

PENERAPAN METODE C4.5 UNTUK KLASIFIKASI WARGA MISKIN PADA DESA MENGANDUNG SARI

Carissa Elma Purnomo¹, Rikendry²

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia¹

Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia²

carissaelmapuromo@gmail.com¹, rikendry@teknokrat.ac.id²

Received: (9 September 2021) Accepted: (16 September 2021) Published: (30 September 2021)

Abstract

Mengandung Sari Village, Sekampung Udik District, East Lampung Regency is a village derived from the name "Mehganung wo Gannung", a division of Toba village. On January 1, 1961 by the head of the village of Tuba named Abdul Gani Title Kerio Pendetta Mego, along with the head of the Minak Brajo hamlet, they started to pioneer or open this area with the village name "Mengandung sari". Mengandung Sari Village is one of the villages located in Sekampung Udik, East Lampung. This village has 6 hamlets in it. Mengandung Sari Village is obliged to participate in programs from the Ministry of Social Affairs for the welfare of the population including Non-Cash Food Assistance (BPNT) in the form of groceries which is given once a month, and others. This program is intended for poor families as part of the government's efforts to reduce poverty. A new problem arose, namely how the Local Environmental Unit (SLS) determined the Head of the Family who was truly entitled to receive the assistance. With the C4.5 algorithm, data on villagers containing extracts can be classified based on "poor" and "not poor" and the results of the accuracy of classifying data for poor people using the C4.5 method are categorized as "superior" with an accuracy of 100%. Based on the results of the previous discussion, it can be concluded that the Kelurahan can provide a decision tree for assessing data on the poor using the C4.5 algorithm method.

Keywords: C4.5, Classification, Poor People, Containing Sari, decision tree.

Abstrak

Desa Mengandung Sari, Kecamatan Sekampung Udik, Kabupaten Lampung Timur merupakan desa yang berasal dari nama "Mehganung wo Gannung" pemekaran dari desa Toba. Pada tanggal 01 Januari 1961 oleh kepala kampung Tuba yang bernama Abdul Gani Gelar Kerio Pendetta Mego, beserta kepala dusun Minak Brajo, mereka mulai merintis atau membuka wilayah ini dengan nama desa "Mengandung sari". Desa Mengandung Sari adalah salah satu desa yang berada di Kecamatan Sekampung Udik Lampung Timur. Desa ini memiliki 6 dusun didalamnya. Desa Mengandung Sari diwajibkan untuk mengikuti dalam program dari Kemeterian Sosial untuk mensejahterakan penduduk diantaranya Bantuan Pangan Non Tunai (BPNT) berupa sembako yang diberikan sebulan sekali, dan lain-lain. Program ini diperuntukan bagi keluarga miskin sebagai upaya pemerintah dalam menanggulangi kemiskinan. Masalah baru muncul yaitu bagaimana Satuan Lingkungan Setempat (SLS) menetapkan Kepala Keluarga yang benar-benar berhak menerima bantuan tersebut. Dengan adanya algoritma C4.5, data warga desa mengandung sari dapat di klasifikasikan berdasarkan "miskin" dan "tidak miskin" dan hasil akurasi pengklasifikasian data warga miskin menggunakan metode C4.5 dikategorikan "unggul" dengan hasil akurasi sebesar 100%. Berdasarkan hasil pembahasan sebelumnya, dapat disimpulkan bahwa pihak Kelurahan dapat memberikan keputusan atau decision tree penilaian data warga miskin menggunakan metode algoritma C4.5.

Kata Kunci: C4.5, Klasifikasi, Warga Miskin, Mengandung Sari, decision tree.

To cite this article:

Purnomo, Rikendry. (2021). Penerapan Metode C4.5 Untuk Klasifikasi Warga Miskin Pada Desa Mengandung Sari. Jurnal Teknologi dan Sistem Informasi, Vol (2) No.3 , 14 - 25

1. Pendahuluan

Indonesia merupakan salah satu Negara berkembang di Asia khususnya Asia Tenggara. Salah satu masalah yang sering dihadapi oleh negara berkembang adalah Kemiskinan. Bagi pemerintah Indonesia masalah kemiskinan merupakan masalah lama yang sulit untuk diselesaikan. Secara teoritis, telah dilakukan upaya dalam pengentasan kemiskinan pada suatu daerah dengan upaya pertumbuhan ekonomi yang berkualitas dengan beberapa program kerja pemerintah untuk meningkatkan pertumbuhan ekonomi yang berkualitas. Pengelompokan warga miskin sangat diperlukan karena sangat membantu Kepala Desa dalam pemilihan warga miskin. Dalam beberapa literatur ada beberapa teknik untuk menentukan warga miskin salah satunya adalah data mining.

Desa Mengandung Sari adalah salah satu desa yang berada di Kecamatan Sekampung Udik Lampung Timur. Desa ini memiliki 6 dusun didalamnya. Desa Mengandung Sari diwajibkan untuk mengikuti dalam program dari Kementerian Sosial untuk mensejahterakan penduduk diantaranya Program Keluarga Harapan (PKH) berupa uang yang diberikan 3 bulan sekali dan Bantuan Pangan Non Tunai (BPNT) berupa sembako yang diberikan sebulan sekali, dan lain-lain. Program ini diperuntukan bagi keluarga miskin sebagai upaya pemerintah dalam menanggulangi kemiskinan. Masalah baru muncul yaitu bagaimana Satuan Lingkungan Setempat (SLS) menetapkan Kepala Keluarga yang benar-benar berhak menerima bantuan tersebut. Hal ini biasanya hanya dilakukan secara manual berdasarkan data yang sudah ada dilapangan dan terkadang sering menimbulkan kecemburuan sosial, karena yang menerima bantuan masih ada unsur-unsur kedekatan kekeluargaan dengan SLS setempat.

