

# Penerapan Algoritma C4.5 Mengklarifikasi Penerimaan Bantuan Sosial Menggunakan Feature Selection

M Wandu Dwi Wirawan\*, Siska Kurnia Gusti, Jasril, Pizain

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia

Email: <sup>1,\*</sup>11950111705@students.uin-suska.ac.id, <sup>2</sup>siskakurniagusti@uin-suska.ac.id, <sup>3</sup>jasril@uin-suska.ac.id, <sup>4</sup>pizaini@uin-suska.ac.id

Email Penulis Korespondensi: 11950111705@students.uin-suska.ac.id

Submitted: 12/08/2023; Accepted: 14/09/2023; Published: 25/09/2023

**Abstrak**—Upaya pemerintahan Indonesia untuk menanggulangi kemiskinan di Indonesia adalah melalui program Kartu Indonesia Pintar (KIP) yang dilakukan pemerintah dalam bentuk memberikan bantuan kepada keluarga yang kurang mampu. Tujuan utama dari penyaluran bantuan KIP agar dapat membantu menyekolahkan anak-anak yang kurang mampu untuk melanjutkan ke jenjang Pendidikan, bentuk kesulitan yang terdapat pada penerimaan KIP, disebabkan banyak warga yang mendaftar, serta data yang memiliki beberapa ketentuan, keterbatasan waktu yang dimiliki dalam pemberian KIP oleh pihak kelurahan pangkalan sesai yang relatif rendah, karena itu pemberian bantuan harus tepat sasaran. Sebab itu tujuan penelitian ini mencari atribut paling berpengaruh dalam penerimaan bantuan KIP agar dapat meningkatkan hasil proses verifikasi data. Setelah melakukan Seleksi Feature menggunakan Information Gain, di dapat atribut paling berpengaruh yaitu Jumlah Art, Jumlah Kamar, Bb Masak, Ada Lemari Es, Ada Motor. Oleh karena itu untuk kita ketahui beberapa atribut paling mempengaruhi dalam pemilihan bantuan KIP supaya didapatkannya nilai akurasi dari pemodelan pohon keputusan memakai algoritma C4.5 atau decision tree. Uji coba tersebut dapat menghasilkan pohon keputusan yang mana atribut Jumlah Art menjadi Atribut paling berpengaruh dengan tingkat keberhasilan Penerimaan KIP. Evaluasi ini memakai confusion matrix untuk diperoleh nilai akurasi sebesar 98.21%, presisi sebesar 98.21%, recall sebesar 99.48%.

**Kata Kunci:** Klasifikasi; C4.5; KIP; Feature Selection

**Abstract**—The Indonesian government's efforts to overcome poverty in Indonesia are through the Smart Indonesia Card (KIP) program which is carried out by the government in the form of providing assistance to underprivileged families. The main aim of distributing KIP assistance is to help send underprivileged children to continue their education, the difficulties found in receiving KIP are due to the large number of residents registering, as well as the data having several conditions, the limited time available in providing KIP by sub-district parties, the completion base is relatively low, therefore the provision of assistance must be right on target. Therefore, the aim of this research is to look for the most influential attributes in receiving KIP assistance in order to improve the results of the data verification process. After carrying out Feature Selection using Information Gain, the most influential attributes can be obtained. The influences are Number of Art, Number of Rooms, Cooking Room, Refrigerator, Motorbike. Therefore, we need to know some of the attributes that most influence the selection of KIP assistance so that we can get accuracy values from decision tree modeling using the C4.5 algorithm or decision tree. Test This experiment can produce a decision tree in which the Number of Art attribute is the most influential attribute with the success rate of KIP acceptance. This evaluation uses a confusion matrix to obtain an accuracy value of 98.21%, precision of 98.21%, recall of 99.48%.

**Keywords:** Classification; C4.5; KIP; Feature Selection

## 1. PENDAHULUAN

Salah satu data induk yang mencakup informasi mengenai pelayanan kesejahteraan sosial adalah Data Terpadu Kesejahteraan Sosial (DTKS), dalam penerimaan bantuan dan pemberdayaan masyarakat [1]. DTKS dijadikan salah satu data dalam menangani program penanganan fakir miskin dan penyelenggaraan kesejahteraan sosial, DTKS adalah basis data yang dikelola langsung oleh Kementerian Sosial (KEMENSOS) dan digunakan dalam menerapkan penerimaan program penanganan kemiskinan. Bentuk dari bantuan dari DTKS salah satunya adalah bantuan berupa Kartu Indonesia Pintar (KIP). Kelurahan pangkalan sesai merupakan salah satu yang termasuk dalam penerima KIP. Pernyataan dari Kemensos yang menyatakan bahwa salah satu bantuan yang terdaftar pada data DTKS adalah Program KIP. Pada saat wabah pandemi korona virus 2019 (COVID19) pemerintah juga menggunakan DTKS sebagai salah satu sumber data penetapan sasaran penerima bantuan untuk menanggulangi dampak COVID-19 [2]. Sebagai respons cepat terhadap dampak dari pandemic, manfaat dari program KIP juga di perluas(). Namun, dua kajian cepat yang dilakukan melalui Social Monitoring and Early Response Unit (SMERU) dari bulan April–Mei 2020 ditemukannya bahwasanya ketidaktuntutan data DTKS berdampak negatif terhadap efektivitas pelaksanaan program bantuan di masa awal pandemi COVID-19 [3].

