the C4.5 Method] The rapid development of information technology has impacted many people getting more and more data every day, even excessive. so that the use of the data is not optimal. Likewise patient medical record data at Cut Meutia General Hospital in North Aceh by serving patients every day. The large number of patients handled automatically makes this hospital accommodate a lot of patient medical record data with various types of diseases so that a method is needed to pattern the patient's disease data. For this reason, this study aims to find trends in patient disease at North Aceh Cut Meutia General Hospital based on ICD-10 diagnostics using data mining techniques by analyzing the C4.5 method where the C4.5 method creates a classification model from a large data set so as to produce patterns the new data pattern forms a decision tree (Decision Tree) useful for exploring data, finding relationships between a number of input variables with a target variable. The data used in this study were obtained from the North Aceh Cut Meutia General Hospital for 2020-2021 based on 4 variable data, namely age, gender, address and ICD-10 diagnosis. ICD-10 is a diagnostic classification with international standards that is compiled based on a category system and reports in disease units according to criteria agreed upon by international experts. ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th revision).*

**Keywords:** Algorithm C4.5; Data Mining; ICD-10; Patient Disease Trends

### **Abstrak**

Pesatnya perkembangan teknologi informasi berimbas banyak orang memperoleh data dengan mudah yang semakin hari bertambah banyak bahkan berlebihan. sehingga penggunaan data tersebut tidak optimal. Begitu pun data rekam medis pasien pada Rumah Sakit Umum Cut Meutia Aceh Utara dengan melayani pasien setiap harinya. Banyaknya pasien yang ditangani secara otomatis membuat rumah sakit ini menampung banyak data rekam medis pasien dengan berbagai jenis penyakit sehingga sangat di butuhkan suatu metode untuk mempolakan data penyakit pasien tersebut. Untuk itu penelitian ini bertujuan untuk mencari tren penyakit pasien di Rumah Sakit Umum Cut Meutia Aceh Utara berdasarkan dianogsa ICD-10 menggunakan teknik data mining dengan menganalisa metode C4.5 dimana metode C4.5 membuat model klasifikasi dari kumpulan data yang besar sehingga menghasilkan pola-pola data yang baru membentuk pohon keputusan (*Decision Tree*) berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Data yang digunakan pada penelitian ini diperoleh dari Rumah Sakit Umum Cut Meutia Aceh Utara tahun 2020-2021 berdasarkan 4 variabel data yaitu usia, jenis kelamin, alamat dan diagnosa ICD-10. ICD-10 adalah klasifikasi diagnostik penyakit dengan standar internasional yang disusun berdasarkan sistem kategori dan dikelompokkan dalam satuan penyakit menurut kriteria yang telah disepakati pakar internasional. Implementasi hasil penelitian ini berupa sistem berbasis web yang akan memberikan informasi tren penyakit pasien di Rumah Sakit Umum Cut Meutia Aceh Utara berdasarkan dianogsis ICD-10 (*International Statistical Classification of Diseases and Related Health Problems* revisi ke 10)

**Kata kunci :** Algoritma C4.5; Data Mining; ICD-10; Tren Penyakit Pasien

## 1. Pendahuluan

Seiring dengan pesatnya perkembangan teknologi informasi telah menyebabkan banyak orang memperoleh data dengan mudah bahkan berlebihan. Data tersebut terakumulasi semakin banyak dari hari ke hari, akibatnya pemanfaatan data yang terakumulasi tersebut menjadi tidak optimal. Begitu pun data rekam medis pasien pada Rumah Sakit Umum Cut Meutia Aceh Utara dengan menerima dan melayani banyak pasien setiap harinya. Banyaknya pasien yang ditangani pihak rumah sakit ini tiap harinya secara otomatis membuat rumah sakit ini menampung banyak data rekam medis pasien dengan berbagai jenis penyakit. Pencatatan medis pada rumah sakit pemerintah maupun dinas kesehatan diwajibkan menggunakan standart ICD- 10 (*International Statistical Classification of Diseases and Related Health Problems* revisi ke 10).