Untuk menanggulangi hal diatas maka dikembangkan suatu solusi melalui sebuah penelitian menggunakan Algoritma C4.5 karena algoritma C4.5 memiliki kelebihan yaitu dapat mengolah data diskrit dan data numerik (kontinyu) [1] [2] [3]. Dalam penelitian ini Algoritma C4.5 untuk melakukan penggalian informasi (informasi disini dimasukkan untuk mencari warga yang benar-benar berhak menerima bantuan). Penggalian informasi berdasarkan kumpulan data dari 6 Dusun yaitu Dusun 1 Sampai 6, yang akan dilakukan dengan menggunakan teknologi Data Mining.

Data Mining dapat diartikan sebagai proses mengekstrak atau menggali knowledge yang ada pada sekumpulan data [4] [5] [6]. Sedangkan Algoritma C.45 untuk mengetahui decision tree penerima bantuan uang atau beras setiap bulan pada Desa Mengandung Sari. Sehingga decision tree yang didapatkan akan digunakan untuk memperbaiki kebijakan -kebijakan yang selama ini dianggap tidak adil karena faktor kedekatan kekeluargaan,

untuk memperbaiki kebijakan tersebut peneliti mengangkat penelitian dengan judul “Penerapan Metode C4.5 Untuk Klasifikasi Warga Miskin Pada Desa Mengandung Sari Menggunakan Parameter Badan Pusat Statistik (BPS)”.

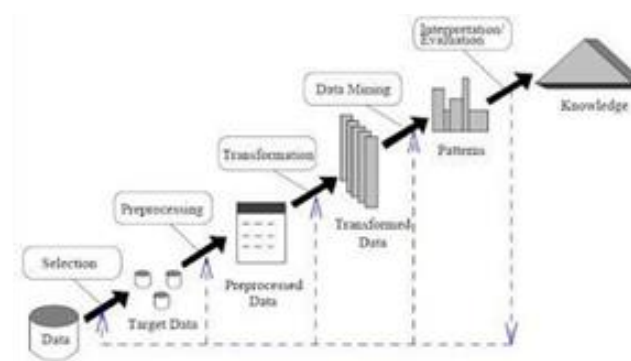
2. Tinjauan Pustaka

2.1 Data Mining

Menurut [7] [8], secara sederhana, data mining dapat diartikan sebagai proses mengekstrak atau menggali knowledge yang ada pada sekumpulan data. Informasi dan knowledge yang didapat tersebut dapat digunakan pada banyak bidang, seperti manajemen bisnis, pendidikan, kesehatan dan sebagainya.

2.2 Knowledge Discovery In Databases (KDD)

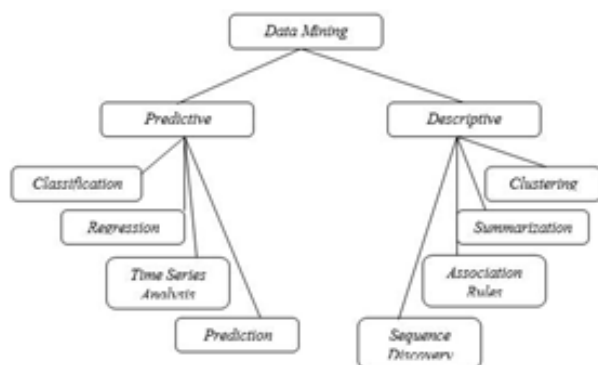
Knowledge Discovery In Databases (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data [9] [10] [11]. Berikut merupakan Tahapan dari KDD:



Gambar 1. Tahapan KDD

2.3 Teknik-Teknik Data mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan basis data [12] [13] [14].



Gambar 2. Teknik Data Mining

2.4 Program Bantuan Pangan Non Tunai (BPNT)

Menurut Sukoco (2016), Program Bantuan Pangan Non Tunai merupakan upaya mereformasi Program Subsidi Rastra yang dilaksanakan berdasarkan arahan Presiden Republik Indonesia untuk meningkatkan Program Bantuan Pangan Non Tunai (BPNT) efektifitas dan ketepatan sasaran program, serta mendorong inklusi keuangan. Penyaluran Bantuan Pangan secara Non Tunai dilaksanakan secara bertahap mulai tahun 2017 pada beberapa daerah terpilih di Indonesia dengan akses dan fasilitas memadai. Selain untuk memberikan pilihan pangan yang lebih luas, penyaluran Bantuan Pangan secara Non Tunai melalui sistem perbankan juga dimaksudkan untuk mendukung perilaku produktif masyarakat melalui fleksibilitas waktu penarikan bantuan dan akumulasi asset melalui kesempatan menabung [15].

2.5 Kriteria Penerima Program Bantuan Pangan Non Tunai (BPNT)

Berikut Syarat atau kriteria calon penerima Bantuan Pangan Non Tunai :

1. Luas lantai bangunan tempat tinggal kurang dari 8 meter persegi untuk masing masing anggota keluarga.
2. Jenis lantai bangunan tempat tinggal terbuat dari tanah, bambu, kayu berkualitas rendah.
3. Jenis dinding bangunan tempat tinggal terbuat dari bambu, rumbia, kayu berkualitas rendah.
4. Fasilitas jamban tidak ada, atau ada tetapi dimiliki secara bersama-sama dengan keluarga lain.
5. Sumber air untuk minum atau memasak berasal dari sumur atau mata air tak terlindung, air sungai, danau, atau air hujan.
6. Sumber penerangan di rumah bukan listrik
7. Bahan bakar yang digunakan memasak berasal dari kayu bakar, arang, atau minyak tanah
8. Dalam seminggu tidak pernah mengonsumsi daging, susu, atau hanya sekali dalam seminggu

