



Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok

Susi Fitryah Damanik¹, Anjar Wanto², Indra Gunawan³

^{1,2,3}Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

¹susidamanik87@gmail.com, ²anjarwanto@amiktunasbangsa.ac.id, ³indra@amiktunasbangsa.ac.id

Received on 2 Desember 2021	Revised on 14 Januari 2022	Accepted on 22 Januari 2022
--------------------------------	-------------------------------	--------------------------------

Abstract

The classification of the level of family welfare in Tiga Dolok Village is a problem experienced by the community in the village. Where the classification of the level of family welfare in the village is not completely accurate so that the distribution of government subsidies is not on target. The issue of welfare classification is the goal of conducting research in order to obtain accurate results in the status of the level of family welfare. To overcome this problem a new model is proposed by utilizing a C4.5 computational method to produce an accurate classification of welfare levels. In this study the algorithm used to classify the level of welfare of the Three Dolok Village is the C4.5 algorithm. This algorithm was chosen because the classification process is simple and fast. The research data used will be the Basic Data of the Three Dolok Village Family Data in 2019. The data source was obtained based on a questionnaire distributed to the Tiga Dolok community. Based on this data a welfare level classification will be carried out using the rapid miner application. With this method a decision tree will be formed so that later it will get the desired classification results.

Keywords: Classification, Tiga Dolok Village Family, Algorithm C4.5

Abstrak

Klasifikasi tingkat kesejahteraan keluarga di Desa Tiga Dolok merupakan permasalahan yang dialami Masyarakat di desa itu. Dimana klasifikasi Tingkat kesejahteraan keluarga di Desa tersebut belum sepenuhnya akurat sehingga mengakibatkan penyaluran subsidi pemerintah tidak tepat sasaran. Permasalahan klasifikasi tingkat kesejahteraan menjadi tujuan dilakukannya penelitian agar mendapatkan hasil yang akurat dalam status tingkat kesejahteraan keluarga. Untuk mengatasi masalah tersebut diusulkan model baru dengan memanfaatkan sebuah metode komputasi C4.5 agar menghasilkan klasifikasi tingkat kesejahteraan yang akurat. Pada penelitian ini algoritma yang digunakan untuk melakukan klasifikasi tingkat kesejahteraan pada Desa Tiga Dolok adalah algoritma C4.5. Algoritma ini dipilih karena proses klasifikasinya sederhana dan cepat. Data penelitian yang digunakan nantinya adalah Data Isian Dasar Keluarga Desa Tiga Dolok Tahun

2019. Sumber data diperoleh berdasarkan kuisisioner yang dibagikan kepada masyarakat Tiga Dolok. Berdasarkan data ini akan dilakukan klasifikasi tingkat kesejahteraan dengan menggunakan aplikasi rapid miner. Dengan metode ini akan dibentuk pohon keputusan agar nantinya mendapatkan hasil klasifikasi yang diinginkan.

Kata Kunci: Klasifikasi, Keluarga Desa Tiga Dolok, Algoritma C4.5

PENDAHULUAN

Keluarga ialah bagian terkecil dari masyarakat yang terdiri dari Ayah, Ibu dan Anak sedangkan kesejahteraan merupakan hal yang didambakan setiap orang. Dimana sejahtera merupakan kondisi kebutuhan di dalam keluarga telah terpenuhi dengan baik. Tingkat kesejahteraan keluarga sangat berpengaruh pada kualitas suatu keluarga, dalam mencapai tujuan keluarga sejahtera itu sendiri dibutuhkan peran anggota keluarga tentunya dalam hal untuk mendapatkan predikat keluarga sejahtera.

Dalam penelitian ini penulis mengamati penyaluran subsidi pemerintah di Desa Tiga Dolok hingga saat ini masih menjadi permasalahan di wilayah itu. Masih sering sekali terjadi, warga yang mendapat subsidi pemerintah adalah warga yang masih mampu sedangkan warga yang ekonominya lemah justru tidak mendapat subsidi.

Dengan demikian tingkat kesejahteraan keluarga itu penting untuk diklasifikasi dengan tujuan penentuan target masyarakat miskin. Agar subsidi pemerintah dapat disalurkan secara tepat sasaran. Oleh sebab itu penulis memberikan solusi untuk mengklasifikasi tingkat kesejahteraan keluarga dengan menggunakan algoritma C4.5. Algoritma ini dipilih karena merupakan salah satu metode pada Decision Tree / Pohon Keputusan yang banyak dimanfaatkan untuk melakukan prediksi terhadap suatu kasus.