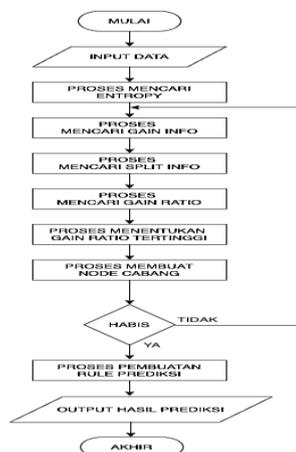
Data rekam medis pasien berisikan catatan dan dokumen tentang identitas pasien, hasil pemeriksaan, pengobatan, tindakan dan pelayanan lain yang telah diberikan kepada pasien tersebut dapat dipakai sebagai acuan untuk pemeriksaan kesehatan pasien selanjutnya, sekaligus sebagai bukti (Bukovský, 2017). Untuk memanfaatkan data rekam medis yang ada di Unit Rekam Medis Rumah Sakit Umum Cut Meutia Aceh Utara, penulis melakukan sebuah pendekatan untuk melakukan penelitian terhadap data rekam medis tersebut guna mendapatkan informasi yang berharga yang bisa digunakan sebagai proyeksi tentang penyakit di daerah tersebut, terutama di Rumah Sakit Umum Cut Meutia Aceh Utara. Hal ini dapat dilakukan dengan menggunakan sebuah teknik yang bernama data mining. Salah satu algoritma yang populer dari teknik data mining klasifikasi adalah algoritma C4.5.

Dari permasalahan di atas, penelitian ini akan mengimplementasikan teknik klasifikasi dari data mining yang menggunakan algoritma C4.5 sebagai metode untuk menganalisa tren penyakit pasien berdasarkan data ICD-10 di Rumah Sakit Umum Cut Meutia Aceh Utara. Adapun judul penelitian ini yaitu “Implementasi Algoritma C4.5 Untuk analisa Tren Penyakit Pasien Di Rumah Sakit Umum Cut Meutia Aceh Utara Berdasarkan Diagnosa ICD-10 (*International Statistical Classification Of Diseases And Related Health Problems Revisi Ke 10*)” .

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk menerapkan algoritma C4.5 dalam sistem penentuan tren penyakit pasien di Rumah Sakit Umum Cut Meutia Aceh Utara. Algoritma C4.5 akan membuat *decision tree* (pohon keputusan) untuk melihat kemungkinan penyakit yang sedang tren saat ini. Pada penelitian ini sistem dibangun menggunakan bahasa pemrograman web. Informasi yang dihasilkan dari sistem ini akan memberikan output dari hasil analisis algoritma C4.5 terhadap variabel yang di gunakan. Dengan adanya penelitian ini diharapkan akan memudahkan user untuk mengetahui tren penyakit di setiap daerah Aceh Utara berdasarkan data di Rumah Sakit Umum Cut Meutia Aceh Utara.

## 2. Metodologi

Dalam penelitian ini menggunakan data yang diperoleh dari RSUD Cut Meutia. Teknik pengumpulan data melalui observasi dengan mengumpulkan dokumen rumah sakit yang berupa data rekam medik pasien yang menjadi data sekunder untuk klasifikasi tren penyakit pasien dengan menggunakan algoritma C4.5. Skema sistem untuk pengolahan data dengan menggunakan algoritma adalah C4.5 sebagai berikut:



**Gambar 1.** Skema Algoritma C4.5

**A. Data Mining**

*Knowledge Discovery in Databases* (KDD) atau sering di sebut data mining, adalah proses pengumpulan informasi dari suatu data besar, penemuan informasi baru dan berpotensi berguna dari sejumlah besar data. Penerapan data mining sebageaian besar berguna di bidang penelitian pendidikan dan penyelidikan. Penggunaan metode statistika, matematika serta memanfaatkan artificial intelligence sering di gunakan untuk proses data mining (Peterson & Baker, n.d.). Data mining merupakan cara yang bisa membantu memperkirakan fenomena suatu objek yang diteliti dengan terlebih dahulu mengetahui atau tidak mengetahui suatu kelas suatu obyek tertentu melalui masukan dari data besar (Natasuwarna, 2019).