9. Dalam setahun paling tidak hanya mampu membeli pakaian baru satu stel
10. Makan dalam sehari hanya satu kali atau dua kali.
11. Tidak mampu membayar anggota keluarga berobat ke puskesmas atau poliklinik.
12. Pekerjaan utama kepala rumah tangga adalah petani dengan luas lahan setengah hektare, buruh tani, kuli bangunan, tukang batu, tukang becak, pemulung, atau pekerja informal lainnya dengan pendapatan maksimal Rp 600 ribu per bulan.
13. Pendidikan tertinggi yang ditamatkan kepala rumah tangga bersangkutan tidak lebih dari SD.
14. Tidak memiliki harta senilai Rp 500 ribu seperti tabungan, perhiasan emas, TV berwarna, ternak, sepeda motor (kredit atau non- kredit), kapal motor, tanah, atau barang modal lainnya.

2.6 Pengertian Penduduk

Menurut Purba (2015), Penduduk adalah orang yang menjadi dirinya pribadi maupun menjadi anggota keluarga, warga negara maupun anggota masyarakat yang memiliki tempat tinggal di suatu tempat di wilayah negara tertentu dan juga pada waktu tertentu.

2.7 Pengeritan Miskin

Menurut Purba (2015), miskin sebagai suatu standar tingkat hidup yang rendah, yaitu adanya suatu tingkat kekurangan pada sejumlah atau segolongan orang dibandingkan dengan standar kehidupan yang rendah ini secara langsung nampak pengaruhnya terhadap tingkat keadaan kesehatan, kehidupan moral dan rasa harga diri mereka yang tergolong sebagai orang miskin.

2.8 Kriteria Pengukuran Kemiskinan

Berikut ini merupakan kriteria pengukuran kemiskinan menurut Isdijoso (2016):

Kriteria Pengukuran Kemiskinan BPS pada PSE05

Variabel Kemiskinan Dalam menentukan rumah tangga miskin, BPS menggunakan 14 variabel untuk menentukan apakah suatu rumah tangga layak dikategorikan miskin. Keempat belas variabel tersebut adalah:

1. Luas bangunan;
2. Jenis lantai;
3. Jenis dinding;
4. Fasilitas buang air besar;
5. Sumber air minum;
6. Sumber penerangan;
7. Jenis bahan bakar untuk memasak;
8. Frekuensi membeli daging, ayam, dan susu dalam seminggu;
9. Frekuensi makan dalam sehari;
10. Jumlah stel pakaian baru yang dibeli dalam setahun;

11. Akses ke puskesmas/poliklinik;
12. Akses ke lapangan pekerjaan;
13. Pendidikan terakhir kepala rumah tangga; dan
14. Kepemilikan beberapa aset.

Dalam PSE05, sebuah rumah tangga dikatakan miskin apabila:

- a. Luas lantai bangunan tempat tinggalnya kurang dari 8 m² per orang;
- b. Lantai bangunan tempat tinggalnya terbuat dari tanah/bambu/kayu murahan
- c. dinding bangunan tempat tinggalnya terbuat dari bambu/rumbia/kayu berkualitas rendah Atau tembok tanpa diplester;
- d. Tidak memiliki fasilitas buang air besar/bersama-sama rumah tangga lain menggunakan Satu jamban
- e. Sumber penerangan rumah tangga tidak menggunakan listrik
- f. Air minum berasal dari sumur/mata air yang tidak terlindung/sungai/air hujan
- g. Bahan bakar untuk memasak sehari-hari adalah kayu bakar/arang/minyak tanah
- h. Hanya mengonsumsi daging/susu/ayam satu kali dalam seminggu
- i. Hanya membeli satu stel pakaian baru dalam setahun
- j. Hanya mampu makan satu/dua kali dalam sehari
- k. Tidak sanggup membayar biaya pengobatan di puskesmas/poliklinik
- l. Sumber penghasilan kepala rumah tangga adalah: petani dengan luas lahan 0,5 ha, buruh Tani, nelayan, buruh bangunan, buruh perkebunan, atau pekerjaan lainnya dengan Pendapatan di bawah rp600.000 per bulan
- m. Pendidikan terakhir kepala rumah tangga: tidak sekolah/tidak tamat sekolah dasar (sd)/hanya sd; dan
- n. Tidak memiliki tabungan/barang yang mudah dijual dengan nilai minimal rp500.000 Seperti sepeda motor (kredit/nonkredit), emas, hewan ternak, kapal motor ataupun Barang modal lainnya.

2.9 Sensus Kemiskinan Pemerintah Daerah

Beberapa pemerintah daerah di Indonesia, seperti Pemerintah Provinsi DKI Jakarta dan Pemerintah Provinsi Kalimantan Timur, pernah melakukan sensus kesejahteraan/kemiskinan. Di DKI Jakarta, variabel yang digunakan untuk menentukan rumah tangga miskin adalah tujuh indikator, yaitu (1) luas lantai rumah per kapita kurang dari 8 meter persegi; (2) lantai rumah berupa tanah atau bambu yang rusak; (3) tidak memiliki fasilitas air bersih; (4) tidak memiliki jamban/WC; (5) konsumsi lauk-pauknya tidak bervariasi; (6) tidak mampu membeli satu stel pakaian dalam setahun untuk setiap anggota rumah tangga; dan (7) tidak memiliki aset rumah tangga yang produktif. Sebuah rumah tangga dikatakan miskin apabila memenuhi tiga dari tujuh variabel tersebut. Indikator atau variabel yang digunakan BPS Provinsi DKI Jakarta ini

dikembangkan dari variabel-variabel yang ditentukan oleh BPS Pusat.