Ada beberapa penelitian terdahulu terkait dengan penggunaan algoritma C4.5 maupun yang berhubungan dengan tingkat kesejahteraan keluarga, diantaranya:

- 1) Dalam pengumpulan data PLKB masih menggunakan cara manual yaitu menulis data per kepala keluarga [1]. Setelah itu, data baru diproses satu persatu untuk menghitung tingkatan keluarga sejahtera. Banyaknya data yang dikumpulkan oleh PLKB menyulitkan untuk dilakukan klasifikasi data kedalam tingkatan kesejahteraan keluarga yang dapat memperlambat kinerja PLKB. Untuk menentukan tingkatan keluarga sejahtera dihitung berdasarkan 21 indikator yang dikelompokkan menjadi 5 tahap keluarga sejahtera. Pola-pola dalam kesejahteraan keluarga dapat diklasifikasikan menggunakan metode data mining dengan algoritma C4.5, menghasilkan aturan klasifikasi yang menghasilkan tahapan keluarga sejahtera. Metode klasifikasi algoritma C4.5 berhasil diimplementasikan pada data kesejahteraan keluarga dusun Joho kecamatan Pringapus, kabupaten Semarang. Dari data tersebut menghasilkan nilai akurasi sebesar 95%. Sehingga penelitian dikatakan berhasil karena dapat membantu meringankan tugas PLKB yaitu mengklasifikasikan data kedalam jenis tingkatan kesejahteraan keluarga menggunakan pohon keputusan data mining berbasis algoritma C4.5.
- 2) Klasifikasi Pemanfaat Program Beras Sejahtera (RASTRA) Berdasarkan Tingkat Kemiskinan Dengan Menggunakan Algoritma Decision Tree C4.5 Berbasis Particle Swarm Optimization [2]. Penulis mengambil

kesimpulan, penyaluran bantuan Pemerintah kepada masyarakat melalui Program Beras Sejahtera (RASTRA) seringkali tidak tepat sasaran, banyak keluarga yang seharusnya memperoleh bantuan tidak mendapatkan bantuan, justru keluarga yang cukup ekonomi mendapat bantuan

- 3) Implementasi algoritma C4.5 dalam studi kasus system evaluasi Suntuk menentukan kelulusan suatu proposal pembukaan program studi baru di Kementerian Riset, Teknologi dan Pendidikan Tinggi [3]. Penulis mengambil kesimpulan bahwa Algoritma C4.5 bertindak sama seperti algoritma ID3 akan tetapi memperbaiki beberapa kekurangan dari ID3 yaitu:
 - a. Memungkinkan untuk menggunakan data yang continues.
 - b. Dapat memprediksi nilai yang hilang.
 - c. Mampu menggunakan atribut dengan bobot yang berbeda.
 - d. Memperbaiki pohon keputusan yang telah ada.

METODE PENELITIAN

Metode yang akan di gunakan peneliti adalah metode survey yaitu pengamatan dan pengambilan sampel data langsung di lapangan dengan membagikan kuisisioner langsung kepada responden. Jenis data yang dikumpulkan dalam penelitian ini berupa data primer yang diperoleh dari hasil wawancara langsung dengan responden menggunakan kuisisioner. Data yang dikumpulkan yaitu Data Dasar Keluarga yang dibagikan kepada warga Desa Tiga Dolok.

Algoritma C4.5 adalah salah satu metode klasifikasi dari data mining yang digunakan untuk mengkonstruksikan pohon keputusan (decision tree) [4]. Algoritma C4.5 adalah program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran [5]. Pohon keputusan tindak lanjut ini kemudian diverifikasi terhadap data uji berlabel yang tidak terlihat untuk menghitung generalisasinya. C4.5 adalah program yang digunakan untuk menghasilkan peraturan taksonomi dengan menggunakan pohon keputusan dari sekumpulan data yang diberikan. Algoritma C4.5 merupakan perpanjangan dari algoritma ID3 dasar dan dirancang oleh Quinlan. C4.5 adalah salah satu algoritma pembelajaran yang banyak digunakan. Algoritma C4.5 membangun pohon keputusan dari serangkaian data pelatihan yang mirip dengan Algoritma ID3, dengan menggunakan konsep entropi informasi. C4.5 juga dikenal sebagai klasifikasi statistik.

Algoritma C4.5 merupakan metode yang menjadi pilihan pertama dan sering digunakan dalam pengembangan Data Mining karena kecepatan dalam pengklasifikasian pohon keputusan disamping dapat mengkonstruksi pengklasifikasian dengan aturan-aturan yang lain [6]. Juga Algoritma C4.5 merupakan kelompok algoritma Decision Tree, Algoritma ini mempunyai input berupa training samples dan samples. training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.