**B. Algoritma C4.5**

Algoritma C4.5 merupakan algoritma untuk membuat *decision tree* (pohon keputusan) berdasarkan training data yang telah diberikan. Algoritma C4.5 diperkenalkan oleh J.Ross Quinlan yang merupakan pengembangan dari algoritma ID3 (Sularno & Anggraini, 2017). Pohon keputusan adalah cara klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengekspolari data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Tujuan pohon keputusan yang di buat dengan algoritma c4.5 untuk meningkatkan akurasi dari prediksi yang sedang dilakukan, algoritma C4.5 banyak di gunakan untuk membuat pohon keputusan karena mudah di pahami (Navia et al., 2016).

Dengan menggunakan algoritma C4.5 dalam pembuatan pohon keputusannya, ada 4 langkah yang harus dilakukan, yaitu (Rosandy, 2016):

1. Memilih atribut sebagai akar (root).
2. Membuat cabang untuk tiap-tiap nilai.
3. Membagi kasus dalam cabang.
4. Mengulangi proses dalam setiap cabang, sampai semua kasus dalam cabang memilikikelas yang sama.

Pada data mining, entropi berfungsi sebagai parameter untuk mengukur keragaman dalamsuatu kumpulan data (Sinaga et al., 2021). Perhitungan nilai *entropy* dapat dilihat pada persamaan berikut:

$$i=1$$

$$Entropy(S) = \sum^n -pi * log_2pi.....(1)$$

S: Himpunan Kasus.  
 N:Jumlah partisi S.  
 pi: proporsi dari Si terhadap S.

Pengukuran nilai efektifitas suatu atribut dalam mengklasifikasikan data baru bisa dilakukan setelah mendapatkan nilai entropi untuk suatu kumpulan data. Untuk memilih atribut akar,didasarkan pada nilai *gain ratio* tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera dalam persamaan berikut:

$$Gain(S, A) = \sum^n \frac{|Si|}{|S|} * Entropy(Si)..... (2)$$

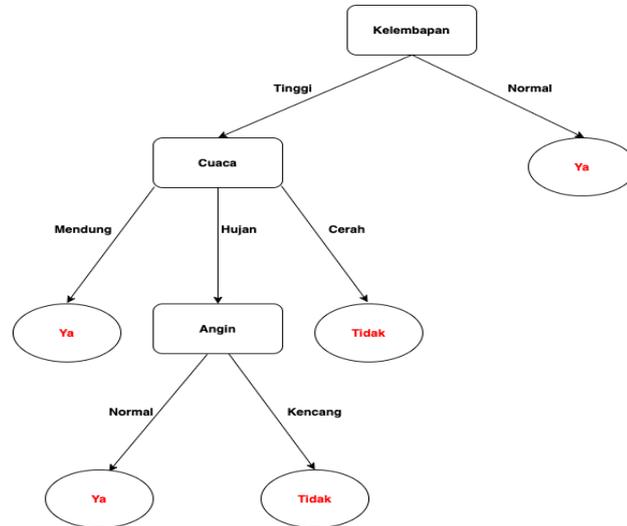
S : Himpunan Kasus  
 A : Atribut  
 N : Jumlah partisi atribut A  
 |Si| : Jumlah kasus pada partisi ke-i  
 |S| : Jumlah kasus dalam

**C. Pohon Keputusan**

Ciri-ciri dari sebuah pohon terdiri dari, akar, batang, daun, sama hal nya juga dengan pohon sebuah keputusan yang berisi simpul akar, cabang, simpul daun. Simpul akar adalah induk dari semua node dan

seperti namanya itu adalah node paling atas di Pohon. Sebuah pohon keputusan adalah pohon di mana setiap node menunjukkan fitur (atribut), setiap tautan (cabang) menunjukkan keputusan (aturan) dan setiap daun menunjukkan hasil. *Decision Tree* merupakan metode yang banyak keunggulan di antaranya dapat memproses data numerik, diskrit, dan mudah tafsirkan (Pujianto et al., 2019).