2.10 Algoritma C4.5

Menurut Kusri dan Luthfi (2009), algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (*kontinyu*) dan *diskret*, dapat menangani nilai atribut yang hilang, menghasilkan aturan - aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain.

Ada beberapa tahap untuk membuat pohon keputusan dengan algoritma C.45 yaitu

1. Menyiapkan data training.

Data training di ambil dari data atau dokument yang tersimpan dan tersusun dalam kelas-kelas tertentu.

2. Menentukan akar pohon.

Akar dipilih dari atribut yang dipilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai tertinggi akan menjadi akar.

Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy dengan rumus:

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

Keterangan :

- S = himpunan kasus

- N = Jumlah partisi

- Pi = Proposi Si terhadap S

3. Menghitung nilai gain menggunakan rumus.

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

Keterangan :

S = Himpunan kasus

A = Fitur

N = Jumlah partisi atribut A

|S_i| = proposi Si terhadap S

|S| = Jumlah kasus dalam S

4. Ulangi langkah ke dua hingga semua record terpartisi

5. Proses partisi berhenti jika :

- a. Semua record dalam simpul N mendapat nilai kelas yang sama.
- b. Tidak ada atribut dalam record yang dipartisi lagi.
- c. Tidak ada record di dalam cabang yang kosong.

2.11 WEKA (Waikato Environment for Knowledge Anaysis)

Menurut Witten dan Frank (2011), WEKA adalah sebuah paket *toolsmachine learning* praktis. "WEKA" merupakan singkatan dari *Waikato Environment for Knowledge Analysis*, yang dibuat di Universitas Waikato,

New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mampu menyelesaikan masalah-masalah data mining di dunia-nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan machine learning.

3. Metode Penelitian

3.1 Kerangka Penelitian



Gambar 3. Kerangka Penelitian



Gambar 4. Tahapan Penelitian

3.2 Metode Pengumpulan Data

Pada tahapan ini peneliti memperoleh data sebagai bahan penyusunan skripsi dengan menggunakan beberapa metode dalam melakukan penelitian yaitu sebagai berikut: Metode Pengamatan. Metode ini dilakukan dengan cara pengumpulan data melalui proses pengamatan secara langsung dalam melakukan pendataan warga pada Desa. Mengandung Sari, Berikut ini contoh data warga sebelum dilakukan seleksi atribut yang ditunjukkan pada Tabel 1.

Tabel 1. Data Warga Sebelum Seleksi Atribut

No	Nama	Jenis Kelamin	Agama	Pendidikan	Pekerjaan	Alamat	Telepon	Alamat Email
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Data Mining

Tahap ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan. Metode yang digunakan pada penelitian ini adalah metode C4.5. Ada beberapa tahapan dalam membuat pohon keputusan dengan Algoritma C4.5 yaitu:

1. Menyiapkan Data Training
Berikut adalah data training yang digunakan untuk mendapatkan hasil perhitungan data mining untuk mendapatkan pola atau informasi yang terdapat dalam kumpulan data, dapat dilihat pada tabel 2:

Tabel 2. Data Trining yang Digunakan

No	Alamat	Agama	Sex	Profesi	Income	Education	Age	Marital	Children	Household Income	Household Size	Household Type	Household Income	Household Size	Household Type	Household Income	Household Size	Household Type	Household Income	Household Size	Household Type	
1

- Menentukan Akar Pohon
Sebelum menghitung gain dari atribut, maka hitung terlebih dahulu nilai entropy. Untuk menghitung nilai entropy dengan rumus:

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

Keterangan :

- S = himpunan kasus
- N = Jumlah partisi
- Pi = Proposi Si terhadap S

Dibawah ini perhitungan entrophy data warga:
Entropy(S) = ((-Jumlah Status Miskin/All Total) * IMLOG2(Jumlah Status Miskin/All Total) + (-Jumlah Status Tidak Miskin/All Total) * IMLOG2(Jumlah Status Tidak Miskin /All Total)).

Setelah menghitung entrophy maka akan dilanjutkan untuk perhitungan gain, dibawah ini merupakan rumus yang digunakan untuk perhitungan gain.

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

Keterangan :

- S = Himpunan kasus
- A = Fitur
- N = Jumlah partisi atribut A
- |Si| = Proposi Si terhadap S
- |S| = Jumlah kasus dalam S

- Proses pembagian data dengan k-fold cross validation.

Pembagian data training dan data testing pada dataset warga miskin menggunakan metode 10-fold cross validation, dengan artian dataset akan dibagi menjadi 10 bagian dari 500 data. Satu bagian (data testing) digunakan untuk pengujian dan sisanya digunakan untuk data training. Berikut adalah penjelasannya:

Tabel 3. Pembagian Data Testing

ID Data Testing	Jumlah Data
K-1	50 Data
K-2	50 Data
K-3	50 Data
K-4	50 Data
K-5	50 Data
K-6	50 Data
K-7	50 Data
K-8	50 Data
K-9	50 Data
K-10	50 Data

Tabel diatas merupakan pembagian data testing warga miskin sebanyak 10 kali dengan keseluruhan data sebanyak 500 data. Setelah data testing dibagi kedalam kelompok sebanyak 10 selanjutnya adalah menentukan pembagian data training bisa didapatkan dari sisa pembagian data testing.

Tabel 4. Pembagian Data Training

ID Data Training	Jumlah Data
DTR-1	450 Data
DTR-2	450 Data
DTR-3	450 Data
DTR-4	450 Data
DTR-5	450 Data
DTR-6	450 Data
DTR-7	450 Data
DTR-8	450 Data
DTR-9	450 Data
DTR-10	450 Data

4. Training

Merupakan penentuan data yang akan dijadikan input untuk pengujian menggunakan metode C4.5. Data yang akan dijadikan input adalah data testing yang telah dibagi sama rata. Data tersebut nantinya akan diuji berdasarkan data training yang sudah dibagi sama rata juga, sehingga menghasilkan skema training sebagai berikut:

Tabel 4. Skema Training

Tabel tersebut merupakan skema training dengan menggunakan 10-fold cross validation, data-data tersebut akan segera bergantian digunakan sebagai training dan testing.