Dari beberapa defenisi tentang algoritma C4.5 diatas dapat dikemukakan bahwa algoritma C4.5 adalah salah satu metode klasifikasi dari data mining yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran yang mempunyai input berupa training samples dan sample training samples

berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.

Adapun atribut yang akan menjadi prediktor untuk menghasilkan atribut target, yaitu (1) Pendapatan (2) Penerima Beras Miskin (3) Penerima BLT / BLSM (4) Peserta BPJS dan (5) Status Kepemilikan Rumah. Dimana atribut target tersebut menjadi class output untuk menentukan penerima subsidi yaitu dibedakan menjadi (Sejahtera dan Tidak Sejahtera) Pada tahap selanjutnya yaitu data ditransformasi kedalam format Microsoft Excel, lalu akan di proses menggunakan aplikasi perangkat lunak Rapid Miner untuk pengujiannya menggunakan cross validation dalam Rapid Miner, maka akan terlihat akurasi, pohon keputusan dan rulenya.

ANALISA DATA

Data yang digunakan peneliti merupakan data berdasarkan survey yang dilakukan pada Nagori Tiga Dolok dengan kuisioner yang dibagikan kepada masyarakat Desa Tiga Dolok. Data yang dikumpulkan sebanyak 357 data kuisioner, dan penulis mengambil sampel sebanyak 249 data.

Dalam melakukan penelitian ini penulis menggunakan Microsoft excel dan Rapid Miner sebagai alat untuk menganalisis data, Microsoft Excel digunakan untuk membantu dalam proses perhitungan. Setelah proses perhitungan selesai maka selanjutnya akan diproses menggunakan Tools Rapid Miner dan untuk pengujiannya menggunakan cross validation yang ada di dalam Rapid Miner.

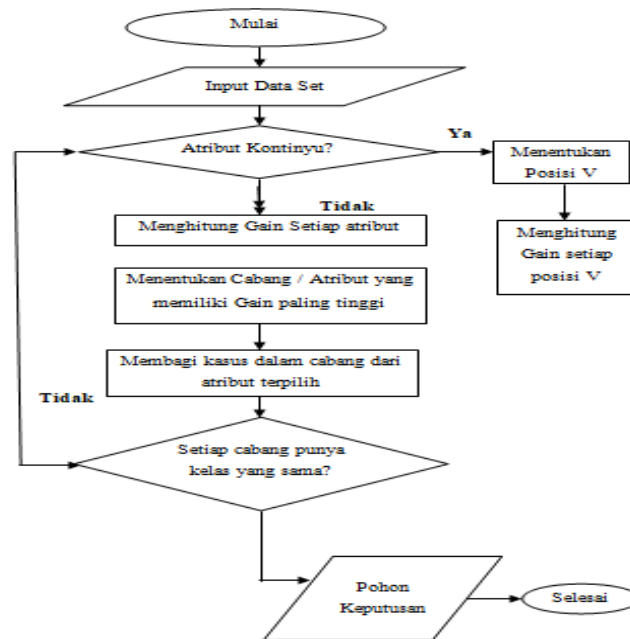
INSTRUMEN PENELITIAN

Peneliti melakukan beberapa proses pengumpulan data dan pengolahan data dengan cara memasukkan data memproses dengan metode decision tree C4.5 sehingga ditemukan klasifikasi Sejahtera dan Tidak Sejahtera. Sebelum diproses data terlebih dahulu di seleksi menggunakan Rapid Miner kemudian disimpan hasil klasifikasi yang telah didapat. Proses percobaan dapat dilihat pada Gambar 1.

Keterangan dari prosedur penelitian diatas adalah sebagai berikut :

- 1) Dari 357 responden, maka penulis mengambil sampel data sebanyak 70,2 % jumlah data kepala keluarga yang ada pada Desa Tiga Dolok yaitu sebanyak 249 orang responden.
- 2) Menentukan atribut sebagai node. Atribut yang digunakan pada penelitian ini yaitu : Nama KP, Pendapatan , Jumlah pengeluaran perbulan, Penerima beras miskin, Penerima BLT/BLSM, Peserta BPJS/Jamkesmas dan Status kepemilikan rumah.
- 3) Menghitung gain untuk setiap atribut.
- 4) Menentukan Cabang tiap nilai maka dilakukan perhitungan rasio gain untuk memilih pilihan percabangan terbaik.
- 5) Membagi kasus cabang pada dari atribut terpilih
- 6) Lakukan proses berulang untuk setiap cabang hingga semua kasus dalam cabang mendapat kelas yang sama.

