

# Implementasi Algoritma C4.5 dalam Klasifikasi Calon Penerima Program Keluarga Harapan (PKH) di Kelurahan Bah Sorma

*by* Vina Tri Putri Agil Purba

---

**Submission date:** 29-May-2024 01:41PM (UTC+0700)

**Submission ID:** 2390611983

**File name:** MERKURIUS\_VOL\_2\_NO\_3\_MEI\_2024\_HAL\_72-97.pdf (619.57K)

**Word count:** 8688

**Character count:** 47673



## Implementasi Algoritma C4.5 dalam Klasifikasi Calon Penerima Program Keluarga Harapan (PKH) di Kelurahan Bah Sorma

60

Vina Tri Putri Agil Purba

Program Studi Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya

Email: [agilpurba82@gmail.com](mailto:agilpurba82@gmail.com)

Fitriyanti Fitriyani

Program Studi Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya

Email: [fitriyani@ars.ac.id](mailto:fitriyani@ars.ac.id)

Jl. Sekolah Internasional No.1-2 Antap, Bandung

Korespondensi penulis: [agilpurba82@gmail.com](mailto:agilpurba82@gmail.com)

**Abstract:** The Family Hope Program (PKH) is a program that provides attention to the community, especially the health category, education category and social welfare category for poor families. The Family Hope Program (PKH) aims to reduce poverty and improve the welfare of the Indonesian population. Due to the large number of residents who want to register themselves as PKH recipients, there are residents who manipulate data or claim to be poor people in order to get PKH. If this continues to happen, and there is no preventive action, it is not impossible that many residents are not right in receiving PKH provided by the Government. One of the efforts that can be made is to the classification of prospective PKH recipients in Bah Sorma Village. This study aims to classify prospective recipients of the Family Hope Program in Bah Sorma Village. The dataset used is data on prospective PKH recipients in Bah Sorma Village, Pematang Siantar City. This research is a comparative study of previous research using the Naïve Bayes method. The method used in this research is Data Mining with the C4.5 method which is used to see the accuracy of the best method than previous research. The accuracy result obtained by this research is 98.18%. Based on the results obtained, research with the case of classification of prospective PKH recipients in Bah Sorma Village using the C4.5 Algorithm gets better accuracy than previous research using Naïve Bayes obtaining an accuracy of 80%.

**Keywords:** Family Hope Program, Data Mining, Classification, Decision Tree, C4.5

**Abstrak:** Program Keluarga Harapan (PKH) merupakan program yang memberikan perhatian kepada masyarakat khususnya kategori kesehatan, kategori pendidikan dan kategori kesejahteraan sosial bagi keluarga miskin. Program Keluarga Harapan (PKH) bertujuan untuk penurunan angka kemiskinan dan peningkatan kesejahteraan penduduk Indonesia. Karena banyaknya warga yang mau mendaftarkan dirinya menjadi penerima PKH, sehingga banyak warga yang memanipulasi data atau mengaku menjadi masyarakat miskin demi mendapatkan PKH. Apabila hal ini terus terjadi, dan tidak ada tindakan pencegahan bukan hal yang mustahil banyak warga yang tidak tepat dalam menerima PKH yang diberikan Pemerintah. Salah satu upaya yang dapat dilakukan adalah melakukan uji klasifikasi calon penerima PKH di Kelurahan Bah Sorma. Penelitian ini bertujuan untuk melakukan klasifikasi penerima calon penerima Program Keluarga Harapan di Kelurahan Bah Sorma. Dataset yang digunakan adalah data calon penerima PKH di Kelurahan Bah Sorma, Kota Pematang Siantar. Penelitian ini merupakan penelitian perbandingan dari penelitian sebelumnya yang menggunakan metode Naïve Bayes. Metode yang digunakan pada penelitian ini adalah Data Mining dan metode C4.5 yang digunakan untuk melihat akurasi metode yang terbaik daripada penelitian sebelumnya. Hasil akurasi yang diperoleh penelitian ini sebesar 97,27%. Berdasarkan hasil yang didapatkan, penelitian dengan kasus klasifikasi calon penerima PKH di Kelurahan Bah Sorma dengan menggunakan Algoritma C4.5 mendapatkan akurasi yang lebih baik daripada penelitian sebelumnya dengan menggunakan Naïve Bayes memperoleh akurasi sebesar 80%.

**Kata Kunci:** Program Keluarga Harapan, Data Mining, Klasifikasi, Decision Tree, C4.5

### PENDAHULUAN

Bantuan sosial yang diberikan Pemerintah Indonesia merupakan salah satu trik dalam menuntaskan kemiskinan. Salah satu program dan kegiatan yang ditempuh Pemerintah Indonesia adalah Program Keluarga Harapan. Program Keluarga Harapan (PKH) merupakan

Received April 30, 2024; Accepted 29 Mei, 2024; Published Mei 31, 2024

\* Vina Tri Putri Agil Purba, [agilpurba82@gmail.com](mailto:agilpurba82@gmail.com)

program yang memberikan perhatian kepada masyarakat khususnya kategori kesehatan, kategori pendidikan dan kategori kesejahteraan sosial bagi keluarga miskin (Muttakin et al., 2021). Program Keluarga Harapan bertujuan untuk penurunan angka kemiskinan dan peningkatan kesejahteraan penduduk Indonesia. Hal ini semakin signifikan pada bulan Maret tahun 2021, karena jumlah keluarga miskin dan prasejahtera di Indonesia berjumlah 10,14% dari total keseluruhan penduduk atau sejumlah 27,54 juta jiwa (Sulaiman, 2022).

Berdasarkan tingkat signifikan penerima PKH banyak warga yang mau mendaftarkan dirinya menjadi calon penerima PKH. Salah satunya wilayah Kelurahan Bah Sorma, yang berada di Kecamatan Siantar Sitalasari Kota Pematang Siantar menjadi salah satu wilayah yang dimana sebagian warganya penerima bantuan PKH dari Pemerintah Indonesia. Karena banyaknya warga yang mau mendaftarkan dirinya menjadi penerima PKH membuat pegawai Kelurahan Bah Sorma harus berhati-hati dalam mengambil keputusan untuk menentukan warganya yang mendaftarkan penerima PKH untuk diterima layak atau tidak. Sehingga terdapat warga yang memanipulasi data atau mengaku menjadi masyarakat miskin demi mendapatkan PKH. Apabila hal ini terus terjadi, dan tidak ada tindakan pencegahan bukan hal yang mustahil banyak warga yang tidak tepat dalam menerima PKH yang diberikan Pemerintah. Salah satu upaya adalah dengan melakukan analisa klasifikasi kelayakan calon penerima PKH di Kelurahan Bah Sorma.

Oleh karena itu, salah satu cara untuk mengetahui kelayakan calon penerima PKH adalah dengan melakukan klasifikasi calon penerima PKH. Proses yang dilakukan tidaklah mudah, dibutuhkan tahapan-tahapan dari permasalahan tersebut. Adapun penelitian ini merupakan uji banding dan sekaligus perkembangan dari penelitian sebelumnya yang dilakukan oleh (Sulaiman, 2022). Penelitian ini menggunakan algoritma yang berbeda dari penelitian sebelumnya, yaitu menggunakan Data Mining dengan Algoritma C4.5 (R.H. Zer et al., 2022). Data penelitian ini adalah data yang digunakan penelitian sebelumnya yaitu calon penerima PKH tahun 2021.

Beberapa penelitian yang menjadi rujukan penelitian ini, diantaranya penelitian sebelumnya dalam mengklasifikasi calon penerima PKH di Kelurahan Bah Sorma dengan menggunakan Algoritma C4.5. Dataset yang digunakan penelitian tersebut adalah data calon penerima PKH di Kelurahan Bah Sorma tahun 2021. Hasil penelitian tersebut menghasilkan nilai akurasi sebesar 80% dengan Prediksi Tidak dengan label Tidak berjumlah 4 dan label Iya berjumlah 2 sehingga nilai class precision sebesar 56,67%. Prediksi Iya dengan label Tidak berjumlah 0 dan label Iya berjumlah 4 sehingga class precision sebesar 100% (Sulaiman, 2022). Penelitian berikutnya adalah dalam menganalisa kelayakan penerima Program Keluarga

Harapan (PKH) dengan menggunakan Algoritma C4.5. Dataset yang digunakan adalah data keluarga yang menerima PKH sebanyak 100 sampel data. Hasil penelitian tersebut adalah Algoritma C4.5 dapat digunakan untuk mengklasifikasi masyarakat yang berhak dan tidak berhak mendapatkan PKH dengan tingkat akurasi 96,67% dengan hasil rules sebanyak 6 rules (M. A. Tanjung et al., 2021).

Penelitian berikutnya adalah klasifikasi penerima bantuan Program Keluarga Harapan dengan menggunakan Algoritma C4.5 dengan studi kasus di Desa Mlirip Kabupaten Mojokerto. Dataset yang digunakan penelitian tersebut adalah data yang diperoleh dari Pemerintah Desa Mlirip Kabupaten Mojokerto dengan melakukan penyebaran kuesioner. Hasil penelitian tersebut adalah menghasilkan *accuracy* sebesar 71,1 %, *precision* 90,9%, *recall* 70,4%, dan *f1-score* 79,3%. Terdapat ketidakseimbangan data hasil prediksi dan kesalahan dalam pelabelan. Hal itu disebabkan karena atribut kepemilikan anak sekolah memiliki pengaruh yang sangat besar pada hasil klasifikasi sehingga kemungkinan terjadi salah prediksi dan ketimpangan pada setiap kriteria yang ada (Nella et al., 2022).

Adapun penelitian selanjutnya prediksi kelulusan Mahasiswa Prodi Informatika dengan menggunakan Algoritma C4.5. Dataset yang digunakan penelitian tersebut adalah data tentang mahasiswa yang lulus Prodi Informatika dari Januari 2013 sampai Agustus 2017. Hasil penelitian tersebut adalah penerapan algoritma C4.5 dalam prediksi kelulusan mahasiswa Prodi Informatika dapat disimpulkan bahwa atribut yang paling dominan dalam kelulusan mahasiswa adalah IPK, kedua adalah TOEFL, ketiga adalah asal daerah, dan yang terakhir adalah jenis kelamin (Putri & Waspada, 2018). Dan penelitian terakhir adalah penelitian memprediksi metode penyakit kutil menggunakan Decision Tree dengan implementasi *Greedy Forward Selection*. Dataset yang digunakan dalam penelitian ini ada dua yaitu: *dataset cryotherapy* dan *immunotherapy*, kedua dataset ini dapat di akses pada situs *UCI Repository*. Hasil penelitian pada *dataset cryotherapy* menggunakan *Decision Tree* mendapatkan nilai akurasi 90%, sedangkan hasil seleksi fitur menggunakan *Greedy Forward Selection* sebesar 92,22%. Hasil penelitian pada *dataset immunotherapy* menggunakan *Decision Tree* tanpa seleksi fitur mendapatkan nilai akurasi sebesar 78,89%, sedangkan menggunakan seleksi fitur *Greedy Forward Selection* mendapat nilai akurasi sebesar 82,22%. Hasil ini menunjukkan bahwa bahwa seleksi fitur *Greedy Forward Selection* dapat meningkatkan kinerja dari model (Fitriyani & Arifin, 2020).