5. Klasifikasi

Tahap ini merupakan tahapan klasifikasi atau pengujian data testing terhadap data training yang sudah dibagi menggunakan metode fold cross validation dengan perhitungan C4.5, pada tahap ini juga akan diketahui akurasi dari masing-masing pengujian. Klasifikasi didasarkan pada skema training yang sudah ditentukan.

Tabel 5. Pengujian Berdasarkan Fold Cross Validation

ID Pengujian	ID Data Training	ID Data Testing
Pgn-1	DTR-1	K-1
Pgn-2	DTR-2	K-2
Pgn-3	DTR-3	K-3
Pgn-4	DTR-4	K-4
Pgn-5	DTR-5	K-5
Pgn-6	DTR-6	K-6
Pgn-7	DTR-7	K-7
Pgn-8	DTR-8	K-8
Pgn-9	DTR-9	K-9
Pgn-10	DTR-10	K-10

6. Training

Pada tahap ini dilakukan perhitungan rata-rata akurasi berdasarkan keseluruhan pengujian. Ini dilakukan untuk mengetahui seberapa besar akurasi dari keseluruhan data, sebagai acuan seberapa baik metode C4.5 dalam mengklasifikasikan warga miskin. Tingkat akurasi menunjukkan tingkat kebenaran klasifikasi terhadap kelas. Semakin rendah akurasi berpengaruh terhadap tingkat kebenarannya sebaliknya untuk tingkat akurasi yang tinggi semakin rendah tingkat kesalahannya. Tingkat akurasi yang baik adalah tingkat presentase yang mendekati angka 100%.

4. Hasil Dan Pembahasan

4.1 Pembahasan

Pada sub bab ini akan membahas mengenai hasil analisis terhadap penggunaan algoritma C4.5 dalam memprediksi tingkat keberhasilan dalam membantu Kepala Desa untuk mendukung pemberian keputusan apakah warga tersebut miskin atau tidak secara objektif dalam bentuk pohon keputusan, jika warga tersebut miskin maka layak untuk mendapatkan Bantuan Pangan Non Tunai (BPNT).

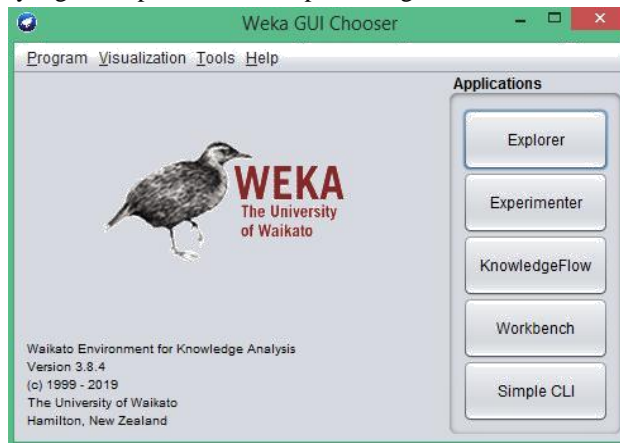
4.2 Implementasi Data Mining Pada Weka Pada sub bab ini akan dijelaskan hasil

pengujian dari aplikasi weka. Serta akan diberikan hasil tingkat akurasi pengujian data menggunakan K-Fold Cross Validation sebanyak 10 fold.

4.2.1. Visualisasi C4.5 pada Weka

2 Tampilan Awal Weka

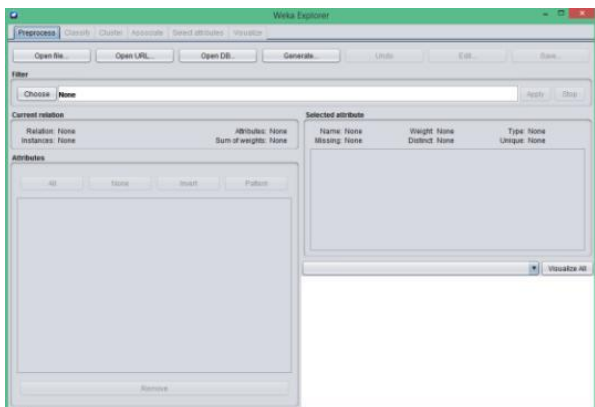
Berikut ini merupakan tampilan awal WEKA versi 3.8.4 yang dapat dilihat pada gambar 5.



Gambar 5. Tampilan Awal Weka

2 Tampilan Awal Explorer

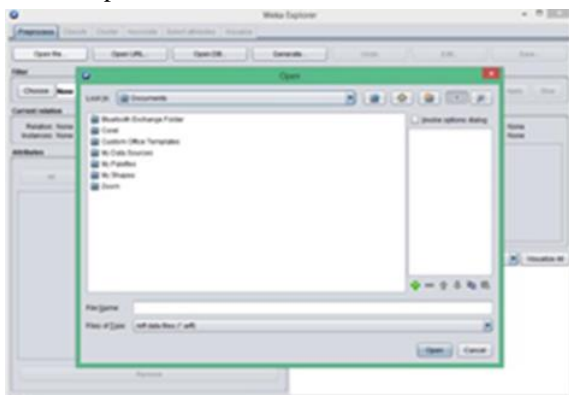
Berikut ini merupakan tampilan awal explorer sebelum diinputkan data, tampilan explorer dapat dilihat pada gambar 6.