Pendataan KIP dalam pemberian bantuan kepada anak-anak yang kurang mampu, harus tepat sasaran, agar anak-anak yang kurang mampu juga dapat melanjutkan Pendidikan ke jenjang yang lebih tinggi lagi [4][5]. Banyaknya antusias warga yang mendaftar mengakibatkan bagian dari pihak kelurahan pangkalan sesai mengalami kualahan dalam menerima bantuan. Keterbatasan waktu merupakan salah satu kendala dalam mendata atribut dan menentukan beberapa ketentuan parameter tertentu. Dibutuhkan proses pengolahan data yang menggunakan pemrosesan data mining, data mining adalah suatu proses ekstraksi pengetahuan yang menggunakan pola atau pengetahuan informasi penting pada data [6]. Pohon keputusan adalah macam prediksi yang

menggambarkan beberapa komponen berpengaruh dalam pengambilan keputusan dan memberikan suatu prediksi akhir pengambilan keputusan[7].Metode decision tree merupakan metode yang membantu algoritma dalam klasifikasikan data dari penduduk dan membandikan data dari penerima bantuan sosial tunai dan membantu kelurahan dipangkalan sesai dan dinas sosial dalam penentuan apakah penduduk layak untuk menerima KIP atau tidak mendapatkan KIP [8].

Berdasarkan penelitian lainnya dengan Judul Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar,telah dilakukan penelitian dalam penerimaan beasiswa KIP,data yang digunakan untuk pemilihan penerimaan KIP yang berdasarkan dari uji coba yang mereka dapat dari salah satu sekolah menengah,metode klasifikasi C4.5,yang menghasilkan akurasi yang sangat tinggi,akurasi yang di peroleh 90,24% dengan errorate 9,75% sehingga C4.5 bisa dibilang sebagai algoritma yang sangat baik dalam penelitian ini[9].penelitian yang sejenis juga dengan judul penerapan algoritma C4.5 atau decision tree dalam penerimaan bantuan COVID-19 : sembako dan uang tunai.bentuk dari penelitian ini menunjukkan rule bahwa algoritma decision tree perbandingan 90 dan 10 memiliki tingkat akurasi yang baik ,yaitu 79,54% dengan menggunakan metode ini,proses bantuan yang diberikan lebih baik dan akurat dalam membantu pemerintah dalam memberikan bantuan selama pandemi COVID 19[10].

Penelitian yang sejenis juga dengan judul penerapan algoritma C4.5 untuk penentuan penerima beasiswa mahasiswa,disini data diklasifikasi dengan data mining sehingga mendapatkan hasil pemberian beasiswa untuk mahasiswa,penelitiannya dibutuhkan 3 kategori penelian untuk menentukan mahasiswa layak menerima bantuan pendidikan,berupa pemasukan orang tua,kewajiban,dan angka indeks presisi(ip) dari semester 3.dengan implementasikan c4.5 pada aplikasi penerima beasiswa menghasilkan pohon keputusan pada penerima bantuan pendidikan yang berdasarkan dataset beasiswa dengan 3 kriteria yang ada di atas agar mendapatkan hasil yang akurat[11]. penelitian sejenis juga penerapan algoritma decision tree untuk klasifikasi seleksi penerima bantuan langsung di desa ciomas,penelitian ini menggunakan 10 variabel untuk melakukan penentuan penerimaan bantuan tidak mampu,pengujiannya dilakukan dengan menggunakan angket,hasil dari angket tersebut menunjukkan kelayakan dari semua aspek penerima bantuan adalah 74,9% bisa dikatakan layak[12].

Penelitian yang sejenis juga dengan judul penerapan decision tree sebagai seleksi fitur dan support vector mechine(svm).hasil penelitian ini menunjukkan penggabungan antara C4.5 dan SVM mendapatkan hasil yang akurat dalam diagnosa permasalahan kanker payudara.penggunaan decision tree sebagai tahap awal seleksi fitur dalam memilih fitur-fitur dalam berkontribusi dalam diagnose kanker payudara,dan selanjutnya SVM memberikan analisis dan akurasi mendalam untuk memprediksi seorang pasien menderita kanker payudara[13].

Penelitian yang sejenis juga dengan judul information gain pada algoritma C4.5 untuk klasifikasi penerimaan bantuan pangan non tunai (BPNT),terdapat suatu masalah yang kerap terjadi dilapangan membuat lambatnya pemberian bantuan,algoritma yang digunakan adalah algoritma C4.5 menggunakan information gain seleksi fitur,ini merupakan salah satu algoritma terbaik,data set yang didapat sebanyak 130 data,penelitian ini memiliki model decision tree yang diperoleh dengan uji coba C4.5 mendapatkan akurasi 91.54% dan AUC 0.986,melalui ini penerima bantuan BNPT memiliki range algoritma sangat baik dan dapat memudahkan pengguna[14].

Melihat dari penelitian sebelumnya,penelitian kali ini akan menambahkan seleksi fitur information gain menggunakan algoritma C4.5 dalam klasifikasi data pada penerimaan KIP.[15][16]penambahan seleksi fitur pada penelitian ini disebabkan karena banyaknya atribut dan data yang dipakai dalam pemilihan calon penerima KIP.Algoritma C4.5 atau decision tree dipilih karena memiliki kemampuan keakuratan dari penelitian yang di lakukan[17][18].Algoritma C4.5 dapat diartikan sebagai algoritma pembelajaran yang sederhana untuk membuat pohon keputusan atau proses merubah bentuk data menjadi pohon, merubah pohon menjadi rule, dan menseederhanakan rule. Algoritma ini dapat mengklasifikasikan data secara akurat dan information gain untuk menentukan atribut mana yang akan menjadi akar pertama atau atribut yang paling berpengaruh, dan algoritma C4.5 dapat mengatasi hal seperti missing value ,continue data,bahkan pruning data[19][20].

## 2. METODOLOGI PENELITIAN

Penjelasan algoritma C4.5 menggunakan Feature Selection untuk mendapatkan tingkat kinerja yang tinggi untuk melakukan penelitian ini penerapan decision tree dengan penggunaan Feature selection,mempunyai ketelitian cukup tinggi dalam menghitung hal yang sederhana.dibawah adalah proses untuk hasil dari tahapan-tahapan penlitian.