Berdasarkan latar belakang tersebut, maka penulis melakukan penelitian dengan tujuan untuk klasifikasi kelayakan calon penerima Program Keluarga Harapan (PKH) di Kelurahan Bah Sorma dengan menggunakan Algoritma C4.5. Berdasarkan uraian dari latar belakang,

maka di usulkan sebuah penelitian dengan judul “Implementasi Algoritma C4.5 Dalam Klasifikasi Calon Penerima Program Keluarga Harapan (PKH) di Kelurahan Bah Sorma”. Hasil dari Penelitian ini diharapkan dapat menjadi bahan pertimbangan lanjutan untuk pihak Pemerintah di Kelurahan Bah Sorma dalam pengambilan keputusan untuk masyarakat yang layak atau tidak mendapatkan bantuan PKH.

## LANDASAN TEORI

### Tinjauan Pustaka

Dalam penyusunan penelitian ini, penulis mengambil teori-teori yang berkaitan dengan penelitian dari berbagai sumber yang relevan. Sehingga dengan teori yang tepat dapat menghasilkan penelitian yang baik dan bermanfaat.

### Pengertian Penerima Program Keluarga Harapan (PKH)

Program Keluarga Harapan (PKH) merupakan program bantuan sosial bersyarat yang diberikan oleh Pemerintah dalam upaya mengentaskan kemiskinan (Alfiah, 2021). Program ini menerapkan Conditional Cash Transfer (CCT) dengan fokus terhadap aspek peningkatan kesehatan, peningkatan kesejahteraan dan pendidikan. Penerima PKH harus melalui proses seleksi yang diperoleh dari analisis dan pertimbangan yang dilakukan oleh Pemerintah (Arifin et al., 2022). Program manfaat PKH dilakukan pada tahun 2007 hingga saat ini dimana sudah mencapai 34 Provinsi yang terdiri dari 512 kabupaten/kota dan 7.214 kecamatan. Dampak yang paling signifikan dari manfaat PKH adalah peningkatan sebanyak 4,8% biaya konsumsi dan penurunan beban pengeluaran pada keluarga prasejahtera. PKH dapat membantu keluarga pembiayaan sekolah anak tingkat SD, SMP dan SMA, sehingga perkembangan penyaluran PKH dilakukan selalu meningkat (Muttakin et al., 2021). Sehingga dalam jangka pendek, program PKH dapat mengurangi biaya pengeluaran keluarga miskin serta mengurangi kemiskinan yang berkepanjangan.

### Pengertian Data Mining

Data Mining merupakan proses menemukan informasi atau pola dari data yang besar. Data Mining meliputi kumpulan dari ekstraksi, analisis, dan statistic data (Arifin et al., 2022). Data Mining bertujuan untuk memanfaatkan data dalam database dengan mengolah data tersebut sehingga menghasilkan informasi baru yang berguna untuk kepentingan perusahaan atau instansi.

Data Mining sering juga disebut *Knowledge discovery in database (KDD)* merupakan meliputi kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan



dalam set data berukuran besar. “Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.” (Azwanti, 2018).

### **Pengertian Data Mining Klasifikasi**

*Data Mining* terdiri beberapa kelompok pekerjaan yaitu : prediksi (*prediction*), klustering, klasifikasi, estimasi dan asosiasi. “Klasifikasi didefinisikan metode *Supervised Learning* yang membutuhkan data training dengan label untuk menghasilkan sebuah model dalam data uji pada kelompok atau kelas yang telah ditentukan”. Beberapa teknik klasifikasi yang digunakan adalah *Decision Tree*, *Rule-Based Classifier*, *Neural-Network*, *Support Machine* Dan *Naïve Bayes Classifier* (Setio et al., 2020).

### **Pengertian Pohon Keputusan (*Decision Tree*)**

Pohon keputusan merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah serangkaian pertanyaan yang sistematis dan diatur sehingga setiap pertanyaan query atribut berdasarkan nilai dari atribut tersebut (Fitriyani & Arifin, 2020).

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon *variabel input* dengan sebuah *variabel target*. Bagian awal dari pohon keputusan ini merupakan titik akar (*root*), sedangkan setiap cabang dari pohon keputusan adalah pembagian berdasarkan hasil uji, dan titik akhir (*leaf*) adalah pembagian kelas yang dihasilkan.

### **Algoritma C4.5**

Algoritma C4.5 merupakan salah satu solusi pemecahan masalah pada teknik klasifikasi. Algoritma C4.5 merupakan algoritma yang digunakan untuk mengeksplorasi data yaitu untuk menemukan hubungan *variable input* atau *attribute* kriteria dengan *variable target* atau *decision attribute* (atribut keputusan) (Sihombing et al., 2021).

### **Pengertian RapidMiner**

Menurut (Elfaladonna & Rahmadani, 2019) ‘*RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*Open Source*) sebuah salah satu solusi untuk melakukan analisis terhadap *Data Mining*’. *RapidMiner* aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin *Data Mining* untuk integrasi kepada penelitian atau suatu produk.

1 Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model dan optimasi

Penelitian yang dilakukan penulis menggunakan tool *RapidMiner Studio versi 5.3*, tampilan *RapidMiner Studio versi 5.3* terdapat pada Gambar 1 sebagai berikut :



Gambar 1. Tampilan *RapidMiner Studio versi 5.3*

Keterangan pada Gambar 1 adalah :

1. *Menu Bar*, merupakan menu utama yang terletak dibawah Title Bar yang berfungsi sebagai menu perintah untuk mengoperasikan tool *RapidMiner Studio versi 5.3*.
2. *Toolbox*, merupakan komponen yang digunakan untuk proses penyimpanan, membuat proses baru, proses eksekusi analisis, dan untuk mencari data, operator dan membuka proses yang sudah ada.
3. *Operator View*, merupakan komponen yang digunakan untuk operator dalam *RapidMiner* yang terdiri dari *Data Access*, *Blending*, *Cleansing*, dan lainnya untuk dalam hal proses analisis.
4. *Repository View*, merupakan komponen penting dalam *Design Perspective* untuk mengolah dan menata proses serta dapat digunakan sebagai sumber data.
5. *Process View*, merupakan langkah-langkah dalam proses analisis, sebagai penghubung langkah-langkah proses dan sebagai media desain proses langkah-langkah analisis.
6. *Parameters View*, merupakan komponen dalam hal menentukan parameter dari operator yang digunakan, beberapa operator membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang baik dan benar.
7. *Help dan Comment View*, merupakan komponen yang menunjukkan penjelasan setiap operator yang digunakan.

### Kinerja Klasifikasi

Dalam pengujian kinerja klasifikasi menggunakan Algoritma C4.5 secara umum menggunakan Matriks Konfusi (21) *Confusion Matrix*. *Confusion Matrix* merupakan visualisasi

yang biasa digunakan pada Algoritma berupa matrik yang memperlihatkan kebenaran dan kesalahan prediksi data dari hasil yang diperoleh seperti Akurasi (*Accuracy*), *Precision*, dan *Recall* (Isra, 2022). Terdapat empat nilai yang dihasilkan dalam *Confusion Matrix*, diantaranya *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN). Berikut ilustrasi dari *Confusion Matrix* yang dapat dilihat pada tabel 1 :

**Tabel 1. Confusion Matrix**

	38 <i>Relevan</i>	<i>Non Relevan</i>
<i>Retrive</i>	<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
<i>Non Retrive</i>	<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

(Sumber : Abidin et al., 2023)

Untuk mengetahui nilai dari Akurasi, *Precision* dan *Recall* dapat dilihat persamaan dari nilai Akurasi, *Precision* dan *Recall* sebagai berikut :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots(2.3)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2.4)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(2.5)$$

Dengan Keterangan :

TP = *True Positive*

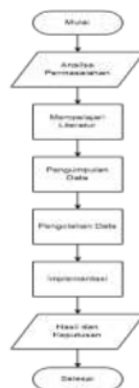
FP = *False Positive*

FN = *False Negative*

TN = *True Negative*

**METODOLOGI PENELITIAN**

Rancangan atau model penelitian ini disajikan dalam rancangan menggunakan diagram *Flowchart*. Diagram *Flowchart* digunakan untuk mengetahui alur penelitian dari awal sampai akhir dan juga memperjelas bagaimana nantinya penelitian ini berjalan yang dapat dilihat pada Gambar 2. berikut :



**Gambar 2. Rancangan Penelitian**



## HASIL DAN PEMBAHASAN

### Hasil

Hasil pada bab ini disajikan sesuai penelitian yang dilakukan. Pada penelitian ini digunakan tools *RapidMiner* dengan algoritma C4.5 sebagai pengujian data yang telah di analisa. Berikut adalah bagian-bagian dari *RapidMiner* serta proses pengolahan data yang digunakan.

### Perhitungan Menggunakan Algoritma C4.5

Data penelitian terdiri dari 5 atribut dan 1 label. Berikut ini adalah rekapitulasi data penelitian yang digunakan pada Tabel 2 dengan keterangan sebagai berikut :

V1 = Aset

V2 = Bantuan Lain

V3 = Pekerjaan

V4 = Penghasilan

V5 = Tanggungan

Label = Penerima PKH

**Tabel 2. Data Penelitian**

Nama Pengurus	V1	V2	V3	V4	V5	Label
Susilawati	Ada	Iya	Wiraswasta	Lebih dari Rp. 3.000.000	Dua	Tidak
Elly Noviani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Dini Mei Riani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Remina Malau	Tidak	Iya	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Suratminah	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Dua	Iya
Nasli Aswani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Suyati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tidak Ada	Iya
Sri Jenti	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Satu	Iya
Masriana Purba	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Dua	Iya
Ida Lesmawati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Nenny Pristari	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Dua	Iya
Fatmawati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Dua	Iya
Siswati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Wenni Ofrika	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Riama Sitohang	Ada	Tidak	Wiraswasta	Lebih dari Rp. 3.000.000	Tiga	Tidak
Sri Rahayu	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Dua	Iya
Dina Susana	Ada	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Satu	Tidak
Agustinus Sirait	Ada	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tidak Ada	Tidak
Idah Mesnawaty	Tidak	Tidak	Wiraswasta	50 awah Rp. 2.000.000	Satu	Iya
Sriwati	Tidak	Iya	Tidak Bekerja	50 h dari Rp. 3.000.000	Dua	Tidak
Saudur Purba	Ada	Iya	Wiraswasta	Lebih dari Rp. 3.000.000	Dua	Tidak
Rosdiana Silitonga	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Riyani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Legiana	Tidak	Iya	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Ismiati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Dua	Iya
Atemi	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Mariana	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tidak Ada	Iya
Pungut	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Satu	Iya
Zubaida Sinaga	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Dua	Iya
Poniaty	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Sulastri	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Dua	Iya
Nurhayati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Dua	Iya
Pintauli	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tiga	Iya
Dermawati Mapaung	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Efa Mala Sari	Ada	Tidak	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Tiga	Tidak