Gambar 6. Tampilan Awal Explorer

3. Penginputan Data

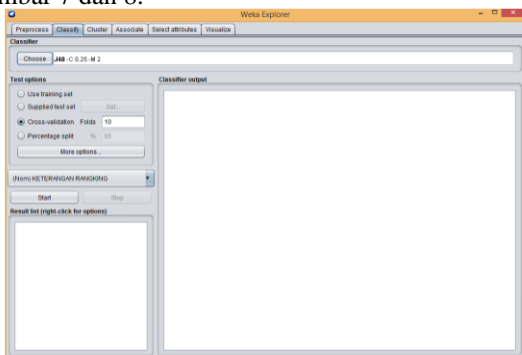
Berikut ini merupakan tampilan awal explorer setelah diinputkan data.



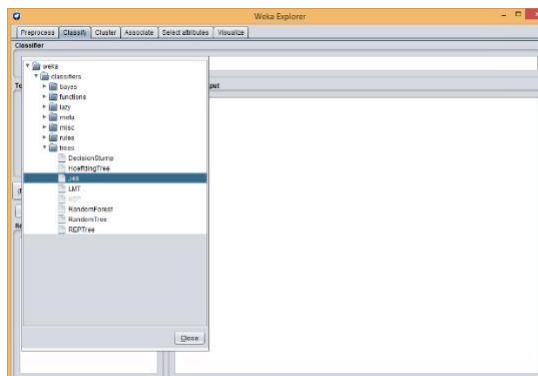
Gambar 6. Penginputan Data

4. Proses Klasifikasi

Proses klasifikasi dilakukan melalui tab *Classify*. Berikut ini merupakan tampilan menu classify dan pemilihan *choose* algoritma yang dapat dilihat pada gambar 7 dan 8.



Gambar 7. Menu Classify



Gambar 8. Pemilihan Algoritma J4.8 atau C4.5

4.2.2 Proses Pengujian

Proses pengujian dilakukan berdasarkan perhitungan data testing k-fold cross validation dari fold 1 sampai dengan fold 10.

1. Perhitungan Pgn-1 atau Fold 1

Tabel 6. Pengujian Data Testing 50 Data Fold 1

Instance	Actual Class	Predicted Class	Confidence	Weight	Correctly Classified	Incorrectly Classified	Percentage Correctly Classified	Percentage Incorrectly Classified	Percentage Correctly Classified (Weighted)	Percentage Incorrectly Classified (Weighted)
1	0	0	1.0	1.0	1	0	100%	0%	100%	0%
2	0	0	1.0	1.0	1	0	100%	0%	100%	0%
3	0	0	1.0	1.0	1	0	100%	0%	100%	0%
4	0	0	1.0	1.0	1	0	100%	0%	100%	0%
5	0	0	1.0	1.0	1	0	100%	0%	100%	0%
6	0	0	1.0	1.0	1	0	100%	0%	100%	0%
7	0	0	1.0	1.0	1	0	100%	0%	100%	0%
8	0	0	1.0	1.0	1	0	100%	0%	100%	0%
9	0	0	1.0	1.0	1	0	100%	0%	100%	0%
10	0	0	1.0	1.0	1	0	100%	0%	100%	0%
11	0	0	1.0	1.0	1	0	100%	0%	100%	0%
12	0	0	1.0	1.0	1	0	100%	0%	100%	0%
13	0	0	1.0	1.0	1	0	100%	0%	100%	0%
14	0	0	1.0	1.0	1	0	100%	0%	100%	0%
15	0	0	1.0	1.0	1	0	100%	0%	100%	0%
16	0	0	1.0	1.0	1	0	100%	0%	100%	0%
17	0	0	1.0	1.0	1	0	100%	0%	100%	0%
18	0	0	1.0	1.0	1	0	100%	0%	100%	0%
19	0	0	1.0	1.0	1	0	100%	0%	100%	0%
20	0	0	1.0	1.0	1	0	100%	0%	100%	0%
21	0	0	1.0	1.0	1	0	100%	0%	100%	0%
22	0	0	1.0	1.0	1	0	100%	0%	100%	0%
23	0	0	1.0	1.0	1	0	100%	0%	100%	0%
24	0	0	1.0	1.0	1	0	100%	0%	100%	0%
25	0	0	1.0	1.0	1	0	100%	0%	100%	0%
26	0	0	1.0	1.0	1	0	100%	0%	100%	0%
27	0	0	1.0	1.0	1	0	100%	0%	100%	0%
28	0	0	1.0	1.0	1	0	100%	0%	100%	0%
29	0	0	1.0	1.0	1	0	100%	0%	100%	0%
30	0	0	1.0	1.0	1	0	100%	0%	100%	0%
31	0	0	1.0	1.0	1	0	100%	0%	100%	0%
32	0	0	1.0	1.0	1	0	100%	0%	100%	0%
33	0	0	1.0	1.0	1	0	100%	0%	100%	0%
34	0	0	1.0	1.0	1	0	100%	0%	100%	0%
35	0	0	1.0	1.0	1	0	100%	0%	100%	0%
36	0	0	1.0	1.0	1	0	100%	0%	100%	0%
37	0	0	1.0	1.0	1	0	100%	0%	100%	0%
38	0	0	1.0	1.0	1	0	100%	0%	100%	0%
39	0	0	1.0	1.0	1	0	100%	0%	100%	0%
40	0	0	1.0	1.0	1	0	100%	0%	100%	0%
41	0	0	1.0	1.0	1	0	100%	0%	100%	0%
42	0	0	1.0	1.0	1	0	100%	0%	100%	0%
43	0	0	1.0	1.0	1	0	100%	0%	100%	0%
44	0	0	1.0	1.0	1	0	100%	0%	100%	0%
45	0	0	1.0	1.0	1	0	100%	0%	100%	0%
46	0	0	1.0	1.0	1	0	100%	0%	100%	0%
47	0	0	1.0	1.0	1	0	100%	0%	100%	0%
48	0	0	1.0	1.0	1	0	100%	0%	100%	0%
49	0	0	1.0	1.0	1	0	100%	0%	100%	0%
50	0	0	1.0	1.0	1	0	100%	0%	100%	0%

Berdasar tabel 6 di atas menunjukkan bahwa nilai total pengujian yang benar mendapatkan hasil 47 data dari 50 data.