**Gambar 1.** Flowchart Tahapan Penelitian

Beberapa tahapan yang terdapat pada flowchart diatas:

### 2.1 Identifikasi Masalah

Identifikasi masalah adalah salah satu hal penting dalam penelitian.mengidentifikasi dan menjelaskan adanya suatu masalah.pada penelitian ini sebagai penulis kita mencari informasi mengenai topik yang akan dibahas.pengidentifikasi masalah yang tepat dapat membantu penuliis untuk mengembangkan suatu rencana dari penelitian

### 2.2 Pengumpulan Data

Mengumpulkan data adalah proses penting dalam suatu penelitian dalam pengumpulan dari bentuk informasi yang kita perlukan untuk melakukan suatu penelitian terkait.dalam penelitian ,data yang di peroleh dari sistem tanya jawab kita lakukan secara langsung atau wawancara dengan narasumber dari dinas sosial kota dumai,provinsi riau,data yang di ambil berasal dari data terpadu kesejahteraan sosial (DTKS) kelurahan pangkalan sesai.

### 2.3 Analisis.

#### 2.3.1 Data Selection

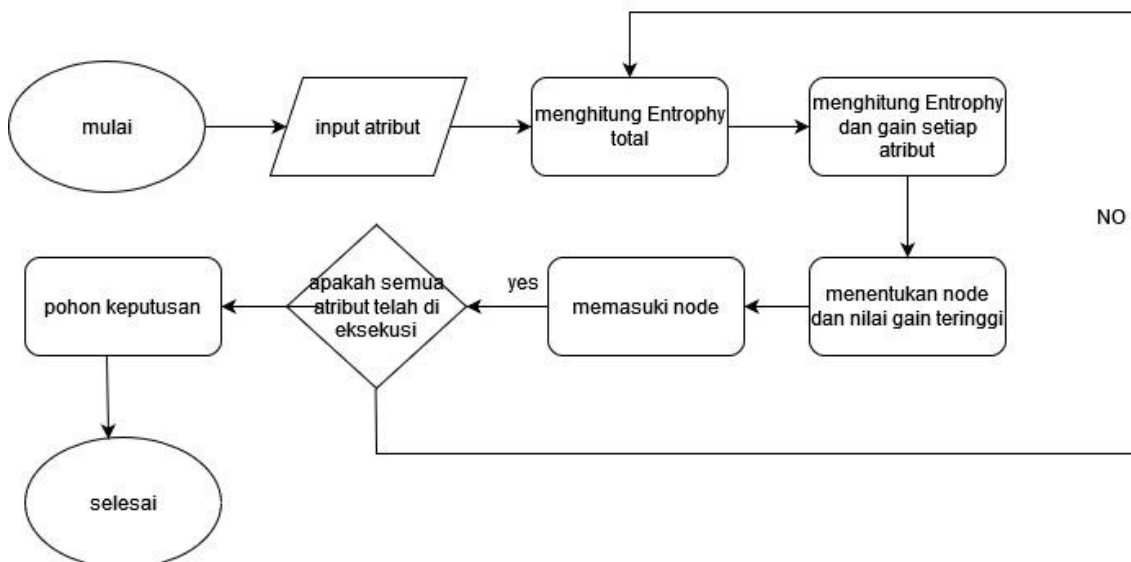
lanjut pengumpulan data,tahapan selanjutnya kita dapatkan mengerjakan data selection atau seleksi data.seleksi data adalah tahapan pemilihan Sebagian dari atribut atau fitur dari dataset.tujuan nya adalah untuk mengurangi atribut yang tidak relevan,berlebihan,atau yang memiliki potensi yang sangat rendah terhadap tujuan dari penelitian.

#### 2.3.2 Data Preprocessing

Praprocessing data yaitu tahapan awal dari analisis data untuk membersihkan data dan mempersiapkan data sebelum dilakukannya proses analisis lebih lanjut.salah satu Langkah dalam praproses data adalah data cleaning dilakukan dengan cara menghapus baris data yang berisi nilai null atau missing value.menghapus nilai null dapat diproses dengan tools rapid miner.operator ini memungkinkan untuk melakukan pemfilteran atau seleksi baris data berdasarkan kondisi tertentu.

#### 2.3.4 Perhitungan Algoritma C4.5

Setelah data berhasil dikumpulkan pada proses dalam mengumpulkan data dan praprocessing data,Langkah selanjutnya adalah tahapan data menggunakan algoritma C4.5.proses tersebut bis akita lihat pada gambar flowchart berikut (gambar 2).



**Gambar 2** .Flowchart Algoritma

Dibawah ini adalah rumus dan proses bagaimana menghitung algoritma C4.5 didasarkan dari gambar 2 di atas.

- penginputan bentuk atribut yang telah di praprocessing
- penghitungan entrophy nilai menggunakan rumus seperti dibawah :

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

Keterangan

S = himpunan kasus

n = Jumlah Pratisi S

Pi = Proporsi Si Terhadap S

- c. Menghitung nilai dari gain masing-masing atribut dengan rumus yang dapat dilihat dibawah ini :

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

Penjelasan :

S = Himpunan Kasus

A = Fitur

n = Jumlah Partisi Atribut A

|S<sub>i</sub>| = Proporsi S<sub>i</sub> Terhadap S

|S| = Jumlah Kasus Dalam S

- d. Dalam membangun suatu pohon keputusan, langkah pertama adalah mendefinisikan node yang paling tinggi (root node).node paling besar adalah titik awal dari pohon keputusan, yang jadi akar pertama dari pohon keputusan.
- e. Dalam pembuatan model decision tree atau C4.5.proses perulangan dilakukan untuk membangun pohon keputusan secara bertahap.pada setiap perulangan, algoritma akan mengecek apakah semua atribut sudah di proses dengan benar atau digunakan sebagai pemisah pada cabang dari pohon keputusan.jika semua atribut telah berhasil di proses, maka model decision tree akan selesai dibuat.
- f. Model dari yang dihasilkan dari decision tree atau C4.5 adalah berupa keterangan, rule dan pola nantinya bakal berguna bagi desa.