Nama Pengurus	V1	V2	V3	V4	V5	Label
Rubiah	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Sumiati Sinaga	Ada	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Tidak
Lestia Manik	Ada	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Tidak
Sulistiana	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya
Muji Andayani	Tidak	Iya	Tidak Bekerja	Rp. 2.000.000 - Rp. 2.999.999	Dua	Iya
Tuti Halimah	Ada	Iya	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Dua	Tidak
Eslinde Simarmata	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Susanti Sitinjak	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Paina	Tidak	Iya	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Tity Usnaini Lubis	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Kartina	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Sri Agus Tina	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Iya
Jumini	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya
Teti Hariati	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Dua	Iya
Siti Rohani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Sunaryatun	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Dua	Iya
Rahmadani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Siti Aisyah	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Berliana Rita	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Mariani Togatorop	Ada	Tidak	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Tiga	Tidak
Yusni Artaida	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Sartika Ayu	Ada	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Tidak
Marlina	Ada	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Tidak
Yusniarti	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya
Srigiat Ningsih	Tidak	Iya	Tidak Bekerja	Rp. 2.000.000 - Rp. 2.999.999	Dua	Iya
Rohani Hayati	Ada	Iya	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Dua	Tidak
Rika Irawati	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Samini	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Nurmi	Tidak	Iya	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Rakan Intan K	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Jami	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Poniti	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Iya
Dermalina Saragih	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya
Nurmala	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Dua	Iya
Sayem	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Sulastrı	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Dua	Iya
Even	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Rosmaida Hutagaol	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Juli Arlina	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Rismawaty	Ada	Tidak	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Tiga	Tidak
Rayani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Arga Simanjuntak	Ada	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Tidak
Wasinah	Ada	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Tidak
Jurita Numawati	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya
Nurbaedah Saragih	Tidak	Iya	Tidak Bekerja	Rp. 2.000.000 - Rp. 2.999.999	Dua	Iya
Dilli Nopriani	Ada	Iya	Wiraswasta	Lebih dari Rp. 3.000.000	Dua	Tidak
Eva Meliana	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Nurul Kumala Sari	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Hormaulina Henrica	Tidak	Iya	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Hetti Tampubolon	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Erlis	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Nemi	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Iya
Misrawati	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya
Susanti	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Dua	Iya
Berlihati Tarigan	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Lina	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Dua	Iya
Asia Saputri	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Siti Halida	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tiga	Iya
Rita Manik	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Satu	Iya
Nurhasana Siregar	Ada	Tidak	Wiraswasta	Lebih dari Rp. 3.000,000	Tiga	Tidak
Suri Asri Yani	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Dua	Iya
Sugiem	Ada	Tidak	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Satu	Tidak
Hotri Sauli Sinaga	Ada	Tidak	Tidak Bekerja	Dibawah Rp. 2.000,000	Tidak Ada	Tidak
Ernawati	Tidak	Tidak	Wiraswasta	Dibawah Rp. 2.000,000	Satu	Iya

Nama Pengurus	V1	V2	V3	V4	V5	Label
Suhaila Siregar	Tidak	Iya	Tidak Bekerja	Rp. 2.000.000 - Rp. 2.999.999	Dua	Iya
Laila	Ada	Tidak	Tidak Bekerja	Rp. 2.000.000 - Rp. 2.999.999	Tiga	Iya
Magdalena	Ada	Tidak	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Satu	Tidak
Bernita Simarmata	Tidak	Iya	Wiraswasta	Dibawah Rp. 2.000.000	Tiga	Iya
Henni Supriaty	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Dua	Iya
Tiorina Simatupang	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Mesiyah	Tidak	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Tidak Ada	Iya
Ngatinem	Ada	Tidak	Tidak Bekerja	Dibawah Rp. 2.000.000	Satu	Iya
Siti Atika	Ada	Tidak	Tidak Bekerja	Rp. 2.000.000 - Rp. 2.999.999	Dua	Iya
Rosenna Rita Purba	Tidak	Iya	Wiraswasta	Rp. 2.000.000 - Rp. 2.999.999	Tiga	Iya
Diana Sipayung	Ada	Tidak	Wiraswasta	Dibawah Rp. 2.000.000	Dua	Tidak

Berdasarkan Tabel 2 di atas dapat digunakan untuk mengklasifikasi penerima PKH dengan menggunakan Algoritma C4.5. Berikut ini langkah-langkah dalam membentuk pohon keputusan menggunakan Algoritma C4.5 :

1. Menghitung nilai *Entropy* dari total data. Dari data penelitian, diketahui jumlah kasus ada 110 data, diketahui Label Tidak ada 23 data dan Label Iya ada 87 data sehingga nilai *Entropy* yang diperoleh :

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-23/110 * \log_2 (23/110)) + (-87/110 * \log_2 (87/110)) \\
 &= 0,7397
 \end{aligned}$$

2. Menghitung nilai *Entropy* pada atribut Aset dengan dengan setiap kategori. Berdasarkan Tabel IV.1 diketahui kategori Ada dengan jumlah kasus 25 data dengan label Tidak sebanyak 22 data dan label Iya sebanyak 3 data. Untuk kategori Tidak diketahui jumlah kasus sebanyak 85 dengan label Tidak sebanyak 1 data dan label Iya sebanyak 84 data. Sehingga nilai *Entropy* untuk setiap kategori pada atribut Aset sebagai berikut :

$$\begin{aligned}
 Entropy(Ada) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-22/25 * \log_2 (22/25)) + (-3/25 * \log_2 (3/25)) \\
 &= 0,5294
 \end{aligned}$$

$$\begin{aligned}
 Entropy(Tidak) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-1/85 * \log_2 (1/85)) + (-84/85 * \log_2 (84/85)) \\
 &= 0,0923
 \end{aligned}$$

3. Selanjutnya menghitung nilai *Gain* pada Atribut Aset dengan jumlah data pada setiap kategori, yaitu kategori Ada dan Tidak pada atribut Aset sebagai berikut :

$$\begin{aligned}
 Gain(Aset) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\
 &= 0,7397 - (((25/110)*0,5294) + ((85/110)*0,0923)) \\
 &= 0,5481
 \end{aligned}$$

4. Menghitung nilai *Entropy* pada atribut Bantuan Lain dengan dengan setiap kategori. Berdasarkan Tabel IV.1 diketahui kategori Iya dengan jumlah kasus 17 data dengan label Tidak sebanyak 6 data dan label Iya sebanyak 11 data. Untuk kategori Tidak diketahui jumlah kasus sebanyak 93 dengan label Tidak sebanyak 17 data dan label Iya sebanyak 76 data. Sehingga nilai *Entropy* untuk setiap kategori pada atribut Bantuan Lain sebagai berikut :

$$\begin{aligned} Entropy(Iya) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-6/17 \cdot \log_2 (6/17)) + (-11/17 \cdot \log_2 (11/17)) \\ &= 0,9367 \end{aligned}$$

$$\begin{aligned} Entropy(Tidak) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-17/93 \cdot \log_2 (17/93)) + (-76/93 \cdot \log_2 (76/93)) \\ &= 0,6862 \end{aligned}$$

5. Selanjutnya menghitung nilai *Gain* pada Atribut Bantuan Lain dengan jumlah data pada setiap kategori, yaitu kategori Iya dan Tidak pada atribut Bantuan Lain sebagai berikut :

$$\begin{aligned} Gain(Bantuan Lain) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,7397 - (((17/110)*0,9367)+(93/110)*0,6862) \\ &= 0,0149 \end{aligned}$$

6. Menghitung nilai *Entropy* pada atribut Pekerjaan dengan dengan setiap kategori. Berdasarkan Tabel 2 diketahui kategori Wiraswasta dengan jumlah kasus 39 data dengan label Tidak sebanyak 17 data dan label Iya sebanyak 22 data. Untuk kategori Tidak Bekerja diketahui jumlah kasus sebanyak 71 dengan label Tidak sebanyak 6 data dan label Iya sebanyak 65 data. Sehingga nilai *Entropy* untuk setiap kategori pada atribut Pekerjaan sebagai berikut :

$$\begin{aligned} Entropy(Wiraswasta) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-17/39 \cdot \log_2 (17/39)) + (-22/39 \cdot \log_2 (22/39)) \\ &= 0,9881 \end{aligned}$$

$$\begin{aligned} Entropy(Tidak Bekerja) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-6/71 \cdot \log_2 (6/71)) + (-65/71 \cdot \log_2 (65/71)) \\ &= 0,4179 \end{aligned}$$

7. Selanjutnya menghitung nilai *Gain* pada Atribut Pekerjaan dengan jumlah data pada setiap kategori, yaitu kategori Wiraswasta dan Tidak Bekerja pada atribut Pekerjaan sebagai berikut :

$$\begin{aligned} Gain(Pekerjaan) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,7397 - (((39/110)*0,9881)+(71/110)*0,4179) \\ &= 0,1197 \end{aligned}$$

8. Menghitung nilai *Entropy* pada atribut Penghasilan dengan dengan setiap kategori. Berdasarkan Tabel 2 diketahui kategori <sup>93</sup> Lebih dari Rp.3.000.000 dengan jumlah kasus 6 data dengan label Tidak sebanyak 6 data dan label Iya sebanyak 0 data. Untuk kategori Rp.2.000.000 – Rp. 2.999.999 diketahui jumlah kasus sebanyak 14 dengan label Tidak sebanyak 7 data dan label Iya sebanyak 7 data. Untuk kategori Dibawah Rp.2.000.000 diketahui jumlah kasus sebanyak <sup>5</sup> 90 dengan label Tidak sebanyak 10 data dan label Iya sebanyak 80 data. Sehingga nilai *Entropy* untuk setiap kategori pada atribut Penghasilan sebagai berikut :

$$\begin{aligned} \text{Entropy}(\text{Lebih dari Rp.3.000.000}) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-6/6 * \log_2 (6/6)) + (-0/6 * \log_2 (0/6)) \\ &= 0,0000 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(\text{Rp.2.000.000 – Rp.2.999.999}) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-7/14 * \log_2 (7/14)) + (-7/14 * \log_2 (7/14)) \\ &= 1,0000 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(\text{Dibawah Rp.2.000.000}) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-10/90 * \log_2 (10/90)) + (-80/90 * \log_2 (80/90)) \\ &= 0,5033 \end{aligned}$$

9. Selanjutnya menghitung nilai *Gain* pada Atribut Penghasilan dengan jumlah data pada setiap kategori, yaitu kategori Lebih dari Rp.3.000.000, Rp.2.000.000 – Rp. 2.999.999 dan Dibawah Rp.2.000.000 pada atribut Penghasilan sebagai berikut :

$$\begin{aligned} \text{Gain}(\text{Penghasilan}) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \\ &= 0,7397 - (((6/110)*0,0000)+((14/110)*1,0000)+((90/110)*0,5033)) \\ &= 0,2007 \end{aligned}$$

10. Menghitung nilai *Entropy* pada atribut Tanggungan dengan dengan setiap kategori. Berdasarkan Tabel 2 diketahui kategori Tiga dengan jumlah kasus 28 data dengan label Tidak sebanyak 5 data dan label Iya sebanyak 23 data. Untuk kategori Dua diketahui jumlah kasus sebanyak 38 dengan label Tidak sebanyak 7 data dan label Iya sebanyak 31 data. Untuk kategori Satu diketahui jumlah kasus sebanyak 33 dengan label Tidak sebanyak 6 data dan label Iya sebanyak 27 data. Untuk kategori Tidak Ada diketahui jumlah kasus sebanyak <sup>5</sup> 11 dengan label Tidak sebanyak 5 data dan label Iya sebanyak 6 data. Sehingga nilai *Entropy* untuk setiap kategori pada atribut Tanggungan sebagai berikut :

$$\begin{aligned} \text{Entropy}(\text{Tiga}) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-5/28 * \log_2 (5/28)) + (-23/28 * \log_2 (23/28)) \\ &= 0,6769 \end{aligned}$$



$$\begin{aligned}
 Entropy(Dua) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-7/38 \cdot \log_2 (7/38)) + (-31/38 \cdot \log_2 (31/38)) \\
 &= 0,6892
 \end{aligned}$$

$$\begin{aligned}
 Entropy(Satu) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-6/33 \cdot \log_2 (6/33)) + (-27/33 \cdot \log_2 (27/33)) \\
 &= 0,6840
 \end{aligned}$$

$$\begin{aligned}
 Entropy(Tidak Ada) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= (-5/11 \cdot \log_2 (5/11)) + (-6/11 \cdot \log_2 (6/11)) \\
 &= 0,9940
 \end{aligned}$$