- 7) Hasil penelitian ini menghasilkan pohon keputusan klasifikasi tingkat kesejahteraan keluarga pada desa Tiga Dolok dengan mendapatkan hasil keputusan sejahtera sebanyak 19 kepala keluarga dan 230 kepala keluarga untuk keputusan tidak sejahtera.



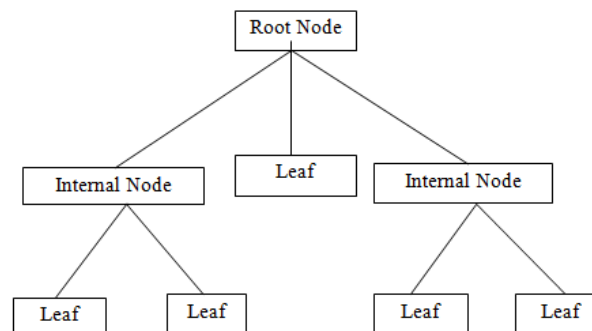
Gambar 1. Proses Percobaan

PEMODELAN METODE

Teknik dalam pohon keputusan ialah data diubah menjadi sebuah pohon keputusan yang di dalamnya terdapat aturan keputusan. Terdapat tiga jenis node, dalam pohon keputusan antara lain:

1. Akar
Menjadi node paling atas setelah dilakukan proses perhitungan gain dan entropy.
2. Internal node
Yaitu nilai dari setiap atribut, merupakan anak dari struktur pohon keputusan tetapi bukan daun.
3. Daun
Merupakan simpul terakhir dari pohon keputusan.

Model pohon keputusan yang akan digunakan tampak seperti pada Gambar 2.



Gambar 2. Model Decision Tree

HASIL DAN PEMBAHASAN

MENYIAPKAN DATA

Berikut adalah Tabel 1 data yang diolah menggunakan Microsoft Office Excel.

Table 1. Data Isian Dasar Keluarga
(Sumber data : Kantor Desa Tiga Dolok)

No	Nama KP	Jumlah Penghasilan Perbulan	Jumlah Pengeluaran Perbulan	Penerima Beras Miskin	Penerima BLT/ BLSM	Peserta BPJS/ Jamkes mas	Status Kepemilikan Rumah
1	Penneria Ambarita	Rendah	Sedang	Tidak	Tidak	Ya	Milik Sendiri
2	Jongir Sidabutar	Sedang	Tinggi	Ya	Tidak	Ya	Milik Sendiri
3	Pangihutan Situmeang	Tinggi	Sedang	Tidak	Tidak	Ya	Milik Sendiri
4	Ramses Purba	Sedang	Tinggi	Ya	Tidak	Ya	Milik Orang Tua
5	Pontas Rajaguk-guk	Tinggi	Sedang	Tidak	Tidak	Tidak	Milik Sendiri
6	Ruhtnawati Parapat	Rendah	Sedang	Ya	Tidak	Tidak	Milik Sendiri
7	Rummina Marpaung	Rendah	Rendah	Tidak	Tidak	Ya	Milik Sendiri
8	Lintong Benekditus Siallagan	Rendah	Rendah	Tidak	Tidak	Tidak	Kontrak
9	Longser Bakkara	Tinggi	Sedang	Ya	Tidak	Ya	Milik Sendiri
10	Hengki Tarunis Sinurat	Sedang	Rendah	Tidak	Tidak	Tidak	Pinjam Pakai
11	Hotnel Samosir	Tinggi	Sedang	Tidak	Tidak	Ya	Milik Sendiri
12	Merwin Gatimor Napitu	Tinggi	Sedang	Tidak	Tidak	Tidak	Milik Sendiri
13	Gusmin Napitu	Tinggi	Sedang	Ya	Ya	Ya	Milik Sendiri

Nb: data selanjutnya tidak semua di tampilkan

PENGOLAHAN DATA

Table 2. Perhitungan Gain dan Entropy Menggunakan Microsoft Excel

	Jumlah (s)	Sejahtera	Tidak Sejahtera	Entropy	Gain
Total	249	150	99	0.969523583	
Pendapatan					0.935396504
Rendah	202	0	202	0	
Sedang	15	2	13	0.566509507	
Tinggi	32	17	1	0	
Penerima Raskin					0.762723141
Ya	89	0	89	0	
Tidak	160	3	134	0.321833187	
Penerima BLT/BLSM					0.724768015
Ya	191	191	0	0	
Tidak	58	19	25	1.050760971	
Peserta BPJS/Jamkesmas					0.522664578
Ya	101	4	97	0.240473407	
Tidak	148	15	119	0.587703231	