### 2.4 pengujian

Tahapan pengujian ini semuanya dapat di uji menggunakan cara manual dan tools dari rapid miner dan algoritma C4.5.

### 2.5 Evaluasi

Kinerjanya dapat kita evaluasikan menggunakan confusion matrix yang terdapat tools rapid miner untuk kita menghitung nilai akurasi, presisi dan recall..

### 2.6 Kesimpulan

Kesimpulan yang kita peroleh dari pemodelan algoritma C4.5 ini kita dapat mengetahui output yang dihasilkan dari pohon keputusannya dan nilai akurasi, presisi, serta nilai recall.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pengumpulan Data

Penelitian ini, menggunakan data yang bersumber dari data DTKS,pada kelurahan pangkalan sesai kota Dumai data tersebut merupakan data calon penerima bantuan KIP data yang didapatkan dalam bentuk format Microsoft excel berisi berbagai atribut yang relevan dengan verifikasi dan validasi DTKS penduduk yang bertempat di kelurahan pangkalan sesai kota dumai dari tabel 2 dibawah ini kita dapat mengetahui atribut data yang didapat dari DTKS

**Tabel 1.** Penjelasan Atribut Data

Atribut	Kode	Keterangan	Deskripsi
Jumlah Art	A1	Jumlah anggota rumah tangga	tangga Menyesuaikan masing-masing anggota keluarga pada setiap rumah.
Sta Bangunan	A2	Status penguasaan bangunan tempat tinggal yang ditempati	1. Milik sendiri 2. Kontrak/sewa 3. Bebas sewa 4. Dinas 5. Lainnya
Sta Lahan	A3	Status lahan tempat tinggal yang ditempati	1. Milik sendiri 2. Milik orang lain 3. Tanah negara 4. Lainnya
Lantai	A4	Jenis lantai terluas	1. Marmmergranit 2. Keramik 3. Parket/vinil/permadani 4. Ubin/tegel/teraso 5. Kayu/papan kualitas tinggi 6.Semestara/bata 7. Bambu 8.Kayu/papan 9. Tanah 10. Lainnya
Dinding	A5	Jenis dinding terluas	1. Tembok 2. Plasteran anyaman bambu 3. Kayu 4. Anyaman bambu 5. Batang Kayu 6. Bambu 7. Lainnya
Kondisi lantai	A6	Kondisi dinding terluas	1. Bagus/kualitas tinggi 2. Jelek/kualitas rendah

Atribut	Kode	Keterangan	Deskripsi
Atap	A7	Jenis atap terluas	1. Beton/genteng beton 2. Genteng keramik 3. Genteng metal 4. Genteng tanah liat 5. Asbes 6. Seng 7. Sirap 8. Bambu 9. Jerami/ijuk/daun 10. Lainnya
Kondisi Atap	A8	Kondisi atap terluas	1. Bagus/kualitas tinggi 2. Jelek/kualitas rendah
Jumlah Kamar	A9	Jumlah kamar tidur	Menyesuaikan jumlah kamar pada masing-masing rumah.
Sumber Air Minum	A10	Sumber air minum	1. Air kemasan bermerk 2. Air isi ulang 3. Leding meteran 4. Leding eceran 5. Sumur bor/pompa 6. Sumur terlindung 7. Sumur tak terlindung 8. Mata air terlindung 9. Mata air tak terlindung 10. Airsungai/danau/waduk 11. Air hujan 12. Lainnya
Sumber Penerangan	A11	Sumber penerangan utama	1. Listrik PLN 2. Listrik non PLN 3. Bukan listrik
Daya	A12	Daya terpasang	1.) 450 watt 2.) 900 watt 3.) 1.300 watt 4.) 2.200 watt 5.) > 2.200 watt 6.) Tanpa meteran
BB masak	A13	Bahan bakar/energi utama untuk memasak	1. Listrik 2. Gas > 3 kg 3. Gas 3 kg 4. Gas kota/biogas 5. Minyak tanah 6. Briket 7. Arang 8. Kayu bakar 9. Tidak memasak dirumah
Fasbab	A14	Penggunaan fasilitas tempat BAB	1. Sendiri 2. Bersama 3. Tidak ada
Kloset	A15	Jenis kloset	1. Leher angsa 2. Plengsengan 3. Cemplung/cubluk 4. Tidak pakai
Buang Tinja	A16	Tempat pembuangan akhir tinja	1. Tangki 2. Plengsengan 3. Lubang tanah 4. Kolam/sawah/sungai/danau/laut 5. Pantai/tanah lapang/kebun 6. Lainnya
Tabung gas	A17	Memiliki tabung gas 5,5 kg atau lebih	1. Ya 2. Tidak
Lemari es	A18	Memiliki lemari es/kulkas	1. Ya 2. Tidak
AC	A19	Memiliki AC	1. Ya 2. Tidak
Pemanas Air	A20	Memiliki pemanas air (water heater)	1. Ya 2. Tidak
Telepon	A21	Memiliki telepon rumah (PSTN)	1. Ya 2. Tidak
Tv	A22	Memiliki televisi	1. Ya 2. Tidak
Emas	A23	Memiliki emas/perhiasan & tabungan (senilai 10 gram 10 gram emas)	1. Ya 2. Tidak
Komputer/laptop	A24	Memiliki komputer/laptop	1. Ya 2. Tidak
Sepeda	A25	Memiliki sepeda	1. Ya 2. Tidak
Motor	A26	Memiliki sepeda motor	1. Ya 2. Tidak
Mobil	A27	Memiliki mobil	1. Ya 2. Tidak
Perahu	A28	Memiliki perahu	1. Ya 2. Tidak
Motor Tempel	A29	Memiliki motor tempel	1. Ya 2. Tidak
Kapal	A30	Memiliki kapal	1. Ya 2. Tidak
Aset Tidak Bergerak	A31	Rumah tangga memiliki aset tidak bergerak	a. Lahan 1. Ya..... 2. Tidak b. Rumah ditempat lain 3. Ya 4. Tidak
Rumah lain	A32	Rumah tangga memiliki aset rumah di tempat lain	1. Ya 2. Tidak
Stat Art Usaha	A33	Apakah ada ART yang memiliki usaha sendiri/milik bersama	1. Ya 2. Tidak
Status	Label atau kelas	Jika YES menerima KIP Jika NO tidak menerima KIP	YES ATAU NO