11. Selanjutnya menghitung nilai *Gain* pada Atribut Tanggungan dengan jumlah data pada setiap kategori, yaitu kategori Tiga, Dua, Satu dan Tidak Ada pada atribut Tanggungan sebagai berikut :

$$\begin{aligned}
 Gain(Tanggungan) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\
 &= 0,7397 - (((28/110)*0,6769)+((38/110)*0,6892)+((33/110)*0,6840) + \\
 &\quad ((11/110)*0,9940)) \\
 &= 0,0247
 \end{aligned}$$

Hasil dari perhitungan nilai *Entropy* dan nilai *Gain*, untuk tiap atribut dapat dilihat pada

Tabel 3 berikut :

**Tabel 3. Gain Node 1**

No	Atribut	Jumlah Kasus	Tidak	Iya	Entropy	Gain
	Total	110	23	87	0,7397	
1	Aset					0,5481
	Ada	25	22	3	0,5294	
	Tidak	85	1	84	0,0923	
2	Bantuan Lain					0,0149
	Iya	17	6	11	0,9367	
	Tidak	93	17	76	0,6862	
3	Pekerjaan					0,1197
	Wiraswasta	39	17	22	0,9881	
	Tidak Bekerja	71	6	65	0,4179	
4	Penghasilan					0,2007
	Lebih dari Rp. 3.000.000	6	6	0	0,0000	
	Rp. 2.000.000 - Rp. 2.999.999	14	7	7	1,0000	
	Dibawah Rp. 2.000.000	90	10	80	0,5033	
5	Tanggungan					0,0247
	Tiga	28	5	23	0,6769	
	Dua	38	7	31	0,6892	
	Satu	33	6	27	0,6840	
	Tidak Ada	11	5	6	0,9940	

Dari Tabel 3 diketahui nilai *Gain* terbesar adalah atribut Aset dengan nilai *Gain* 0,5481 dengan kategori Ada dengan Nilai *Entropy* 0,5294 dan kategori Tidak dengan Nilai *Entropy* 0,0923 kemudian kedua kategori tersebut dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Berdasarkan penjelasan pada Tabel IV.2 dapat diperoleh pohon keputusan pada Gambar 3 sebagai berikut :



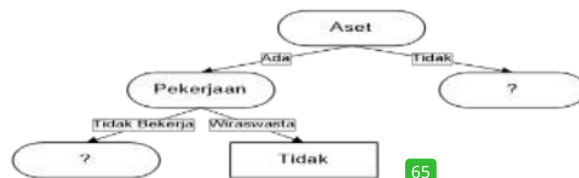
Gambar 3. Pohon Keputusan Node 1

Selanjutnya perhitungan nilai *Entropy* dan nilai *Gain* untuk Node 1.1 lanjutan dari atribut Aset dengan kategori Ada dengan Nilai *Entropy* 0,5294 yang dapat dilihat rekapitulasi *Gain* Node 1.1. pada Tabel 4 sebagai berikut :

Tabel 4. Gain Node 1.1

No	Atribut	Jumlah Kasus	Tidak	Iya	Entropy	Gain
	Aset = Ada	25	22	3	0,5294	
2	Bantuan Lain					0,0415
	Iya	5	5	0	0,0000	
	Tidak	20	17	3	0,6098	
3	Pekerjaan					0,2239
	Wiraswasta	17	17	0	0,0000	
	Tidak Bekerja	8	5	3	0,9544	
4	Penghasilan					0,0609
	Lebih dari Rp. 3.000.000	5	5	0	0,0000	
	Rp. 2.000.000 - Rp. 2.999.999	9	7	2	0,7642	
	Dibawah Rp. 2.000.000	11	10	1	0,4395	
5	Tanggungans					0,1980
	Tiga	6	5	1	0,0000	
	Dua	7	6	1	0,5917	
	Satu	7	6	1	0,5917	
	Tidak Ada	5	5	0	0,0000	

Dari Tabel 4 diketahui nilai *Gain* terbesar adalah atribut Pekerjaan dengan nilai *Gain* 0,2239 dengan kategori Wiraswasta dengan Nilai *Entropy* 0,0000 dan kategori Tidak Bekerja dengan Nilai *Entropy* 0,9544. Kemudian kategori Tidak Bekerja dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya untuk Node 1.1.1. Berdasarkan penjelasan pada Tabel 4 dapat diperoleh pohon keputusan pada Gambar 4 sebagai berikut :



Gambar 4. Pohon Keputusan Node 1.1

Selanjutnya perhitungan nilai *Entropy* dan nilai *Gain* untuk Node 1.1.1 lanjutan dari atribut Pekerjaan dengan kategori Tidak Bekerja dengan Nilai *Entropy* 0,9544 yang dapat dilihat rekapitulasi *Gain* Node 1.1.1. pada Tabel 5 sebagai berikut :

Tabel 5. Gain Node 1.1.1

No	Atribut	Jumlah Kasus	Tidak	Iya	Entropy	Gain
	Aset = Ada & Pekerjaan = Tidak Bekerja	8	5	3	0,9544	
2	Bantuan Lain					0,0000
	Iya	0	0	0	0,0000	
	Tidak	8	5	3	0,9544	
4	Penghasilan					0,4669

		79	0	0	0	0,0000	
		Lebih dari Rp. 3.000.000	2	0	2	0,0000	
		Rp. 2.000.000 - Rp. 2.999.999	6	5	1	0,6500	
		Dibawah Rp. 2.000.000					
5	Tanggungjan						0,9544
		Tiga	1	0	1	0,0000	
		Dua	1	0	1	0,0000	
		Satu	1	0	1	0,0000	
		Tidak Ada	5	5	0	0,0000	

Dari Tabel 5 diketahui nilai *Gain* terbesar adalah atribut Tanggungan dengan nilai *Gain* 0,8113 dengan kategori Tiga dengan Nilai *Entropy* 0,0000, kategori Dua dengan Nilai *Entropy* 0,0000, kategori Satu dengan Nilai *Entropy* 0,0000, dan kategori Tidak Ada dengan Nilai *Entropy* 0,0000. Berdasarkan nilai *Entropy* keseluruhan kategori mendapat nilai 0,0000, maka Node 1.1.1 berhenti. Pada Tabel 5 dapat diperoleh pohon keputusan pada Gambar 5 berikut :



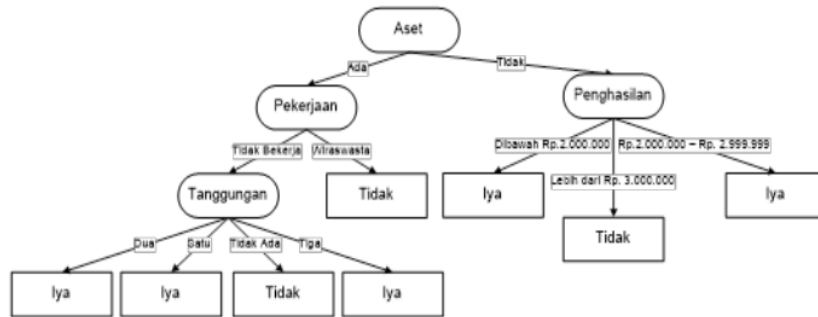
Gambar 5. Pohon Keputusan Node 1.1.1

Selanjutnya perhitungan nilai *Entropy* dan nilai *Gain* untuk Node 1.2 lanjutan dari atribut Aset dengan kategori Tidak dengan Nilai *Entropy* 0,0923 yang dapat dilihat rekapitulasi *Gain* Node 1.2. pada Tabel 6 sebagai berikut :

Tabel 6 Gain Node 1.2

No	Atribut	Jumlah Kasus	Tidak	Iya	Entropy	Gain
	Aset = Tidak	85	1	84	0,0923	
2	Bantuan Lain					0,0339
	Iya	12	1	11	0,4138	
	Tidak	73	0	73	0,0000	
3	Pekerjaan					0,0051
	Wiraswasta	22	0	22	0,0000	
	Tidak Bekerja	63	1	62	0,1176	
4	Penghasilan					0,0923
	Lebih dari Rp. 3.000.000	1	1	0	0,0000	
	Rp. 2.000.000 - Rp. 2.999.999	5	0	5	0,0000	
	Dibawah Rp. 2.000.000	79	0	79	0,0000	
5	Tanggungjan					0,0173
	Tiga	22	0	22	0,0000	
	Dua	31	1	30	0,2056	
	Satu	26	0	26	0,0000	
	Tidak Ada	6	0	6	0,0000	

Dari Tabel 6 diketahui nilai *Gain* terbesar adalah atribut Penghasilan dengan nilai *Gain* 0,00923 dengan kategori Lebih dari Rp. 3.000.000 dengan Nilai *Entropy* 0,0000, kategori Rp. 2.000.000 - Rp. 2.999.999 dengan Nilai *Entropy* 0,0000, dan kategori Dibawah Rp. 2.000.000 dengan Nilai *Entropy* 0,0000. Berdasarkan nilai *Entropy* keseluruhan kategori mendapat nilai 0,0000, maka Node 1.2 berhenti. Pada Tabel 6 dapat diperoleh pohon keputusan pada Gambar 6 berikut :



Gambar 6. Pohon Keputusan Node 1.2

Berdasarkan hasil akhir pada Gambar 6 pohon keputusan, dapat diperoleh model aturan (*rule*) berdasarkan pohon keputusan tersebut sebagai berikut :

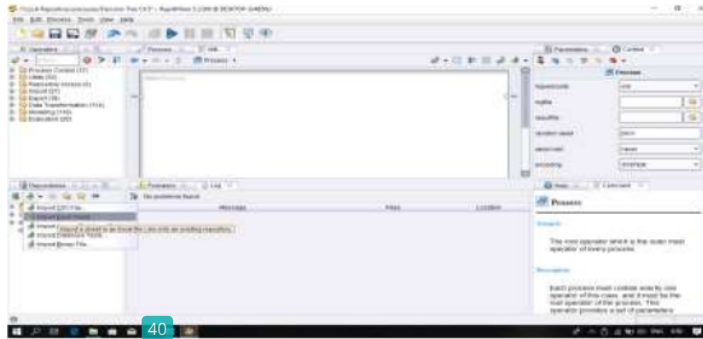
1. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Dua, maka hasilnya = Iya (Tidak = 0 dan Iya = 1).
2. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Satu, maka hasilnya = Iya (Tidak = 0 dan Iya = 1).
3. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Tidak Ada, maka hasilnya = Tidak (Tidak = 5 dan Iya = 0).
4. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Tiga, maka hasilnya = Iya (Tidak = 0 dan Iya = 1).
5. Jika Aset = Ada dan Pekerjaan = Wiraswasta, maka hasilnya = Tidak (Tidak = 17 dan Iya = 0).
6. Jika Aset = Tidak dan Penghasilan = Dibawah Rp. 2.000.000, maka hasilnya = Iya (Tidak = 0 dan Iya = 79).
7. Jika Aset = Tidak dan Penghasilan = Lebih dari Rp. 3.000.000, maka hasilnya = Tidak (Tidak = 1 dan Iya = 0).
8. Jika Aset = Tidak dan Penghasilan = Rp. 2.000.000 – Rp. 2.999.999, maka hasilnya = Iya (Tidak = 0 dan Iya = 5).