Pada Tabel 2 diuraikan secara rinci hasil perhitungan *entropy* dan *gain* dengan menggunakan *microsoft excel*. Setelah dilakukan perhitungan maka didapatkan nilai gain tertinggi terdapat pada variabel Pendapatan dengan nilai *gain* sebesar 0.9355396504 maka Pendapatan dipilih sebagai *node* akar. Perhitungan algoritma C4.5 untuk memperoleh hasil klasifikasi pohon keputusan dapat diuraikan sebagai berikut :

Langkah 1 : Menghitung jumlah kasus, jumlah kasus untuk keputusan Sejahtera dan jumlah kasus untuk keputusan Tidak Sejahtera.

Langkah 2 : Menghitung *Entropy* untuk semua kasus dan kasus yang dibagi berdasarkan kelas atribut. Selanjutnya dilakukan perhitungan *Gain* untuk untuk masing-masing atribut dengan. Berikut ini adalah perhitungan nilai *entropy* dan *gain*.

Menghitung entropy total :

$$\begin{aligned} \text{Entropy [Total]} &= \left(-\frac{150}{249} * \log_2 \left(\frac{150}{249}\right)\right) + \left(-\frac{99}{249} * \log_2 \left(\frac{99}{249}\right)\right) \\ &= 0.440472 + 0.529052 \\ &= 0.9695236 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Pendapatan-Rendah]} &= \left(-\frac{0}{202} * \log_2 \left(\frac{0}{202}\right)\right) + \left(-\frac{202}{202} * \log_2 \left(\frac{202}{202}\right)\right) \\ &= 0.510059 + 0.488173 \\ &= 0.9982314 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Pendapatan-Sedang]} &= \left(-\frac{2}{15} * \log_2 \left(\frac{2}{15}\right)\right) + \left(-\frac{13}{15} * \log_2 \left(\frac{13}{15}\right)\right) \\ &= 0.389975 + 0.528321 \\ &= 0.9182958 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Pendapatan-Tinggi]} &= \left(-\frac{17}{32} * \log_2 \left(\frac{17}{32}\right)\right) + \left(-\frac{1}{32} * \log_2 \left(\frac{1}{32}\right)\right) \\ &= 0 + 0 \\ &= 0 \end{aligned}$$

Menghitung Gain untuk nilai Pendapatan

$$\begin{aligned} &= 0.9695236 - \left(\frac{0}{249}\right) * 0.9982314 - \left(\frac{202}{249}\right) * 0.9182958 - 0 \\ &= 0.10439431 \end{aligned}$$

Menghitung Entropy untuk atribut Penerima Raskin

$$\begin{aligned} \text{Entropy [Penerima Raskin-Ya]} &= \left(-\frac{0}{89} * \log_2 \left(\frac{0}{89}\right)\right) + \left(-\frac{89}{89} * \log_2 \left(\frac{89}{89}\right)\right) \\ &= 0 + 0 \\ &= 0 \end{aligned}$$

Entropy [Penerima Raskin-Tidak]

$$\begin{aligned} &= \left(-\frac{97}{160} * \log_2 \left(\frac{97}{160}\right)\right) + \left(-\frac{63}{160} * \log_2 \left(\frac{63}{160}\right)\right) \\ &= 0.437722 + 0.529455 \end{aligned}$$

$$= 0.967177$$

Menghitung Gain untuk nilai Penerima Raskin

$$= 0.9695236 - \left(\frac{0}{89}\right) * 0 - \left(\frac{97}{160}\right) * 0.967177$$

$$= 0.304418736$$

Menghitung Entropy untuk atribut Penerima BLT/BLSM

Entropy [Penerima BLT/BLSM-Ya]

$$= \left(-\frac{0}{73} * \log_2 \left(\frac{0}{73}\right)\right) + \left(-\frac{73}{73} * \log_2 \left(\frac{73}{73}\right)\right)$$

$$= 0 + 0$$

$$= 0$$

Entropy [Penerima BLT/BLSM-Tidak]

$$= \left(-\frac{113}{176} * \log_2 \left(\frac{113}{176}\right)\right) + \left(-\frac{63}{176} * \log_2 \left(\frac{63}{176}\right)\right)$$

$$= 0.410429 + 0.530543$$

$$= 0.9409722$$

Menghitung Gain untuk nilai Penerima BLT/BLSM

$$= 0.9695236 - \left(\frac{73}{249}\right) * 0 - \left(\frac{176}{249}\right) * 0.9409722$$