Pada tabel 1 diatas kita di perlihatkan atribut data penerima kartu Indonesia pintar (KIP)

**Tabel 2.** sampel data KIP

N0	A1	A2	A3	A4	A5	...	A29	A30	A31	A32	A33	Status
D1	4	2	2	2	1	...	2	2	2	4	0	YES
D2	4	1	1	2	1	...	2	2	1	4	2	YES
D3	6	2	2	6	1	...	2	2	2	4	2	NO
D4	1	NULL	NULL	NULL	NULL	...	NULL	NULL	NULL	NULL	NULL	NO
D5	6	1	1	2	1	...	2	2	1	4	2	NO
D6	1	NULL	NULL	NULL	NULL	...	NULL	NULL	NULL	NULL	NULL	YES
D7	3	1	1	2	1	...	2	2	2	4	0	YES
D8	4	1	1	6	1	...	2	2	1	4	2	NO
D9	4	1	1	2	1	...	2	2	1	4	2	YES
D10	4	2	2	2	1	...	2	2	2	4	0	YES
...	...	...	...	...	...	...	...	...	...	...	...	...
D1998	2	3	2	4	1	...	2	2	1	4	1	NO
D1999	2	3	2	4	1	...	2	2	2	4	2	YES

### 3.2 Selection fitur information gain dan Data Preprocessing

Sebelum kita melakukan penelitian menerapkan Algoritma C4.5 atau decision tree,peneliti menjalankan tahapan data selection fitur information gain,cleaning data dan data praprocessing data sesuai dengan permasalahan yang di dapat.data yang telah melewati tahapan selection data dan praprocessing data dipaparkan pada tabel 3 dibawah ini.

**Tabel 3.** Data Clening

No	A1	A2	A3	A4	...	A30	A31	A32	A33	Status
D1	4	2	2	2	...	4	2	4	0	YES
D2	4	1	1	2	...	4	1	4	2	YES
D3	6	2	2	6	...	4	2	4	2	NO
D4	6	1	1	2	...	4	1	4	2	NO
D5	3	1	1	2	...	4	2	4	0	NO
D6	4	1	1	6	...	4	1	4	2	YES
D7	4	1	1	2	...	4	1	4	2	YES
D8	2	1	1	2	...	4	1	4	1	NO
D9	4	2	2	2	...	4	2	4	0	YES
...	...	...	...	...	...	...	...	...	...	...
D1674	7	1	1	6	...	4	1	4	1	NO
D1675	4	1	1	8	...	4	2	4	2	YES

Sistem yang digunakan dalam mengklasifikasi tingkat pendataan KIP,kita dapat memakai data yang numeric dan non-numeric.pada saat data telah dikumpulkan,lanjut proses penerapan algoritma C4.5 untuk memproses data,dan dihasilkannya output berupa pohon keputusan dan penilaian akurasi,presisi,dan recall.dibawah ini dapat kita lihat langkah pengolahan data menggunakan algoritma C4.5.

a. Menentukan nilai Entrophy dan Information Gain

Dalam pemrosesan decision tree dibutuhkan perhitungan entrophy dan nilai gain.nilai gain yang terbesar akan menjadikannya root node decision tree.perhitungan entrophy yang dijalankan dalam mencari kategori atribut,dibawah ini bentuk dari contoh perhitungan yang di hasilkan menggunakan algoritma C4.5. Mencari nilai Entrophy penelitian ini, disini kita memiliki 2 label,maka kita akan menggunakan rumus log2. Adapun cara pengerjaannya dapat dilihat dibawah ini :

Total kasus : 1675

Jumlah penerima KIP : 384

Jumlah tidak penerima KIP :1291

$$\text{Entrophy}(\text{total}) = \left(-\frac{384}{1675} \times \log_2\left(\frac{384}{1675}\right)\right) + \left(-\frac{1291}{1675} \times \log_2\left(\frac{1291}{1675}\right)\right) = 0,77670812 \quad (3)$$

Setelah kita mendapatkan nilai dari entrophy nya,kita masukkan ke tahapan seleksi fiturnya,tahapan ini menggunakan tools dari rapid miner untuk mencari seleksi fitur information gainnya, hasilnya dapat kita lihat dari tabel 4 dan 5 dibawah ini

**Tabel 4.** Nilai Gain Atribut

Atribut	Nilai gain
A1	0,195
A9	0,013

Atribut	Nilai gain
A26	0,011
A18	0,009
A13	0,007
...	...
A21	0.00

Setelah dilakukan seleksi fitur information gain menggunakan tool ramid miner keseluruhan atribut pada data, maka kita dapatkan 5 atribut tertinggi yang akan kita pilih untuk perhitungan C4.5, bisa kita lihat pada tabel 5 dibawah ini.