### Implementasi Algoritma C4.5

Pada bagian ini berisikan tampilan sistem dengan menggunakan Algoritma C4.5 dengan menggunakan *tools RapidMiner* sebagai berikut :

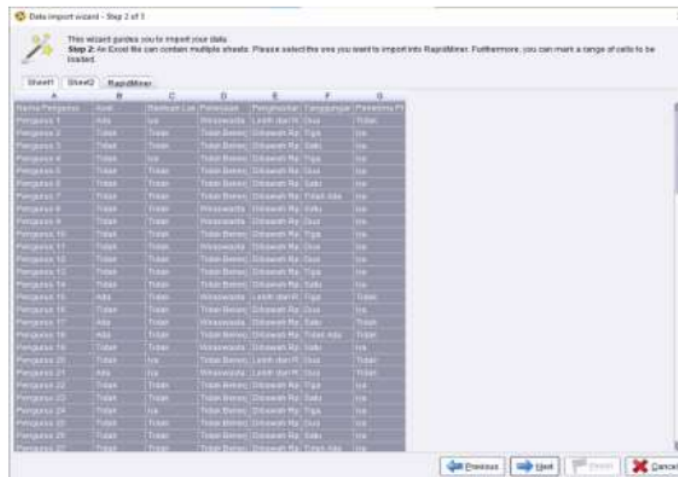
#### 1. Desain Model Algoritma C4.5

Desain Model Algoritma C4.5 menggunakan *tools RapidMiner* yang digunakan sebagai tampilan instalasi Algoritma C4.5. Berikut tahapan-tahapan gambar desain Model Algoritma C4.5 pada gambar 7 berikut :



**Gambar 7.** Alur Proses *Import Data Step 1*

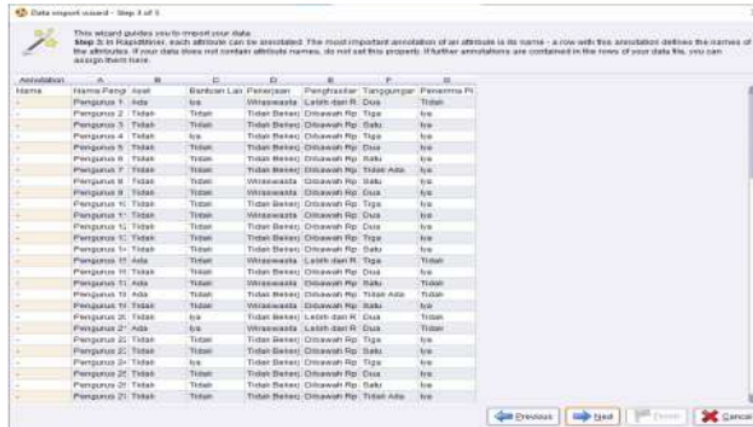
Gambar 7 dapat dilihat adalah tampilan *software RapidMiner* yang digunakan. Untuk mengimpor data yang digunakan dalam penelitian, digunakan menu pada tool *Import data into an existing repository* di menu *Repositories*. Data yang akan di *import* dalam bentuk *format Microsoft Excel*, pilih tool *Import Excel Sheet*. Fungsi tool ini untuk mengimpor data *format Microsoft Excel* dalam sebuah lembar (*sheet*) yang digunakan untuk penyeleksian dan pengolahan data. Pilih data yang akan digunakan untuk di *import* kemudian klik *Next* untuk mengelola data yang akan di *import*, seperti pada Gambar 8 berikut :



**Gambar 8.** Alur Proses *Import Data Step 2*

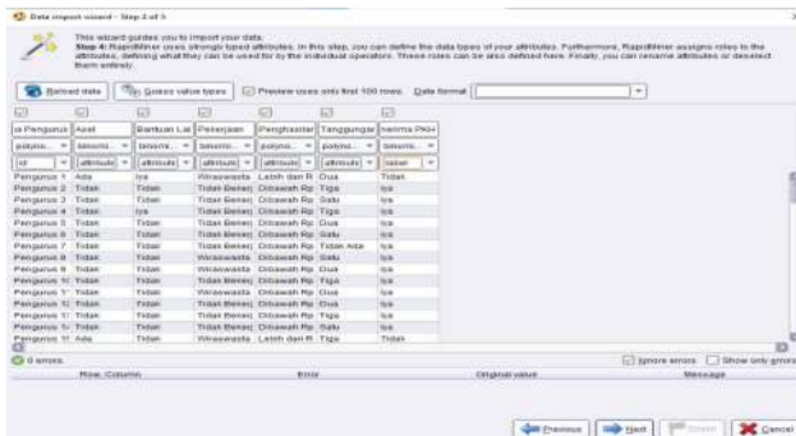
Gambar 8 dapat dilihat sebuah data yang dipilih dalam *format Microsoft Excel*. Data dalam *Microsoft Excel* dapat berisi banyak lembar (*sheet*), pilih lembar yang akan di *import* ke *RapidMiner* untuk pengolahan data. Sebuah lembar (*sheet*) yang di *import* dapat ditandai rentang atau di blok sel yang akan digunakan untuk pengolahan data. Sel yang diblok dapat berpengaruh untuk ke tahap penyeleksian data untuk menentukan id, atribut dan label yang akan digunakan. Tahapan untuk menentukan atribut dapat dilihat pada Gambar 9 berikut:





Gambar 9. Alur Proses Import Data Step 3

Gambar 9 dapat dijelaskan bahwa setiap atribut dapat dijelaskan dan di anotasi. Atribut yang di anotasi yang paling penting adalah namanya baris dengan anotasi ini untuk menentukan nama atribut pada tahapan selanjutnya. Gambar IV.7 dapat dilihat atribut yang digunakan terdiri dari Nama Pengurus, Aset, Bantuan Lain, Pekerjaan, Penghasilan, Tanggungan dan Penerima PKH. Atribut yang digunakan ditentukan nama atribut untuk pengolahan data. Tahapan selanjutnya dapat dilihat pada Gambar 10 berikut:



Gambar 10. Alur Proses Import Data Step 4

Gambar 10. dapat dilihat pemilihan bentuk dan jenis data sangat diperlukan. Penelitian ini menggunakan Nama Pengurus sebagai jenis *id* dengan tipe *Polynomial* sebagai penanda dalam setiap data, atribut Aset, Bantuan Lain, Pekerjaan, Penghasilan, dan Tanggungan menggunakan jenis *Attribute* dan Penerima PKH menggunakan jenis *Label* dengan tipe *Binominal* yang merupakan nilai *String* yang hanya memiliki 2 kategori yaitu Iya dan Tidak. Setelah ditentukan tipe atribut yang dipilih maka klik *Finish*, kemudian tentukan tempat untuk menyimpan data yang di *importing* pada *local repository*.

No	Nama Pengurus	Penerima PKH	...	...	...	...	...
1	Pengurus 1	Tidak	Ya	Wawancara	Lainnya	Ya	Ya
2	Pengurus 2	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
3	Pengurus 3	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
4	Pengurus 4	Ya	Tidak	Ya	Tidak	Tidak	Tidak
5	Pengurus 5	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
6	Pengurus 6	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
7	Pengurus 7	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
8	Pengurus 8	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
9	Pengurus 9	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
10	Pengurus 10	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
11	Pengurus 11	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
12	Pengurus 12	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
13	Pengurus 13	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
14	Pengurus 14	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
15	Pengurus 15	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
16	Pengurus 16	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
17	Pengurus 17	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
18	Pengurus 18	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
19	Pengurus 19	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
20	Pengurus 20	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
21	Pengurus 21	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
22	Pengurus 22	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
23	Pengurus 23	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
24	Pengurus 24	Ya	Tidak	Tidak	Tidak	Tidak	Tidak

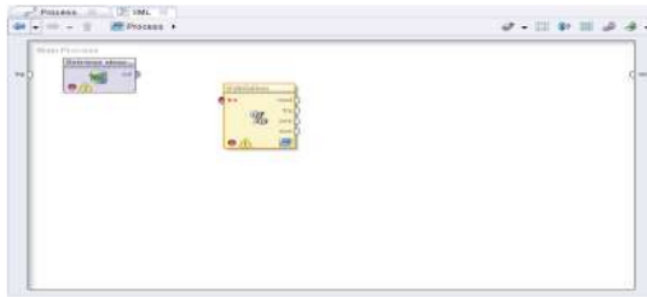
Gambar 11. Hasil *Import Data*

Gambar 11 dapat dilihat hasil dari *import* data menghasilkan informasi terdapat 110 *example* data yang merupakan jumlah sampel sebanyak 110 item, 2 *special attributes* yang merupakan 2 atribut *special* pada data yaitu atribut Nama Pengurus yang menggunakan jenis *id* dengan tipe *Polynomial* dan Penerima PKH yang menggunakan jenis *label* dengan tipe *Binominal*, dan 5 *regular attributes* yang merupakan 5 atribut regular yang digunakan sebagai *node* akar pada pohon keputusan dengan jenis *attribute*. Langkah selanjutnya kembali ke tahapan desain pada *main process* untuk memproses data yang akan digunakan untuk mencari hasil keputusan. Langkah pertama *drag and drop* data yang telah di *importing* kedalam *main process* di menu *Repositories* pada *RapidMiner* yang dapat dilihat pada Gambar 12 berikut :



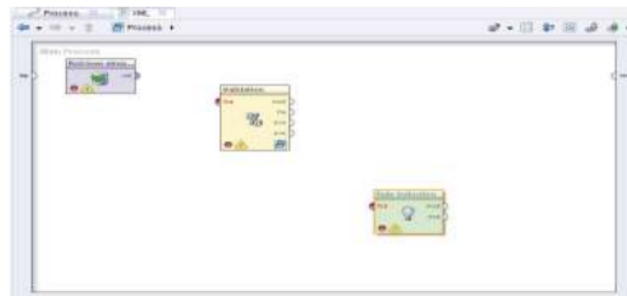
Gambar 12. *Drag and Drop Operator Retrieve Data*

Gambar 12 dapat dilihat hasil *import* yang telah disimpan di *Repositories* menjadi data yang siap diolah. Untuk Operator *Retrieve Data* yang dimasukkan ke dalam *main process* pastikan menu *repository entry* pada tab *Parameter View* di sebelah kanan sudah sesuai dengan nama repositori yang telah di *import*. Langkah selanjutnya *drag and drop* operator *X-Validation* yang diambil dari tab *Operators*. Operator *X-Validation* yang digunakan untuk membagi dua area yaitu *training* dan *testing*. Area *Training* digunakan untuk membuat *rule* dengan menggunakan *Set Role* dan menggunakan *Decision Tree* untuk mendapatkan *rule*-nya. Area *testing* menggunakan *Apply Model* untuk hasil dari algoritma *C4.5* dan *Performance* untuk mendapatkan akurasi data yang digunakan. Hasil *drag and drop* operator *X-Validation* dapat dilihat pada Gambar 13 berikut :



**Gambar 13. Drag and Drop Operator X-Validation**

Gambar 13 dapat dilihat hasil *drag and drop* operator *X-Validation* pada *main process RapidMiner*. Selanjutnya *drag and drop* operator *Rule Induction* yang digunakan untuk menyederhanakan model pada Gambar 14 berikut :



**Gambar 14. Drag and Drop Operator Rule Induction**

Gambar 14 dapat dilihat hasil *drag and drop* operator *Rule Induction* pada *main process RapidMiner*. Selanjutnya didalam operator *X-Validation* dilakukan *drag and drop* proses selanjutnya untuk menambahkan operator *Decision Tree* yang merupakan operator dari Algoritma C4.5 dengan pengaturan parameter sebagai berikut :

- 1) *Criterion = information gain*
- 2) *Minimal size for split = 2*
- 3) *Minimal leaf size = 1*
- 4) *Minimal Gain = 0,001*
- 5) *Maximal Depth = 20*
- 6) *Confidence = 0,25*

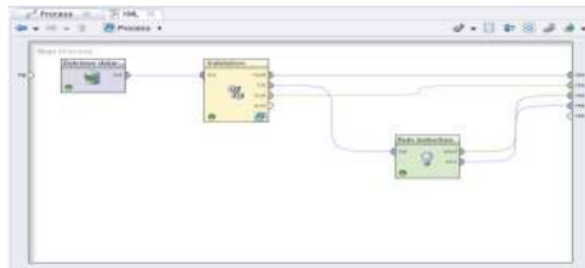
Operator *Apply Model* yang digunakan untuk menerapkan model yang sudah dipelajari atau dilatih pada dataset yang digunakan. Operator *Performance* digunakan untuk mengevaluasi kinerja statistik dari klasifikasi yang dihasilkan berupa nilai akurasi.