$$= 0.30441874$$

Menghitung Entropy untuk atribut Peserta BPJS/Jamkesmas

Entropy [Peserta BPJS/Jamkesmas-Ya]

$$= \left(-\frac{63}{108} * \log_2 \left(\frac{63}{108}\right)\right) + \left(-\frac{45}{108} * \log_2 \left(\frac{45}{108}\right)\right)$$

$$= 0.453604 + 0.526264$$

$$= 0.9798688$$

Entropy [Peserta BPJS/Jamkesmas-Tidak]

$$= \left(-\frac{87}{141} * \log_2 \left(\frac{87}{141}\right)\right) + \left(-\frac{54}{141} * \log_2 \left(\frac{54}{141}\right)\right)$$

$$= 0.429822 + 0.530297$$

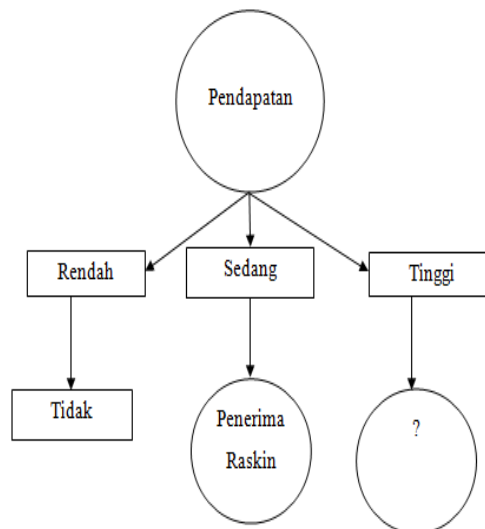
$$= 0.9601187$$

Menghitung Gain untuk nilai Peserta BPJS/Jamkesmas

$$= 0.9695236 - \left(\frac{108}{249}\right) * 0.979869 - \left(\frac{141}{249}\right) * 0.960119$$

$$= 0.00083861$$

Dari rangkuman perhitungan entropy dan gain pada tabel 2. diatas maka diperoleh root node seperti terlihat pada Gambar 3.



Gambar 3. Menentukan node 1

Berdasarkan Gambar 3 atribut Pendapatan digunakan sebagai node akar. Nilai atribut Rendah telah terklasifikasi yaitu Tidak Sejahtera sedangkan nilai atribut Sedang terklasifikasi yaitu dengan atribut Penerima Raskin dan nilai atribut Tinggi belum terklasifikasi. Demikian juga untuk atribut yang lainnya akan dilakukan proses perhitungan yang sama.

PROSES PENGUJIAN DENGAN RAPID MINER

Proses pengujian dengan menggunakan metode decision tree C4.5 pada aplikasi rapidminer dapat dilihat seperti tahapan dibawah ini :

- Langkah 1 : Membuka aplikasi Rapidminer.
- Langkah 2 : Pilih new process untuk melakukan process baru. Tampilan new process dapat dilihat seperti gambar 5
- Langkah 3 : Memilih Read Excel dibagian operator kemudian drag and drop operator Read Excel ke dalam lembar kerja main process sebanyak 2 operator untuk menstransformasi data training dan data testing yang sebelumnya telah diinput di microsoft excel. Data yang telah berhasil ditransformasi dapat dilihat seperti gambar 6. Untuk mentransformasi data dapat dilakukan dengan mengklik Import Configuration Wizard pada bagian Parameters kemudian mengimport data training untuk operator Read Excel 1 dan data testing pada operator Read Excel 2.
- Langkah 4 : Pada proses import data, atribut yang nilainya masih kosong akan di cleaning. Serta yang nilainya yang yang masih kosong dapat digantikan dengan men-drag and drop Replace Missing Values pada operator Missing. Berdasarkan data keluarga yang diteliti, atribut yang akan dihilangkan ialah Nama KP, dan Status Kepemilikan Rumah. Serta atribut Jumlah Penghasilan Perbulan dan Jumlah Pengeluaran Perbulan akan digantikan dengan atribut “Pendapatan”.

- Langkah 5 : Langkah selanjutnya yaitu menetapkan atribut “Kategori” sebagai variabel keputusan. Sedangkan untuk variabel penentu diantaranya : Pendapatan, Penerima Raskin, Penerima blt/blsm, dan Peserta BPJS.
- Langkah 6 : Data yang sudah berhasil di import kemudian akan diimplementasikan ke dalam decision tree C4.5 untuk mendapatkan klasifikasi tingkat kesejahteraan keluarga. Tahap selanjutnya, men-drag and drop Set Role, Decision Tree, beserta Apply Model pada menu operators ke menu Process. Pada parameters Set Role menetapkan “Kategori” menjadi variabel keputusan. Tahap terakhir pada proses ini yaitu menghubungkan setiap operator yang ada pada menu process.
- Langkah 7 : Setelah operator yang ada di dalam menu process terhubung dan tidak error pada koneksi operator kemudian klik tombol Run pada toolbox untuk memperoleh hasil klasifikasi decision tree C4.5.