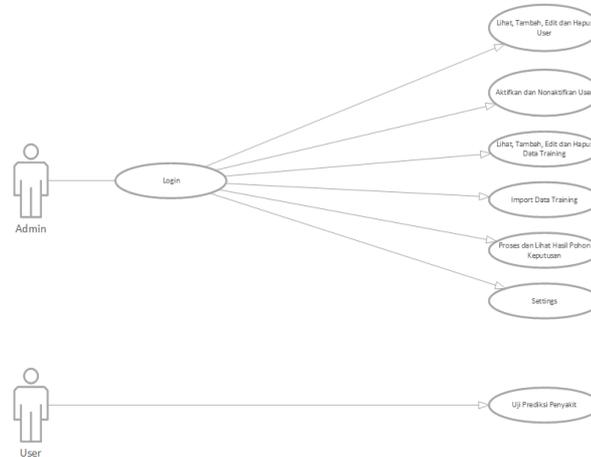
Data yang di berikan pada pohon keputusan di nyatakan dalam bentuk table, atribut dan record. Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model *tree*, mengubah model *tree* menjadi *rule* yang sudah di sederhanakan.



Gambar 2. Pohon Keputusan

#### D. Sistem yang di usulkan

Sistem yang akan di usul pada penelitian ini mampu melakukan perhitungan untuk mencari tren penyakit pasien yang terjadi di sebuah daerah dengan memanfaatkan data *medical record* yang ada di Rumah Sakit. Data



Gambar 3. Use Diagram

### 3. Hasil Dan Pembahasan

#### A. Proses Klasifikasi dengan C4.5

Pada tahap ini penulis menggunakan data medical record dengan variabel jenis kelamin, umur, wilayah. Adapun detail dari variabel-variabel tersebut adalah sebagai berikut:

Tabel 1. Dianogsa

Kode Diagnosa	Diagnosa
---------------	----------

A00-B99	Penyakit infeksi dan parasitik tertentu
C00-D48	Neoplasma
D50-D89	Penyakit darah, organ pembentuk darah dan gangguan yang melibatkan mekanisme imunitas
E00-E90	Penyakit endokrin, nutrisi dan metabolik
F00-F90	Gangguan mental dan perilaku
G00-G99	Penyakit sistem saraf
H00-H59	Penyakit mata dan adneksa mata
H60-H95	Penyakit telinga dan prosesus mastoid
.....	.....
S00-T98	Cedera, keracunan dan konsekuensi-konsekuensi lain akibat sebab luar
V01-Y98	Sebab-sebab luar mortalitas dan morbiditas
Z00-Z99	Faktor-faktor yang mempengaruhi status kesehatan dan kontak dengan pelayanan kesehatan

Tabel 2. Gabungan data

No	Jenis Kelamin	Umur	Alamat	Diagnosa
1	L	Bayi	Baktiya	R00-R99
2	L	Anak-anak	Baktiya Barat	F00-F90
...	...	...	...	...
20	P	Dewasa	Cot Girek	G00-G99

Proses selanjutnya adalah dengan mencari *node* akar (*root*) yaitu dengan melakukan perhitungan entropi total dan perhitungan terhadap entropi masing-masing atribut dengan menggunakan persamaan (2.1).

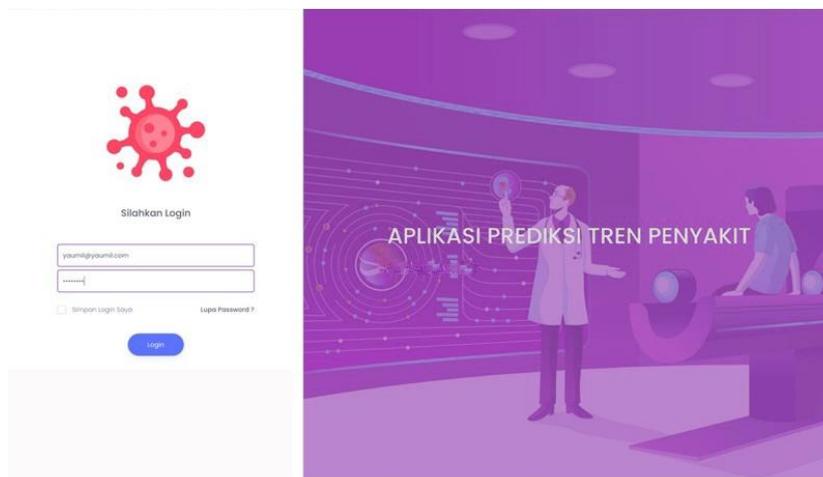
$$Entropi(Total) = \sum_{i=1}^n -p_i * \log_2 p_i \dots \dots \dots (3)$$