Operator-operator tersebut yang telah dimasukkan ke *main process* di dalam *X-Validation* dapat dilihat pada Gambar 15 berikut :

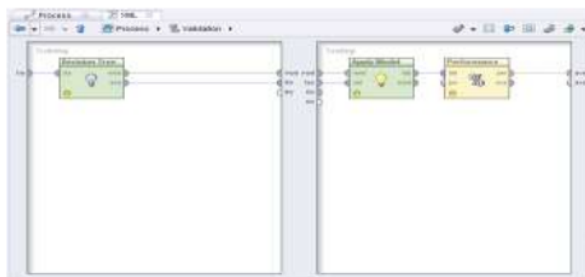


Gambar 15. Drag and Drop Operator didalam X-Validation

Selanjutnya menghubungkan antar operator yang telah dimasukkan kedalam *main process* pada *RapidMiner*. Hasil hubungan antar operator dapat dilihat pada Gambar 15 dan Gambar 16 sebagai bentuk Rancangan Model dengan menggunakan Algoritma C4.5:



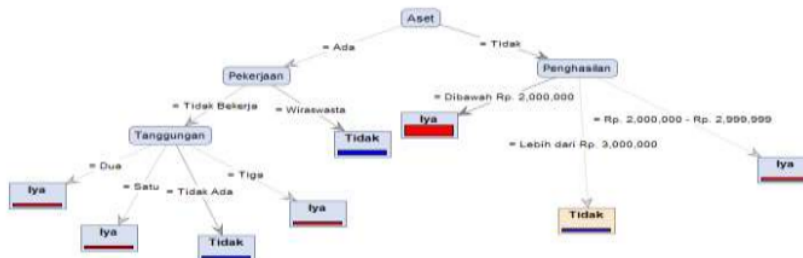
Gambar 16. Rancangan Model Algoritma C4.5 di RapidMiner



Gambar 17. Rancangan Model di dalam Operator X-Validation

2. Hasil Algoritma C4.5

Setelah membuat model Algoritma C4.5 di *tools RapidMiner*, diperoleh hasil klasifikasi calon penerima Program Keluarga Harapan (PKH) di Kelurahan Bah Sorma yang dapat dilihat pada Gambar IV.16 berikut :



Gambar 18. Hasil Algoritma C4.5

Gambar 18 merupakan gambar pohon keputusan yang diperoleh dari proses di *RapidMiner* dengan menggunakan Algoritma C4.5. Hasil aturan atau *rule* yang dapat dilihat seperti Gambar 19 berikut :

```
Tree
Aset = Ada
| Pekerjaan = Tidak Bekerja
| | Tanggungan = Dua: Iya (Tidak=0, Iya=1)
| | Tanggungan = Satu: Iya (Tidak=0, Iya=1)
| | Tanggungan = Tidak Ada: Tidak (Tidak=5, Iya=0)
| | Tanggungan = Tiga: Iya (Tidak=0, Iya=1)
| Pekerjaan = Wiraswasta: Tidak (Tidak=17, Iya=0)
Aset = Tidak
| Penghasilan = Dibawah Rp. 2,000,000: Iya (Tidak=0, Iya=79)
| Penghasilan = Lebih dari Rp. 3,000,000: Tidak (Tidak=1, Iya=0)
| Penghasilan = Rp. 2,000,000 - Rp. 2,999,999: Iya (Tidak=0, Iya=5)
```

**Gambar 19. Hasil Aturan atau Rule**

Jika dilihat berdasarkan *text view* hasil pohon keputusan pada Gambar IV.17, hasil model aturan berupa teks berdasarkan Gambar IV.17 dengan penjelasan seperti berikut :

- a. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Dua, maka hasilnya = Iya (Tidak = 0 dan Iya = 1).
- b. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Satu, maka hasilnya = Iya (Tidak = 0 dan Iya = 1).
- c. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Tidak Ada, maka hasilnya = Tidak (Tidak = 5 dan Iya = 0).
- d. Jika Aset = Ada, Pekerjaan = Tidak Bekerja dan Tanggungan = Tiga, maka hasilnya = Iya (Tidak = 0 dan Iya = 1).
- e. Jika Aset = Ada dan Pekerjaan = Wiraswasta, maka hasilnya = Tidak (Tidak = 17 dan Iya = 0).
- f. Jika Aset = Tidak dan Penghasilan = Dibawah Rp. 2.000.000, maka hasilnya = Iya (Tidak = 0 dan Iya = 79).
- g. Jika Aset = Tidak dan Penghasilan = Lebih dari Rp. 3.000.000, maka hasilnya = Tidak (Tidak = 1 dan Iya = 0).
- h. Jika Aset = Tidak dan Penghasilan = Rp. 2.000.000 – Rp. 2.999.999, maka hasilnya = Iya (Tidak = 0 dan Iya = 5).

### 3. Hasil Rule Model

Berikut ini tampilan hasil *rule model* yang dihasilkan oleh operator *Rule Induction* pada *tools RapidMiner* pada gambar 20 sebagai berikut :



**RuleModel**

```

if Aset = Tidak then Iya (1 / 84)
if Pekerjaan = Wiraswasta then Tidak (17 / 0)
if Tanggungan = Tidak Ada then Tidak (5 / 0)
else Iya (0 / 2)

correct: 108 out of 109 training examples.

```

**Gambar 20. Hasil Rule Model**

Berdasarkan Gambar 20 dapat dijelaskan bahwa model terbaik dari hasil *tools RapidMiner* adalah :

- a. Jika Aset = Tidak, maka hasilnya Iya, dengan Tidak = 1 dan Iya = 84.
  - b. Jika Pekerjaan = Wiraswasta, maka hasilnya Tidak, dengan Tidak = 17 dan Iya = 0
  - c. Jika Tanggungan = Tidak Ada, maka hasilnya Tidak, dengan Tidak = 5 dan Iya = 0
  - d. Maka Iya dengan Tidak = 0 dan Iya = 2.
4. Akurasi

Setelah diperoleh hasil algoritma C4.5 pada *RapidMiner*, berikut ini hasil akurasi dari *tools RapidMiner* pada Gambar 21 berikut :

	true Tidak	true Iya	class precision
pred: Tidak	22	2	91,67%
pred: Iya	1	85	98,84%
class recall	95,85%	87,70%	

accuracy: 97,27% (+/- 4,17% (mkro: 97,27%))

**Gambar 21. Hasil Akurasi di RapidMiner**

Gambar 21 dapat dilihat perolehan akurasi dari Algoritma C4.5 pada *tools RapidMiner* sebesar 97,27%. Dapat dilihat Prediksi Tidak dengan label Tidak berjumlah 22 dan label Iya berjumlah 2 sehingga nilai *class precision* sebesar 91,67%. Prediksi Iya dengan label Tidak berjumlah 1 dan label Iya berjumlah 85 sehingga *class precision* sebesar 98,84%.

**Pembahasan**

Pembahasan pada bab ini menyajikan hal-hal yang membahas hasil validasi data dan solusi pemecahan dari permasalahan dai penelitian ini. Berikut adalah bagian-bagian dari pembahasan pada penelitian ini :

**Validasi Data**

Hasil yang dilakukan peneliti dalam perhitungan Algoritma C4.5 menghasilkan sebanyak 8 model aturan atau *rule* klasifikasi calon penerima PKH di Kelurahan Bah Sorma. Hasil yang diberikan oleh *RapidMiner* menghasilkan sebanyak 8 model aturan atau *rule* klasifikasi calon penerima PKH di Kelurahan Bah Sorma. Model aturan yang dihasilkan *RapidMiner* dapat dilihat dari keterangan Gambar 21 menggambarkan hasil dari proses yang

dilakukan peneliti dan *RapidMiner* adalah sesuai dan sama dari proses yang dilakukan. Kesamaan Model aturan yang dilakukan dapat dijadikan salah satu solusi pemecahan dari permasalahan yang diteliti. Berdasarkan pernyataan hasil antara perhitungan yang dilakukan penulis dengan hasil yang diperoleh *tools RapidMiner* memiliki kesamaan hasil dari klasifikasi calon penerima PKH di Kelurahan Bah Sorma dengan menggunakan Algoritma C4.5.

### Solusi Pemecahan Permasalahan

Hasil yang didapatkan oleh penulis dan *tools RapidMiner* adalah Algoritma C4.5 menghasilkan sebanyak 8 model aturan atau *rule* klasifikasi calon penerima PKH di Kelurahan Bah Sorma. Pada *tools RapidMiner* diperoleh hasil akurasi sebesar 97,27% dengan Prediksi Tidak dengan label Tidak berjumlah 22 dan label Iya berjumlah 2 sehingga nilai *class precision* sebesar 91,67% Prediksi Iya dengan label Tidak berjumlah 1 dan label Iya berjumlah 85 sehingga *class precision* sebesar 98,84%. Adapun model aturan yang telah diinduksi *tools RapidMiner* adalah Jika Aset = Tidak, maka hasilnya Iya, dengan Tidak = 1 dan Iya = 84, Jika Pekerjaan = Wiraswasta, maka hasilnya Tidak, dengan Tidak = 17 dan Iya = 0, Jika Tanggungan = Tidak Ada, maka hasilnya Tidak, dengan Tidak = 5 dan Iya = 0, dan Maka Iya dengan Tidak = 0 dan Iya = 2. Hasil yang diperoleh dari Data Mining Algoritma C4.5 dapat menjadi rujukan untuk Pemerintah di Kelurahan Bah Sorma dalam menyeleksi calon penerima untuk mendapatkan bantuan Program Keluarga Harapan (PKH).

## PENUTUP

### Kesimpulan

Hasil dari penelitian yang dilakukan oleh penulis menyimpulkan bahwa Data Mining dengan Algoritma C4.5 dapat mengklasifikasi calon penerima PKH di Kelurahan Bah Sorma. Penelitian ini diperoleh hasil bahwa masyarakat yang tidak mendapatkan PKH yaitu dengan ketentuan salah satu kriteria yaitu, Aset = Tidak, Pekerjaan = Wiraswasta, dan Tanggungan = Tidak Ada. Sedangkan jika tidak ada selain yang disebutkan, maka berhak mendapatkan PKH. Hasil akurasi yang diperoleh penelitian ini sebesar 97,27%. Berdasarkan hasil yang didapatkan, penelitian dengan kasus klasifikasi calon penerima PKH di Kelurahan Bah Sorma dengan menggunakan Algoritma C4.5 mendapatkan akurasi yang lebih baik daripada penelitian sebelumnya dengan menggunakan *Naïve Bayes* memperoleh akurasi sebesar 80%. Sehingga dari perbandingan kedua hasil akurasi tersebut, menyatakan teknik Klasifikasi dengan menggunakan Algoritma C4.5 dapat lebih baik mengklasifikasi calon penerima PKH di Kelurahan Bah Sorma. Model dari hasil penelitian ini menjadi rujukan kepada Pemerintah Kelurahan Bah Sorma untuk mengklasifikasi calon penerima PKH dengan baik.