**Tabel 5** Data selection feature

NO	A1	A9	A13	A18	26	Status
D1	4	2	3	3	3	YES
D2	4	3	3	3	3	YES
D3	6	1	8	3	4	NO
...	...	...	...	...	...	...
1674	7	4	3	4	3	NO
1675	4	2	2	3	3	YES

### 3.3 Pengolahan data

Setelah menghitung Entropy dari setiap masing-masing atribut maka, selanjutnya kita mencari nilai gain dari setiap atribut. dibawah ini adalah cara pengerjaan dalam mencari nilai gain.

Entropy total : 0,77670812

Entropy jumlah art 1 : 0,135853315

Entropy jumlah art 2 : 0,113737206

Entropy jumlah art 3 : 0,117594666

Entropy jumlah art 4 : 0

Entropy jumlah art 5 : 0,143262593

Entropy jumlah art 6 : 0,04619244

Entropy jumlah art 7 : 0,110580845

Entropy jumlah art 8 : 0

Entropy jumlah art 9 : 0

Entropy jumlah art 10 : 0

$$\begin{aligned}
 \text{Gain (Total.jumlah art)} &= 0,77670812 = \left( \left( \frac{263}{1675} \times 0,1358533 \right) + \left( \frac{197}{1675} \times 0,1137372 \right) + \left( \frac{315}{1675} \times 0,1175946 \right) + \right. \\
 &\left( \frac{364}{1675} \times 0 \right) + \left( \frac{246}{1675} \times 0,1432625 \right) + \left( \frac{196}{1675} \times 0,0461924 \right) + \left( \frac{68}{1675} \times 0,11058084 \right) + \left( \frac{14}{1675} \times 0 \right) + \left( \frac{9}{1675} \times \right. \\
 &\left. 0 \right) + \left( \frac{3}{1675} \times 0 \right) = 0,688950636 \tag{4}
 \end{aligned}$$

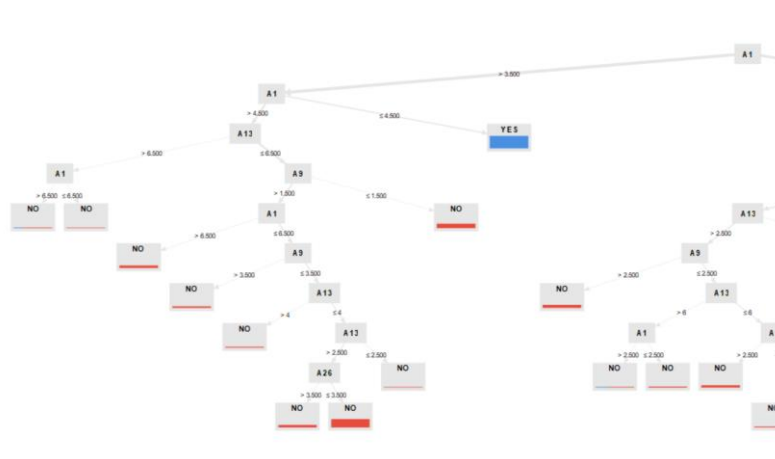
Pengerjaan dari pemilihan atribut yang diatas dapat kita ringkas secara rinci dari tabel 5 dibawah ini.

**Tabel 6.** Presentase Hasil perhitungan Entropy dan Gain

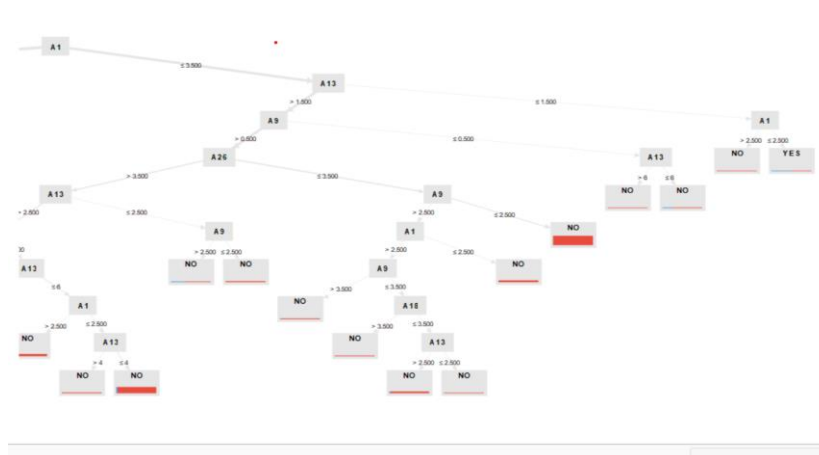
Node	Jumlah kasus	YES	NO	Entropy	Gain
1 Total	1675	364	1291	0,77670812	
A1					<b>0,688950636</b>
1	263	258	5	0,135853315	
2	197	194	3	0,113737206	
3	315	310	5	0,117594666	
4	364	0	364	0	
5	246	241	5	0,143262593	
6	196	195	1	0,04619244	
7	68	67	1	0,110580845	
8	14	14	0	0	
9	9	9	0	0	
10	3	3	0	0	
A9					0,014118747
0	7	6	1	0,591672779	
1	426	319	107	0,813133115	
2	832	630	202	0,799646724	
3	353	279	74	0,740778212	

Node	Jumlah kasus	YES	NO	Entropy	Gain
	4	40	0	0	
	5	17	0	0	
A13					0,008604507
	1	7	6	1	0,591672779
	2	80	59	21	0,830495523
	3	1465	1128	337	0,778070795
	5	74	51	23	0,894107184
	7	4	3	1	0,811278124
	8	39	38	1	0,172036949
	9	6	6	0	0
A18					0,008859268
	3	927	676	251	0,842561525
	4	748	615	133	0,675257076
A26					0,085391766
	3	1104	810	294	0,836111478
	4	571	187	384	0,912347173