4.3 Evaluasi Metode Klasifikasi Dengan Confusion Matrix

Confusion matrix digunakan untuk mendeskripsikan evaluasi dari hasil prediksi.

Beberapa metode evaluasi dapat diperoleh dari confusion matrix seperti akurasi, error rate, sensitivity (recall), specificity, precision, dan false positive rate. Tabel 11 berikut menunjukkan gambaran confusion matrix.

Tabel 11 Gambar Confusion Matrix

		Nilai yang Sebenarnya	
		TP (True Positive)	FP (False Positive)
Nilai Prediksi	FALSE	FN (False Negative)	TN (True Negative)
	TRUE		

Keterangan Tabel 11:

1. TP (true positive)
Hasil prediksi sesuai dengan nilai sebenarnya yaitu sama sama benar.
2. TN (true negative)
Hasil prediksi benar, tetapi nilai yang sebenarnya salah.
3. FP (false positive)
Hasil prediksi salah, tetapi hasil yang sebenarnya benar.
4. FN (false negative)
Hasil prediksi salah, tetapi hasil yang sebenarnya juga salah.

4.3.1 Hasil Confusion Matrix

Berikut ini adalah ringkasan dari confusion matrix tiap pengujian yang memiliki nilai berbeda-beda.

Tabel 12. Hasil Confusion Matrix

Pengujian	Akurasi	Error Rate	Sensitivity (Recall)/True Positive Rate	False Positive Rate (FPR)	Precision	Specificity (true Negative Rate)
Pengujian 1	94%	6%	100%	9%	85%	90%
Pengujian 2	90%	10%	100%	20%	73%	86%
Pengujian 3	90%	0%	100%	100%	20%	80%
Pengujian 4	100%	0%	100%	0%	100%	100%
Pengujian 5	82%	18%	100%	45%	76%	55%
Pengujian 6	96%	4%	100%	66%	95%	33%
Pengujian 7	82%	18%	100%	52%	78%	47%
Pengujian 8	100%	0%	100%	0%	100%	100%
Pengujian 9	88%	12%	100%	66%	87%	33%
Pengujian 10	94%	6%	100%	50%	93%	50%

4.3.2 Hasil Pengujian

Dari beberapa pengujian di atas, data yang diambil untuk menjadi acuan dalam menentukan data warga miskin adalah nilai accuracy dari hasil pengujian. Berikut ini adalah ringkasan dari tingkat akurasi tiap pengujian yang memiliki nilai berbeda-beda.

Tabel 13. Tingkat Akurasi Pengujian

Pengujian	Akurasi Testing
Pengujian 1	94%
Pengujian 2	90%
Pengujian 3	90%
Pengujian 4	100%
Pengujian 5	82%
Pengujian 6	96%
Pengujian 7	82%
Pengujian 8	100%
Pengujian 9	88%
Pengujian 10	94%

Berdasar tabel 13 menunjukkan bahwa dari beberapa tahap pengujian terhadap performa klasifikasi metode C4.5 dalam memprediksi data warga miskin dengan kombinasi pemilihan data training dan data testing berdasarkan skema k-fold cross validation memiliki tingkat akurasi yang beragam yaitu:

1. Pengujian 1 & 2 menghasilkan nilai akurasi sebesar 94% & 90%, dengan hasil 47 data diprediksi benar dan 3 data diprediksi salah. Selain itu kesalahan data tersebut dikarenakan ada data yang diprediksi "tidak miskin" pada data sebenarnya "miskin" dilihat dari jenis dinding tembok dengan plaster.
2. Pengujian 3 menghasilkan nilai akurasi sebesar 90%, dengan hasil 45 data diprediksi benar dan 5 data diprediksi salah. Kesalahan data pada data testing yaitu diprediksi "Miskin" padahal pada data sebenarnya "tidak miskin" karena dinilai dari jenis dinding yaitu tembok dengan plaster, penghasilan < 5.000.000 dan pendidikan terakhir SD dan SMP.
3. Pengujian 4 menghasilkan nilai akurasi sebesar 100%, karna data yang dihasilkan dari data testing dengan data sebenarnya diprediksi benar
4. Pengujian 5 menghasilkan nilai akurasi sebesar 82%, dengan hasil 41 data diprediksi benar dan 9 data diprediksi salah. Kesalahan pada data testing yaitu prediksi "Miskin" padahal data sebenarnya "tidak miskin" karena dinilai dari jenisdindingtembokdengan plaster,penghasilan <5.000.000 dan kepemilikan aset motor,emas senilai 10gr, dan hewan ternak.
5. Pengujian 6 menghasilkan nilai akurasi sebesar 98 % dengan hasil 48 data diprediksi benar dan 2 data diprediksi salah. Kesalah pada data

- testing yaitu nilai sebenarnya tidak sama dengan hasil prediksi berdasarkan pendidikan terakhir, jenis dinding.
6. Pengujian 7 menghasilkan nilai akurasi sebesar 82%, dengan hasil 41 data diprediksi benar dan 9 data diprediksi salah. Kesalahan pada data testing yaitu nilai sebenarnya tidak sama dengan hasil prediksi berdasarkan jenis dinding dan pendidikan terakhir.
 7. Pengujian 8 menghasilkan nilai akurasi sebesar 100%, karna data yang dihasilkan dari data testing dengan data sebenarnya diprediksi benar.
 8. Pengujian 9 menghasilkan nilai akurasi sebesar 88%, dengan hasil 44 data diprediksi benar dan 6 data diprediksi salah. Kesalahan pada data testing yaitu nilai sebenarnya tidak sama dengan hasil prediksi berdasarkan jenis dinding dan kepemilikan aset.
 9. Pengujian 10 menghasilkan nilai akurasi sebesar dengan hasil 47 data diprediksi benar dan 3 data diprediksi salah. Selain itu kesalahan data tersebut dikarenakan pada data testing tidak sama dengan data sebenarnya berdasar jenis dinding dan pendidikan terakhir kepala keluarga.