## Saran

Berdasarkan penelitian dan implementasi sistem yang dilakukan, maka diberikan beberapa saran sebagai berikut.

1. Perlu adanya evaluasi kedepan untuk menambah jumlah kriteria yang digunakan agar dapat mendukung dalam proses klasifikasi *Data Mining*.
2. Ada baiknya sistem dikembangkan berbasis *android* atau Sistem Informasi sehingga dapat lebih banyak menyediakan informasi dan lebih mudah diakses semua kalangan masyarakat, terkhusus Pemerintah Kelurahan Bah Sorma.
3. Perlu adanya pengembangan lebih lanjut terhadap sistem yaitu membangun sistem yang lebih aman dan *user-friendly* dengan memperhatikan aspek – aspek Interaksi Manusia dan Komputer.
4. Perlu adanya dilakukan perbandingan dengan algoritma Data Mining lainnya yang mendukung dalam proses klasifikasi.

## DAFTAR PUSTAKA

- Alfiah, N. (2021). Klasifikasi penerima bantuan sosial Program Keluarga Harapan menggunakan metode Naive Bayes. *Jurnal Teknologi Informasi*, 16(1), 32. <https://doi.org/10.35842/jtir.v16i1.386>
- Arifin, A. A. A., Handoko, W., & Efendi, Z. (2022). Implementasi metode Naive Bayes untuk klasifikasi penerima Program Keluarga Harapan. *J-Com (Journal of Computer)*, 2(1), 21–26. <https://doi.org/10.33330/j-com.v2i1.1577>
- Azwanti, N. (2018). Analisa algoritma C4.5 untuk memprediksi penjualan motor pada PT. Capella Dinamik Nusantara cabang Muka Kuning. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 13(1), 33. <https://doi.org/10.30872/jim.v13i1.629>
- Elfaladonna, F., & Rahmadani, A. (2019). Analisa metode classification-decision tree dan algoritma C4.5 untuk memprediksi penyakit diabetes dengan menggunakan aplikasi Rapid Miner. *SINTECH (Science and Information Technology) Journal*, 2(1), 10–17. <https://doi.org/10.31598/sintechjournal.v2i1.293>
- Fitriyani, F., & Arifin, T. (2020). Implementasi Greedy Forward Selection untuk prediksi metode penyakit kutil menggunakan decision tree. *JST (Jurnal Sains Dan Teknologi)*, 9(1), 76–85. <https://doi.org/10.23887/jstundiksha.v9i1.24896>
- Muttakin, F., Fatwa, K. N., & Sarbaini, S. (2021). Implementasi additive ratio assessment model untuk rekomendasi penerima manfaat Program Keluarga Harapan. *Jurnal Sains, Teknologi dan Industri*, 19(1), 40–48.
- Nella, N. I., Setiawan, N. Y., & Ratnawati, D. E. (2022). Klasifikasi penerima bantuan Program Keluarga Harapan menggunakan algoritme decision tree C4.5 (Studi kasus: Desa Mlirip Kabupaten Mojokerto). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(3), 1332–1339.

- R.H. Zer, P. P. P. A. N. F. I., Wahyuni, M., Rangga, A., & Situmorang, Z. (2022). Analisis model kepuasan mahasiswa terhadap cara pengajaran dosen menggunakan algoritma C4.5. *JIKO (Jurnal Informatika dan Komputer)*, 6(1), 58–64.
- Setio, P. B. N., Saputro, D. R. S., & Winarno, B. (2020). Klasifikasi dengan pohon keputusan berbasis algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Sihombing, I. A., Hartama, D., Parlina, I., Gunawan, I., & Kirana, I. O. (2021). Analisis keberhasilan pembelajaran daring pada masa pandemi Covid-19 menggunakan algoritma C4.5 dan Naive Bayes. *JUKI*, 3(November), 89–96.
- Sulaiman, S. (2022). Penerapan algoritma Naive Bayes untuk mengklasifikasi calon penerima Program Keluarga Harapan (PKH) di Kelurahan.
- Tanjung, M. A., P, P., & Qurniawan, H. (2021). Analisa kelayakan penerima Program Keluarga Harapan (PKH) menggunakan algoritma C4.5. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 6(1), 217. <https://doi.org/10.30645/jurasik.v6i1.286>

# Implementasi Algoritma C4.5 dalam Klasifikasi Calon Penerima Program Keluarga Harapan (PKH) di Kelurahan Bah Sorma

## ORIGINALITY REPORT

17%

SIMILARITY INDEX

12%

INTERNET SOURCES

11%

PUBLICATIONS

5%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://digilib.unila.ac.id">digilib.unila.ac.id</a> Internet Source	<1%
2	<a href="http://www.pelita-informatika.com">www.pelita-informatika.com</a> Internet Source	<1%
3	<a href="http://www.ejournal.pelitaindonesia.ac.id">www.ejournal.pelitaindonesia.ac.id</a> Internet Source	<1%
4	Jefry Antonius Karlia, Wawan Nurmansyah. "Application of C4.5 Algorithm for Late Payment Classification of Insurance Premiums", Tekinfo: Jurnal Ilmiah Teknik Industri dan Informasi, 2021 Publication	<1%
5	Suherman, Marlia Purnamasari, Fitriani Dwi Hastuti. "KLASIFIKASI SISWA BERDASARKAN MATA PELAJARAN LINTAS MINAT MENGGUNAKAN METODE DECISION TREE C4.5", JSiI (Jurnal Sistem Informasi), 2021 Publication	<1%

6	<a href="http://a-research.upi.edu">a-research.upi.edu</a> Internet Source	<1 %
7	<a href="http://ejournal.kemenperin.go.id">ejournal.kemenperin.go.id</a> Internet Source	<1 %
8	Basiroh Basiroh, Henry Irjananto. "C4.5 ALGORITHM AS A DECISION SUPPORT SYSTEM FOR SOCIAL WELFARE AID RECIPIENTS", Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN), 2024 Publication	<1 %
9	<a href="http://ilmudatapy.com">ilmudatapy.com</a> Internet Source	<1 %
10	<a href="http://jurnal.fmipa.unmul.ac.id">jurnal.fmipa.unmul.ac.id</a> Internet Source	<1 %
11	<a href="http://www.poltekindonusa.ac.id">www.poltekindonusa.ac.id</a> Internet Source	<1 %
12	<a href="http://publikasi.dinus.ac.id">publikasi.dinus.ac.id</a> Internet Source	<1 %
13	Submitted to British University In Dubai Student Paper	<1 %
14	Baktiar Setiyaji, Teguh Susyanto, Dwi Remawati. "SISTEM PAKAR DIAGNOSA KERUSAKAN MOTOR 4 TAK HONDA VARIO DENGAN ALGORITMA C4.5", Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN), 2018 Publication	<1 %



15

Submitted to Konsorsium Perguruan Tinggi Swasta Indonesia II

Student Paper

<1 %

---

16

Muhammad Fikri Indriawan, Julius Panda Putra Naibaho, Marlinda Sanglise. "Analisis Komparasi Algoritma Naïve Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa (Studi Kasus: Jurusan S1 Teknik Informatika Universitas Papua)", JISTECH: Journal of Information Science and Technology, 2023

Publication

<1 %

---

17

[journal.itk.ac.id](http://journal.itk.ac.id)

Internet Source

<1 %

---

18

[universitasbumigora.ac.id](http://universitasbumigora.ac.id)

Internet Source

<1 %

---

19

Hardiyan Hardiyan, Nadiyah Nadiyah. "Penerapan Algoritma C4.5 Dalam Pemberian Kelayakan Kredit Motor", Reputasi: Jurnal Rekayasa Perangkat Lunak, 2022

Publication

<1 %

---

20

[www.polgan.ac.id](http://www.polgan.ac.id)

Internet Source

<1 %

---

21

Amrin Amrin, Irawan Satriadi, Oki Rosanto. "ALGORITMA C4.5 UNTUK DIAGNOSA PENYAKIT TUBERKULOSIS", Jurnal Khatulistiwa Informatika, 2019

Publication

<1 %

---

22	<a href="http://digilibadmin.unismuh.ac.id">digilibadmin.unismuh.ac.id</a> Internet Source	<1 %
23	<a href="http://scite.ai">scite.ai</a> Internet Source	<1 %
24	<a href="http://ars.itenas.ac.id">ars.itenas.ac.id</a> Internet Source	<1 %
25	<a href="http://citee.ft.ugm.ac.id">citee.ft.ugm.ac.id</a> Internet Source	<1 %
26	<a href="http://edoc.pub">edoc.pub</a> Internet Source	<1 %
27	<a href="http://jti.respati.ac.id">jti.respati.ac.id</a> Internet Source	<1 %
28	<a href="http://mafiadoc.com">mafiadoc.com</a> Internet Source	<1 %
29	<a href="http://journal.unimar-amni.ac.id">journal.unimar-amni.ac.id</a> Internet Source	<1 %
30	<a href="http://semnas.univbinainsan.ac.id">semnas.univbinainsan.ac.id</a> Internet Source	<1 %
31	<a href="http://www.sciencegate.app">www.sciencegate.app</a> Internet Source	<1 %
32	Hamdi Indra, Devi Yuliana, Ulia Ulfa, Riki Adriant. "Strategi Promosi Penjualan Menggunakan Algoritma C4.5 (Studi Kasus Di	<1 %

# Restoran Plaza Pekanbaru)", Jurnal Minfo Polgan, 2023

Publication

33

Irwan Agus Sobari, Robi Aziz Zuama. "Pendekatan Machine Learning dalam Memprediksi Keluarga Penerima Program PKH", Jurnal Teknik Komputer, 2023

Publication

<1 %

34

Submitted to St. Petersburg High School

Student Paper

<1 %

35

indrakwary.wordpress.com

Internet Source

<1 %

36

Submitted to Australian National University

Student Paper

<1 %

37

Submitted to Lambung Mangkurat University

Student Paper

<1 %

38

Muh. Falach Achsan Yusuf. "Klasifikasi Gambar Burung Konservasi di Wilayah Papua Barat Menggunakan Transfer Learning", Indonesian Journal of Computer Science, 2024

Publication

<1 %

39

Submitted to Universitas Negeri Medan

Student Paper

<1 %

40

journal.amikmahaputra.ac.id

Internet Source

<1 %

mx2.atmajaya.ac.id

41

Internet Source

&lt;1 %

42

Fitriyani Fitriyani. "Metode Bagging Untuk Imbalance Class Pada Bedah Toraks Menggunakan Naive Bayes", Jurnal Kajian Ilmiah, 2018

Publication

&lt;1 %

43

Submitted to Fulton County School District

Student Paper

&lt;1 %

44

I Made Dananjaya Priyatama, Ridwansyah Ridwansyah. "KLASIFIKASI ANAK BERKEBUTUHAN KHUSUS TUNAGRAHITA MENGGUNAKAN METODE ALGORITMA C4.5", Paradigma - Jurnal Komputer dan Informatika, 2022