Hasil berdasarkan dari tabel 6 diatas menghitung entropy dan nilai gain yang dapat kita lihat pada tabel 5 diatas agar atribut dari A1 mempunyai nilai gain yang lebih tinggi,yaitu 0,688950636,sebab dari itu A1 terpilih untuk menjadi akar dari decision tree atau pohon keputusan yang akan dibuat.tahapan berikutnya ialah dengan cara membuat perhitungan sama dengan sebelumnya,menggunakan persamaan entropy,untuk ,kita mencari nilai dari entropy tersebut dan persamaan nilai dari gain yang mana untuk memperoleh nilai gain,maka kita dapat menghasilkan pohon keputusan atau decision tree,ada beberapa aturan dan rule pada pohon keputusan atau decision tree dapat kita lihat dari gambar dibawah 3 dan 4 dibawah ini



**Gambar 3.** Decision Tree (terlihat dari kiri)



**Gambar 4.** Decision Tree (terlihat dari kanan)

Hasil yang diperoleh dari menghitung pohon keputusan pada gambar 3 dan 4,beberapa perhitungan yang di hasilkan.dibawah ini adalah aturan yang di hasilkan pada gambar diatas.

- a. Jika  $A1 > 3.500, A1 \leq 4.500$  maka yes

- b. Jika  $A1 > 3.500, A1 > 4.500, A13 > 6.500, A1 \leq 6.500$  maka NO
- c. Jika  $A1 \leq 3.500, A13 > 1.500, A9 > 0,500, A26 \leq 3.500, A9 \leq 2.500$  maka NO
- d. jika  $A1 \leq 3.500, A13 > 1.500, A9 \leq 0,500, A13 \leq 6$  maka NO
- e. jika  $A1 \leq 3.500, A13 > 1.500, A9 > 0,500, A26 > 3.500, A13 \leq 2.500, A9 \leq 2500$  maka NO
- f. jika  $A1 \leq 3.500, A13 > 1.500, A9 \leq 0,500, A13 > 6$  maka NO
- g. jika  $A1 > 3.500, A1 > 4.500, A13 \leq 6.500, A9 \leq 1.500$  maka NO
- h. jika  $A1 \leq 3.500, A13 \leq 1.500, A1 \leq 2.500$  maka YES
- a. jika  $A1 \leq 3.500, A13 > 1.500, A9 > 0.500, A26 \leq 3.500, A9 > 2.500, A1 \leq 2500$  maka NO
- i. jika  $A1 \leq 3.500, A13 > 1.500, A9 > 0.500, A26 > 3.500, A13 > 2.500, A9 > 2.500$  maka NO
- j. K. Jika  $A1 > 3.500, A1 > 4.500, A13 > 6.500, A1 > 6.500$  maka NO
- k. l. jika  $A1 \leq 3.500, A13 > 1.500, A9 > 0,500, A26 > 3.500, A13 \leq 2.500, A9 > 2500$  maka NO

### 3.4 Evaluasi

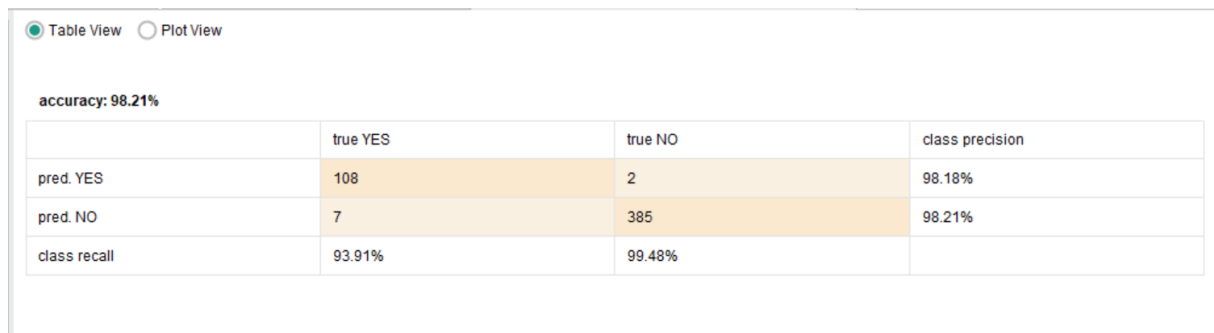


Table View Plot View

accuracy: 98.21%

	true YES	true NO	class precision
pred. YES	108	2	98.18%
pred. NO	7	385	98.21%
class recall	93.91%	99.48%	

**Gambar 5** nilai dari presisi, recal dan akurasi

Pada gambar 5 diatas kita masuk kedalam tahapan evaluasi yang memanfaatkan confusion matrix pada tools rapid miner agar kita dapat menghitung nilai dari presisi,recall,dan akurasi menggunakan pembagian data latih dan uji sebesar 70:30 pada gambar di atas menghasilkan nilai dari akurasi 98.21%,pada nilai presisi YES 98.18%,presis NO 98.21%,serta untuk nila recall dari YES 93.91% dan NO 99.48% dengan hasil yang ditimbulkan disimpulkan penggunaan dari algoritma C4.5 mampu mengklasifikasikan dengan baik tingkat kesuksesan penerimaan KIP di pangkalan sesai.

## 4. KESIMPULAN

Dengan demikian penelitian ini didapatkan pada penelitian penerimaan KIP pada pangkalan sesai di atas pada penelitian ini memberikan kemudahan dalam meningkatkan proses verifikasi data untuk menyalurkan bantuan KIP di wilayah pangkalan sesai. atribut yang paling memperngaruhi nya adalah A1 atau ART dari 1675 data awal yang diuji menggunakan perbandingan 70% data uji dan 30% data latih diperoleh lah akurasi sebesar 98.21%,nilai presisi sebesar 98.18% dan nilai recall 93.91% dengan demikian hasil dari algoritma C4.5 dapat bekerja sangat baik dalam menentukan penerimaan KIP dipangkalan sesai,dengan hasil akurasi sebesar 98.21% dari sini kita dapat menyimpulkan penggunaan algoritma C4.5 mempunyai akurasi yang sangat baik,sekiranya dapat dibandingkan juga dengan algoritma lain agar dapat menghasilkan hasil yang mungkin lebih baik lagi.