$$Entropi(Total) = 0.204787859 + 0.079443311 + 0.01928602 + 0.180133564 + 0.530116789 + 0.088245464 + 0.168384811 + 0.032042525 + 0.348381965 + 0.14390423 + 0.132412716 + 0.112043792 + 0.227487821 + 0.375554607 + 0.007075586 + 0.002104523 + 0.052287734 + 0.185022718 + 0.203276144 + 0.011433881 + 0.191413374$$

Dari perhitungan ini didapatkan entropi total keseluruhan atribut adalah 3.294839434. Nilai inilah yang akan digunakan untuk menghitung nilai gain terhadap masing-masing atribut nantinya.

## B. Implementasi

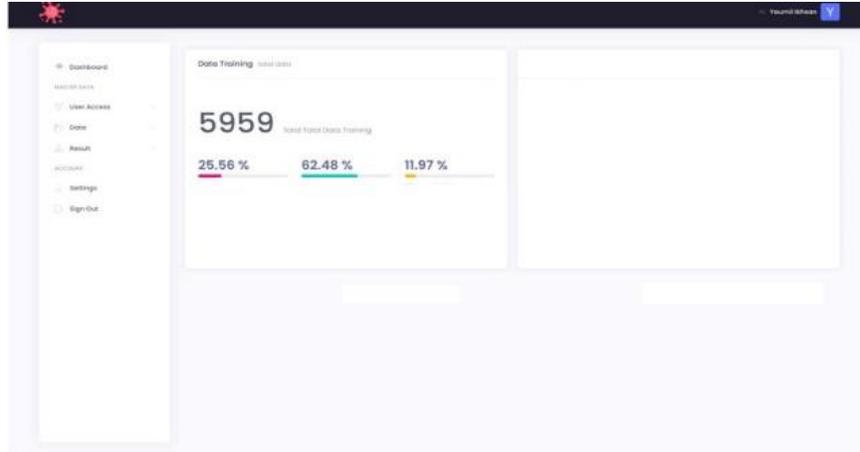
Implementasi sistem klasifikasi tren penyakit pasien pada RSUD Cut Meutia berdasarkan dianogsa ICD-10, mempunyai beberapa interface (antar muka). Berikut adalah tampilan halaman login:



Gambar 4. Menu Login

User harus melakukan login terlebih dahulu untuk menggunakan system, dengan cara memasukan *username* dan *password* yang telah dibuat sebelumnya. System akan memvalidasi apakah data yang di inputkan sesuai dengan data yang sudah terdaftar, jika datanya benar, maka system akan mengarahkan user ke halaman dashboard.