POHON KEPUTUSAN (DECISION TREE)

Hasil yang dicapai dari penelitian ini yaitu sebuah pohon keputusan klasifikasi (decision tree) C4.5. Tampak seperti pada gambar 4.



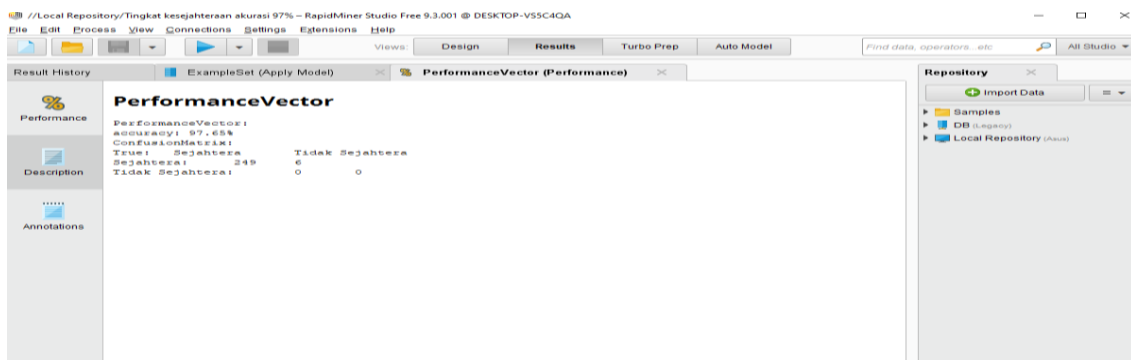
Gambar 4. Pohon Keputusan (*Decision Tree*)

Berdasarkan Gambar 4, hasil klasifikasi digunakan untuk membantu dalam hal penyaluran subsidi pemerintah tepat sasaran pada Desa Tiga Dolok. Dari gambar 4.8. diatas dapat disimpulkan bahwa Pendapatan memiliki kelas tertinggi. Gambar 10 diatas menghasilkan aturan atau rule yang di dapatkan dari RapidMiner menu Decision Tree seperti Tabel 3.

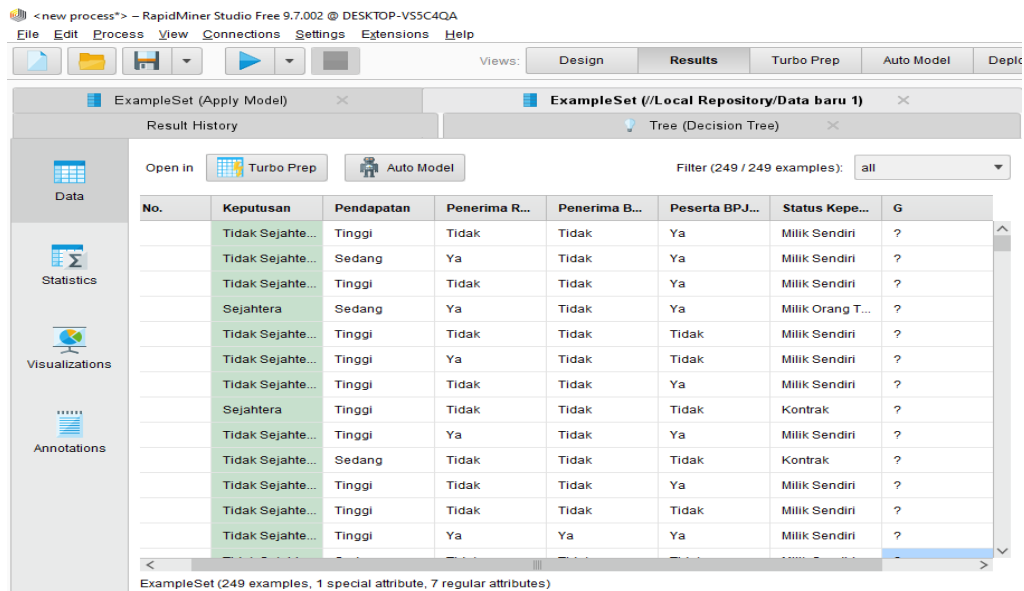
Table 3. Aturan atau Rule pada RapidMiner