## REFERENCES

- [1] W. D. Wulandari, A. Afrisawati, and F. Dristyan, "Penentuan Kelayakan Penerima Kartu Indonesia Pintar (KIP) Menggunakan Metode AHP," *Journal Of Computer Science And Technology (JOCSTEC)*, vol. 1, no. 2, pp. 69–75, May 2023, doi: 10.59435/jocstec.v1i2.60.
- [2] G. Taufik and D. Jatmika, "Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang," vol. 6, no. 1, p. 2021.
- [3] E. E. Barito, ) Jap, T. Beng, and D. Arisandi, "PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA PENERIMA BANTUAN SOSIAL COVID-19," 2022. Accessed: Aug. 08, 2023. [Online]. Available: <https://journal.untar.ac.id/index.php/jiksi/article/view/17819>
- [4] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, "PENENTUAN PENERIMA BANTUAN BEASISWA KARTU INDONESIA PINTAR (KIP) KULIAH MENGGUNAKAN NAÏVE BAYES CLASSIFIER," *JSiI (Jurnal Sistem Informasi)*, vol. 9, no. 2, pp. 193–199, Sep. 2022, doi: 10.30656/jsi.v9i2.4376.
- [5] B. D. S. I. Rizky Haqmanullah Pambudi, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] N. H. Purnomo, B. Pamungkas, and C. Juliane, "JURNAL MEDIA INFORMATIKA BUDIDARMA Penerapan Algoritma C4.5 Untuk Klasifikasi Tren Pelanggaran Kendaraan Angkutan Barang dengan Metode CRISP-DM," vol. 7, pp. 30–40, 2023, doi: 10.30865/mib.v7i1.5247.
- [7] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," vol. 3, pp. 64–71, 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>

- [8] N. Wayan Oktha Pratiwi, N. Widya Utami, and I. Gede Juliana Eka Putra, “Klasifikasi Penentu Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 Di desa Keramas, Gianyar Bali,” *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, vol. 4, no. 3, pp. 101–107, 2022, Accessed: Jul. 25, 2023. [Online]. Available: <https://doi.org/10.51401/jinteks.v4i3.1667>
- [9] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, “Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP),” *Paradigma - Jurnal Komputer dan Informatika*, vol. 23, no. 2, Oct. 2021, doi: 10.31294/p.v23i2.11395.
- [10] A. Riadi, S. Retno Andani, R. Rafiqah Dewi, S. Tunas Bangsa, and J. A. Jend Sudirman Blok No, “Data Mining Algoritma C4.5 menentukan Kelayakan Penerima Bantuan Covid 19,” vol. 3, 2021.
- [11] E. Hasmin and S. Aisa, “Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa Mahasiswa Application of C4.5 Algorithm For Determining Student Scholarship Recipients,” *Cogito Smart Journal* |, vol. 5, no. 2, 2019, Accessed: Aug. 07, 2023. [Online]. Available: <https://cogito.unklab.ac.id/index.php/cogito/article/view/219>
- [12] A. A. F. Irmayansyah, “penerapan algoritma c4.5 klasifikasi penerima bantuan,” vol. 8, 2018, Accessed: Jul. 25, 2023. [Online]. Available: <https://media.neliti.com/media/publications/288963-penerapan-algoritma-c45-untuk-klasifikasi-d7417740.pdf>
- [13] P. Riswanto and A. Aziz, “PENERAPAN DECISION TREE C4.5 SEBAGAI SELEKSI FITUR DAN SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA KANKER PAYUDARA,” 2019. Accessed: Aug. 01, 2023. [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/JurnalInformatika/article/view/1442/pdf>
- [14] R. A. Saputra, S. Wasiantyanti, and D. Pribadi, “Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (Bpnt),” *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 4, no. 1, p. 25, 2021, doi: 10.21927/ijubi.v4i1.1757.
- [15] E. Sufarnap, Analisis Seleksi Fitur dengan Menggunakan Klasifikasi C4.5 dan Density Based Feature Selection (DBFS) dalam Memprediksi Kelulusan Mahasiswa, vol. 1–3. 2018. Accessed: Aug. 01, 2023. [Online]. Available: [https://citisee.amikompurwokerto.ac.id/assets/proceedings/paper/73\\_\\_Proceeding\\_CITISEE\\_2018\\_TI10.pdf](https://citisee.amikompurwokerto.ac.id/assets/proceedings/paper/73__Proceeding_CITISEE_2018_TI10.pdf)
- [16] L. Bachtiar and M. Mahradianur, “Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai,” *Jurnal Informatika*, vol. 10, no. 1, pp. 28–36, Mar. 2023, doi: 10.31294/inf.v10i1.15115.
- [17] S. F. Damanik, A. Wanto, and I. Gunawan, “Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok,” *Volume 1 Nomor 2, Januari 2022, 2022.* <https://ejournal.catuspata.com/index.php/jkdn/index>
- [18] yuyun,rabiatul adawiyah mansyur, “Penerapan Algoritma c4.5 Untuk Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial Di Kabupaten Bulukumba,” 2019.
- [19] dony oscar hendri, “Penerapan Algoritma C4.5 Dalam Mengukur Kepuasan Pengunjung,” 2021. [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infotech73>
- [20] S. W. Siahaan, K. D. R. Sianipar, P. P. P. A. N. W. F. I. R.H Zer, and D. Hartama, “Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa,” *PETIR*, vol. 13, no. 2, pp. 229–239, Sep. 2020, doi: 10.33322/petir.v13i2.1029.