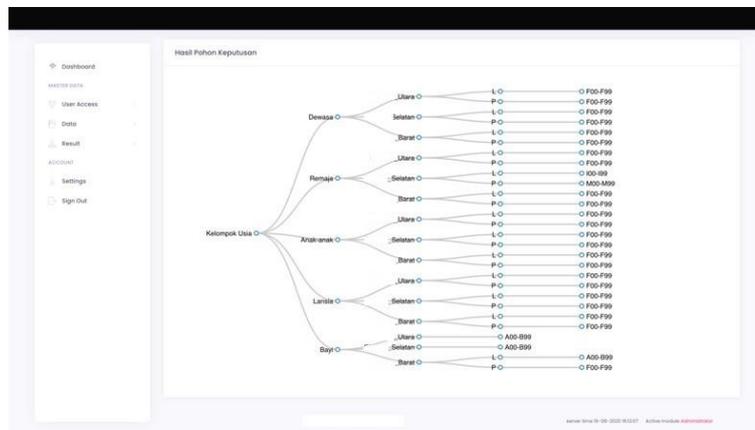
a. Dashboard



Gambar 5. Menu Dashboard

Halaman utama dashboard adalah halaman pertama yang akan dilihat oleh *user* ketika berhasil melakukan *login* ke sistem untuk pertama kali. Pada halaman ini sistem akan menampilkan informasi mengenai data yang sudah ada pada sistem dan informasi tentang klasifikasi tren penyakit pasien.

b. Pohon Keputusan



Gambar 6. Menu Pohon Keputusan

#### 4. Kesimpulan

Uraian di atas menyangkut penjelasan penelitian terhadap pembahasan terkait, maka dapat di petik kesimpulan di mana menurut analisis serta pengembangan sistem Klasifikasi Tren Penyakit Pasien Berdasarkan Dianogsa ICD-10 di RSUD Cut Meutia yang dapat digunakan oleh pihak berwenang untuk melihat proyeksi penyakit di daerah tersebut sebagai informasi guna mengambil kebijakan pencegahan dan penanganan.

1. Hasil akhir penelitian ini adalah sebuah sistem Analisa Tren Penyakit Pasien Berdasarkan Dianogsa ICD-10 di RSUD Cut Meutia yang dapat digunakan oleh pihak berwenang untuk melihat proyeksi penyakit di daerah tersebut sebagai informasi guna mengambil kebijakan pencegahan dan penanganan.
2. Algoritma C4.5 diimplementasikan untuk menganalisa tren penyakit pasien berdasarkan data *medical record* dengan menggunakan 3 variabel yang dijadikan acuan yaitu jenis kelamin, usia dan alamat yang dibagi berdasarkan wilayah serta 1 variabel yang digunakan sebagai label yaitu diagnosa.

## Daftar Pustaka

- Bukovský, L. (2017). Generic extensions of models of ZFC. *Commentationes Mathematicae Universitatis Carolinae*, 58(3), 347–358. <https://doi.org/10.14712/1213-7243.2015.209>
- Natasuwarna, A. P. (2019). *Seminar Nasional Hasil Pengabdian Kepada Masyarakat 2019 SINDIMAS 2019 STMIK Pontianak* (Vol. 29).
- Navia, L., Dosen, R., Informasi, S., Yptk, U. ", Padang, ", Raya, J., Begalung, L., & Barat, P.-S. (2016). *Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit*. 1(2).
- Peterson, B., & Baker, P. S. J. D. (n.d.). *Data Mining for Education*.
- Pujianto, U., Setiawan, A. L., Rosyid, H. A., & Salah, A. M. M. (2019). *Comparison of Naïve Bayes Algorithm and Decision Tree C4.5 for Hospital Readmission Diabetes Patients using HbA1c Measurement*. 2(2), 58–71. <https://doi.org/10.17977/um017v2i22019p58-71>
- Rafiska, R., Defit, S., & Nurcahyo, G. W. (2018). *Analisis Rekam Medis untuk Menentukan Pola Kelompok Penyakit Menggunakan n Algoritma C4.5*. 2(1), 391–396. <http://jurnal.iaii.or.id>
- Rosandy, T. (2016). *PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus : KSPPS / BMT AL-FADHILA)*. 02.
- Sinaga, T. H., Wanto, A., Gunawan, I., Sumarno, S., & Nasution, Z. M. (2021). Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM. *Journal of Computer Networks, Architecture, and High-Performance Computing*, 3(1), 9–20. <https://doi.org/10.47709/cnahpc.v3i1.923>
- Sularno, S., & Anggraini, P. (2017). PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI TINGKAT KEGANASAN HAMA PADA TANAMAN PADI (Studi Kasus : Dinas Pertanian Kabupaten Kerinci). *Jurnal Sains Dan Informatika*, 3(2), 161. <https://doi.org/10.22216/jsi.v3i2.2779>