Publication

&lt;1 %

45

Maylanny Christin, Rico Kurnia Yudhaswara, Dasrun Hidayat. "Description of Selective Behavioral Experience of Choosing Covid-19 information In Mass Media Television", JURNAL PENELITIAN KOMUNIKASI DAN OPINI PUBLIK, 2021

Publication

&lt;1 %

46

Submitted to University of Hong Kong

Student Paper

&lt;1 %

47

[eprints.sinus.ac.id](http://eprints.sinus.ac.id)

Internet Source

&lt;1 %

48	<a href="https://fdocuments.us">fdocuments.us</a> Internet Source	<1 %
49	<a href="https://journals.sagepub.com">journals.sagepub.com</a> Internet Source	<1 %
50	<a href="https://lib.geo.ugm.ac.id">lib.geo.ugm.ac.id</a> Internet Source	<1 %
51	<a href="https://ppjp.ulm.ac.id">ppjp.ulm.ac.id</a> Internet Source	<1 %
52	Bayu Ferdiansyah, Leonard Goeirmanto. "Prediksi Loyalitas dalam Keterikatan Karyawan terhadap Perusahaan Menggunakan Algoritma C4.5* (Studi Kasus PT.XYZ)", Jurnal Sistem dan Teknologi Informasi (JUSTIN), 2020 Publication	<1 %
53	<a href="https://ojs.unud.ac.id">ojs.unud.ac.id</a> Internet Source	<1 %
54	<a href="https://repository.unpad.ac.id">repository.unpad.ac.id</a> Internet Source	<1 %
55	Aji Supriyanto, Jeffry Alfa Razaq, Purwatingtyas Purwatingtyas, Agus Ariyanto. "Keputusan Pemberian Bantuan Sosial Program Keluarga Harapan Menggunakan Metode AHP dan SAW", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022	<1 %

56

D Arifin, A Hadiana. "Computer-based Techniques for Predicting the Failure of Student Studies Using the Decision Tree method", IOP Conference Series: Materials Science and Engineering, 2019

Publication

---

<1 %

57

Edi Sukmara, Idan Setiari. "Nilai Didaktis Folklor "Lutung Kasarung" Karya Ki Raksa Sunda", Jurnal Soshum Insentif, 2019

Publication

---

<1 %

58

Julkari Sinta Parapat, Anita Sindar RMS. "Data Mining Klasifikasi Data Nasabah Kredit KSU Taman Mandiri Menggunakan Algoritma C4.5", Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, 2019

Publication

---

<1 %

59

Sevni Tomatala, Wiclif Sepnath Pinoa, Johan Riry. "Peran Program Keluarga Harapan Dalam Memberikan Perlindungan Sosial Kepada Masyarakat Di Negeri Kamal Kecamatan Kairatu Barat Kabupaten Seram Bagian Barat", Jurnal Pendidikan Geografi Unpatti, 2023

Publication

---

<1 %

60

[journal.widyakarya.ac.id](http://journal.widyakarya.ac.id)

Internet Source

---

<1 %



61	<a href="http://jurnal.unai.edu">jurnal.unai.edu</a> Internet Source	<1 %
62	<a href="http://jurnal.univbinainsan.ac.id">jurnal.univbinainsan.ac.id</a> Internet Source	<1 %
63	<a href="http://jurnal.uts.ac.id">jurnal.uts.ac.id</a> Internet Source	<1 %
64	<a href="http://zombiedoc.com">zombiedoc.com</a> Internet Source	<1 %
65	Fietri Setiawati Sulaeman, Mufti Ahmad Rilmansyah. "Aplikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Berbasis Web Menggunakan Algoritma C45", Jurnal Media Teknik dan Sistem Industri, 2021 Publication	<1 %
66	I Gede Iwan Sudipa, I Nyoman Alit Arsana, Made Leo Radhitya. "PENENTUAN TINGKAT PEMAHAMAN MAHASISWA TERHADAP SOCIAL DISTANCING MENGGUNAKAN ALGORITMA C4.5", SINTECH (Science and Information Technology) Journal, 2020 Publication	<1 %
67	Muhammad Rizki Fahdia, Dwiza Riana, Fachri Amsury, Irwansyah Saputra, Nanang Ruhyana. "Komparasi Algoritma Klasifikasi untuk Orientasi Minat Mahasiswa dalam	<1 %

# Penuntasan Studi", JIRA: Jurnal Inovasi dan Riset Akademik, 2021

Publication

68

Sri Widaningsih. "PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4,5, NAÏVE BAYES, KNN DAN SVM", Jurnal Tekno Insentif, 2019

Publication

<1 %

69

Yohana Tri Widayati, Yani Prihati, Stephanus Widjaja. "ANALISIS DAN KOMPARASI ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI LOYALITAS PELANGGAN MNC PLAY KOTA SEMARANG", Jurnal Transformatika, 2021

Publication

<1 %

70

[digilib.unimed.ac.id](http://digilib.unimed.ac.id)

Internet Source

<1 %

71

[idoc.tips](http://idoc.tips)

Internet Source

<1 %

72

[ieomsociety.org](http://ieomsociety.org)

Internet Source

<1 %

73

[journal.lembagakita.org](http://journal.lembagakita.org)

Internet Source

<1 %

74

[repository.ar-raniry.ac.id](http://repository.ar-raniry.ac.id)

Internet Source

<1 %

75	<a href="https://repository.pnj.ac.id">repository.pnj.ac.id</a> Internet Source	<1 %
76	<a href="https://teknosi.fti.unand.ac.id">teknosi.fti.unand.ac.id</a> Internet Source	<1 %
77	Ardhin Primadewi. "Using Data Mining with C4.5 Algorithm for Student Department Selection at MTs N Kaliangkrik", Borobudur Informatics Review, 2021 Publication	<1 %
78	Donni Prabowo, Firman Hidayat, Gagah Gumelar, Dewa Qintoro, Aji Setiawan. "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN C4.5 DALAM MENENTUKAN TINGKAT PENJUALAN MOTOR HONDA", Jurnal Informatika Komputer, Bisnis dan Manajemen, 2023 Publication	<1 %
79	Istikhomah, Radenrara Imro'atun. "Implementasi supervisi klinis dalam pengembangan kemampuan menyusun kisi-kisi evaluasi pembelajaran Pendidikan Agama Islam di SMP Negeri 1 Baturaden", Universitas Islam Negeri Saifuddin Zuhri (Indonesia), 2024 Publication	<1 %
80	Muhammad Kamil, Widya Cholil. "Analisis Perbandingan Algoritma C4.5 dan Naive	<1 %

Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang", Jurnal Informatika, 2020

Publication

---

81

Nurul Afni, Melan Susanti, Zulfajri Zulfajri. "Prediksi Kepuasan Tenant Pada Gedung Wisma Keiai Menggunakan Algoritma C4.5", Jurnal Infortech, 2020

Publication

---

<1 %

82

Sandrawira Anggraini, Sarjon Defit, Gunadi Widi Nurcahyo. "Data Mining Analysis Sale Transaction With Algorithm C4.5", Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, 2019

Publication

---

<1 %

83

Sri Lestari, Bela Dina. "KLASIFIKASI KETEPATAN KELULUSAN SISWA PADA SMK YADIKA 9 BINTARA JAYA KOTA BEKASI MENGGUNAKAN ALGORITMA C4.5", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2023

Publication

---

<1 %

84

Tabitha Salsabilla, Sulastri Sulastri. "IMPLEMENTASI ALGORITMA C4.5 UNTUK KLASIFIKASI PRODUK LARIS SEPEDA MOTOR HONDA PADA CV CENDANA MOTOR CEPIRING", Rabbit : Jurnal Teknologi dan Sistem Informasi Univrab, 2022

Publication

<1 %

---

85	<a href="http://akbidbinahusada.ac.id">akbidbinahusada.ac.id</a> Internet Source	<1 %
86	<a href="http://cah-mbudur.blogspot.com">cah-mbudur.blogspot.com</a> Internet Source	<1 %
87	<a href="http://fti.ars.ac.id">fti.ars.ac.id</a> Internet Source	<1 %
88	<a href="http://gpriority.co.id">gpriority.co.id</a> Internet Source	<1 %
89	<a href="http://idr.uin-antasari.ac.id">idr.uin-antasari.ac.id</a> Internet Source	<1 %
90	<a href="http://journal.sekawan-org.id">journal.sekawan-org.id</a> Internet Source	<1 %
91	<a href="http://jurnal.harapan.ac.id">jurnal.harapan.ac.id</a> Internet Source	<1 %
92	<a href="http://jurnal.uinsu.ac.id">jurnal.uinsu.ac.id</a> Internet Source	<1 %
93	<a href="http://jurnalmahasiswa.unesa.ac.id">jurnalmahasiswa.unesa.ac.id</a> Internet Source	<1 %
94	<a href="http://openarchives.library.cornell.edu">openarchives.library.cornell.edu</a> Internet Source	<1 %
95	<a href="http://qjtd.atu.ac.ir">qjtd.atu.ac.ir</a> Internet Source	<1 %
96	<a href="http://repository.bsi.ac.id">repository.bsi.ac.id</a> Internet Source	<1 %

---

97

Ade Purwanto, Aji Primajaya, Apriade Voutama. "Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang", Jurnal Sistem dan Teknologi Informasi (Justin), 2020

Publication

&lt;1 %

98

Anggita Safitri Febriarini, Erna Zuni Astuti. "Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang", Eksplora Informatika, 2019

Publication

&lt;1 %

99

Listy Julfani, Ilham Mirzaya Putra. "Dampak Program Keluarga Harapan (PKH) terhadap Peningkatan Kesejahteraan Keluarga Miskin di Kelurahan Kerasaan I Kecamatan Pematang Bandar Kabupaten Simalungun", Al-Kharaj : Jurnal Ekonomi, Keuangan & Bisnis Syariah, 2023

Publication

&lt;1 %

100

Muhamad Yunus, Hanandriya Ramadhan, Dimas Rizki Aji, Agus Yulianto. "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)", Paradigma - Jurnal Komputer dan Informatika, 2021

Publication

&lt;1 %



101 A A Ayu Wulan Agustini, I Putu Mahendra Adi Wardana, Kadek Oky Sanjaya. "Penerapan Data Mining Dengan Algoritma C4.5 Dalam Penentuan Kelayakan Nasabah Kredit", RESI : Jurnal Riset Sistem Informasi, 2023  
Publication <1 %

---

102 Syahrul Mubarak Nur Muhammad, Ferdi Ahmad Mauladi, Risky Kurniawan, Rangga Sanjaya. "Pengembangan Sistem Informasi Kawasan Agrowisata Menggunakan Konsep Model View Control berbasis Web", Edumatic: Jurnal Pendidikan Informatika, 2022  
Publication <1 %

---

103 Taufik Hidayatulloh, Anisa Fajria, Rida Nutria Lestari, Neng Sella Zakiatun Nufus. "Algoritma C4.5 Untuk Menentukan Kelayakan Pemberian Kredit (Studi kasus: Bank Mandiri Taspen Kantor Kas Sukabumi)", Jurnal Larik: Ladang Artikel Ilmu Komputer, 2022  
Publication <1 %

---

104 [ijns.org](http://ijns.org)  
Internet Source <1 %

---

105 [jumanji.unjani.ac.id](http://jumanji.unjani.ac.id)  
Internet Source <1 %

---

Exclude bibliography On