Rule	Ket Rule	Keputusan
1	Jika Pendapatan = Rendah, Penerima Raskin = Ya, Penerima BLT/BLSM = Ya, Peserta BPJS = Tidak	Tidak Sejahtera
2	Jika Pendapatan = Sedang, Penerima Raskin = Tidak, Penerima BLT/BLSM = Tidak, Peserta BPJS = Tidak	Sejahtera
3	Jika Pendapatan = Sedang, Penerima Raskin = Tidak, Penerima BLT/BLSM = Tidak, Peserta BPJS = Ya	Tidak Sejahtera
4	Jika Pendapatan = Sedang, Penerima Raskin = Tidak, Penerima BLT/BLSM = Ya	Tidak Sejahtera
5	Jika Pendapatan = Sedang, Penerima Raskin = Ya, Penerima BLT/BLSM = Ya, Peserta BPJS = Ya	Tidak Sejahtera
6	Jika Pendapatan = Tinggi, Penerima BLT/BLSM = Tidak, Peserta BPJS = Tidak, Penerima Raskin = Tidak	Sejahtera

Berdasarkan Tabel 3 aturan atau rule diatas diperoleh untuk keputusan Sejahtera sebanyak 2 dan keputusan tidak sejahtera sebanyak 4 rule. Hasil *performance decision tree* dapat dilihat seperti pada Gambar 5.



Gambar 5. Performance Decision Tree



Gambar 7. Contoh Hasil Prediksi Dari Proses Yang Dijalankan

Selanjutnya dilakukan penerapan model yang digunakan untuk mengetahui model yang telah dilatih sebelumnya menggunakan data training yang tujuannya untuk mendapatkan prediksi. Jika proses dijalankan maka akan di dapatkan hasil seperti Gambar 6. Berdasarkan Gambar 13 terlihat model tersebut memberikan hasil prediksi yang hampir benar.

Metode pohon keputusan mempunyai beberapa kelebihan, diantaranya sebagai berikut :

1. Dapat menghilangkan bahkan mengganti atribut yang kita inginkan.
2. Dapat mengubah keputusan dari yang besar menjadi lebih spesifik.
3. Hasil desain keputusan yang dihasilkan sangat menarik.
4. Kualitas keputusan yang dihasilkan sangat fleksibel.

KESIMPULAN

Penyaluran subsidi pemerintah terhadap masyarakat seringkali tidak tepat sasaran, banyak keluarga yang seharusnya mendapat bantuan tidak memperoleh bantuan sedangkan yang ekonomi cukup mendapat bantuan. Masalah tersebut dapat diatasi dengan adanya pohon keputusan terhadap kriteria yang sudah ditetapkan. Analisa dan perhitungan yang tepat terhadap atribut kesejahteraan yang ada dapat membantu proses penentuan keputusan. Dengan penggunaan pohon keputusan (decision tree) memberikan gambaran secara jelas prioritas atribut mana yang digunakan sebagai penentu keputusan. Pohon keputusan (decision tree) berupa tingkatan atribut dari kategori kesejahteraan yang ada.

DAFTAR PUSTAKA

- [1] A. D. Lilik and H. Hanny, "Metode pohon keputusan menggunakan algoritma c4.5 untuk pengelompokan data penduduk pada tingkatan kesejahteraan keluarga 1,2," pp. 1–8, 2016.
- [2] S. H. Waluyo and Prihandoko, "Klasifikasi Pemanfaat Program Beras Sejahtera (RASTRA) Berdasarkan Tingkat Kemiskinan Dengan Menggunakan Algoritma Decision Tree C4 . 5 Berbasis Particle Swarm Optimization," *J. Ilm. Ilmu-Ilmu Tek.*, vol. 7, no. 2, pp. 19–24, 2017.
- [3] I. Akil, "Implementasi Algoritma C4.5 Pada Sistem Evaluasi Perijinan Pembukaan Program Studi," vol. 19, no. 1, pp. 80–85, 2017.
- [4] D. A. N. A. C and Kasus, "Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner Dan Algoritma C4.5 (Studi Kasus Upt Puskesmas Ponjong I, Gunungkidul)," *Stud. Puskesmas, U P T*, vol. 5, no. 2, pp. 299–309, 2016.
- [5] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, "Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, pp. 87–96, 2018.
- [6] A. R. Febie Elfaladonna, "Analisa Metode Classification-Decission Tree Dan Algoritma," *Sci. Inf. Technol.*, vol. 2, no. 1, pp. 10–17, 2019.