4.3.3. Hasil Rata-Rata Akurasi

Untuk mengetahui rata-rata dari beberapa hasil pengujian dalam hal ini sebagai titik acuan untuk mengetahui keberhasilan dari metode C4.5 dalam memprediksi data warga miskin maka didapatkan perhitungannya sebagai berikut:

$$\text{Rata-rata akurasi} = \frac{\sum \text{Jumlah Hasil Akurasi Setiap Pengujian}}{\sum \text{Jumlah Total Pengujian}}$$

$$\text{Rata-rata akurasi} = \frac{94 + 90 + 90 + 100 + 82 + 96 + 82 + 100 + 88 + 94}{10}$$

Berdasarkan data dari pengujian dan setelah dihitung rata-rata akurasi dari setiap pengujian didapatkan hasil akhir akurasi sebesar 91,6%. Ini mendefinisikan dalam penelitian ini, bahwa metode C4.5 termasuk dalam performa yang unggul.

5. Kesimpulan Dan Saran

5.1 Kesimpulan

Berdasarkan pembahasan dari penulisan laporan skripsi tentang penerapan metode C4.5 untuk klasifikasi warga miskin pada desa mengandung sari dapat diambil kesimpulan yaitu:

1. Dengan adanya algoritma C4.5, data warga desa mengandung sari dapat di klasifikasikan

berdasarkan “miskin” dan “tidak miskin” dan hasil akurasi pengklasifikasian data warga miskin menggunakan metode C4.5 dikategorikan “unggul” dengan hasil rata-rata akurasi sebesar 91,6%.

2. Berdasarkan hasil pembahasan sebelumnya, dapat disimpulkan bahwa pihak kelurahan dapat memberikan keputusan atau *decision tree* penilaian data warga miskin menggunakan metode algoritma C4.5.

5.2 Saran

Berdasarkan kesimpulan dari hasil penelitian yang telah diuraikan diatas, maka saran yang dapat diberikan untuk penerapan metode C4.5 untuk klasifikasi warga miskin pada desa mengandung sari yaitu pada penelitian ini, peneliti hanya menggunakan satu metode prediksi, sehingga diharapkan pada penelitian selanjutnya dapat melakukan komparasi dengan metode prediksi yang lain untuk pengujian menggunakan data warga desa mengandung sari.

DAFTAR PUSTAKA

- [1] A. A. Aldino and H. Sulistiani, “Decision Tree C4. 5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia),” *Eduitic-Scientific J. Informatics Educ.*, vol. 7, no. 1, 2020.
- [2] L. N. Rani, “Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit,” *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
- [3] A. F. O. Pasaribu, “ANALISIS POLA MENGGUNAKAN METODE C4. 5 UNTUK PEMINATAN JURUSAN SISWA BERDASARKAN KURIKULUM (studi kasus: SMAN 1 NATAR),” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021.
- [4] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, “ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.
- [5] M. Bakri, “Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means,” *vol.*, vol. 11, pp. 1–4, 2017.
- [6] S. Alim, P. P. Lestari, and R. Rusliyawati, “Sistem Pakar Diagnosa Penyakit Tanaman Kakao Menggunakan Metode Certainty Factor Pada

- Kelompok Tani Pt Olam Indonesia (Cocoa) Cabang Lampung,” *J. Data Min. Dan Sist. Inf.*, vol. 1, no. 1, pp. 26–31, 2020.
- [7] A. Mittal and S. Patidar, “Sentiment analysis on twitter data: A survey,” *ACM Int. Conf. Proceeding Ser.*, pp. 91–95, 2019, doi: 10.1145/3348445.3348466.
- [8] D. Ariyanti and K. Iswardani, “Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes,” *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [9] R. Napianto, Y. Rahmanto, and R. I. B. D. O. Lestari, “Software Development Sistem Pakar Penyakit Kanker Pada Rongga Mulut Berbasis Web,” 2019.
- [10] M. R. Handoko and N. Neneng, “SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021.
- [11] A. Nurkholis, A. Riyantomo, and M. Tafrikan, “Sistem pakar penyakit lambung menggunakan metode forward chaining,” *J. Ilm. MOMENTUM*, vol. 13, no. 1, 2017.
- [12] H. Sulistiani, K. Muludi, and A. S. Admi Syarif, “Peer Review: Implementation of Various Artificial Intelligence Approach for Prediction and Recommendation of Personality Disorder Patient.”
- [13] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [14] N. Kurniati, Y. Yanitasari, D. A. Lantana, I. S. Karima, and E. R. Susanto, “Sistem Pakar Untuk Mendiagnosa Penyakit Kulit Pada Kucing Menggunakan Certainty Factor,” *Ilk. J. Ilm.*, vol. 9, no. 1, pp. 34–41, 2017.
- [15] E. R. Susanto, A. S. Puspaningrum, and N. Neneng, “Model Rekomendasi Penerima Bantuan Sosial Berdasarkan Data Kesejahteraan Rakyat,” *J. Tekno Kompak*, vol. 15, no. 1, pp. 1–12, 2021.