

KLASIFIKASI RUMAH TANGGA MISKIN DESA SINEY KECAMATAN TINOMBO SELATAN MENGGUNAKAN METODE *ORDINAL CLASS CLASSIFIER* (OCC)

Alfiani¹, Agusman Sahari², dan Abdul Mahatir Najar³

^{1,2,3}Program Studi Matematika Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tadulako

Jalan Soekarno – Hatta Km. 09 Tondo, Palu 94118, Indonesia

¹alfianihalide92@gmail.com, ²mukhlisahagus79@gmail.com, ³mahatirnajar@gmail.com

ABSTRACT

The Central Bureau of Statistics (BPS) categorizes the status of poor households into three, category that are Very Poor Households (RTSM), Poor Households (RTM), and Nearly Poor Households (RTHM). This classified poor households in Siney Village using the Ordinal Class Classifier (OCC) method . OCC is a meta classifier that can generate a predictive class and can assume nominal class values. This study used data available at the Siney village office of 300 data with 14 indicators of poverty. The test results show that OCC can classify poor household status well. This can be seen from the results of the highest precision test of 98.10%, 87.50% recall and 96.6% OCC accuracy.

Keywords : Poverty, Classification, Ordinal Class Classifier

ABSTRAK

Badan Pusat Statistik (BPS) mengkategorikan status rumah tangga miskin menjadi tiga yaitu Rumah Tangga Sangat Miskin (RTSM), Rumah Tangga Miskin (RTM), dan Rumah Tangga Hampir Miskin (RTHM). Mengklasifikasikan golongan rumah tangga miskin desa Siney menggunakan metode Ordinal Class Classifier (OCC). OCC merupakan meta classifier yang dapat menghasilkan suatu kelas prediksi dan dapat mengasumsikan nilai kelas nominal. Penelitian ini menggunakan data yang tersedia di kantor desa Siney sebanyak 300 data dengan 14 indikator kemiskinan. Tujuannya dapat mengklasifikasikan dan memperoleh akurasi yang optimal dari pembagian data testing dan data training dalam klasifikasi RTSM, RTM, dan RTHM berdasarkan metode OCC. Berdasarkan hasil analisis baik program maupun manual menunjukkan bahwa OCC dapat melakukan klasifikasi status rumah tangga miskin dengan baik. Hal ini dapat dilihat dari hasil Pengujian tertinggi precision sebesar 98,10%, recall 87,50% dan akurasi 96,6% OCC.

Kata Kunci : Kemiskinan, Klasifikasi, Ordinal Class Classifier

I. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan bagi negara berkembang termasuk Indonesia. 14% masyarakat dikategorikan miskin dari 240 juta jiwa rakyat Indonesia dengan menggunakan indikator berpendapatan 1\$ (Rp. 14.358,-) per hari dianggap miskin. Selanjutnya, jika kita menggunakan indikator bank dunia dimana masyarakat miskin di Indonesia adalah yang berpendapatan kurang dari 2\$ (Rp. 28.716,-) per hari, maka angka tersebut melonjak menjadi 35% (Pratama, 2014).

Data mining merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, *association rule*, *clustering*, deskripsi dan visualisasi (Turban, 2007).

Menurut (Setyawan dan Aripin, 2016) Teknik klasifikasi digunakan untuk proses pengolahan data. Klasifikasi merupakan proses yang digunakan untuk mencari model atau fungsi yang dapat menjelaskan dan membedakan suatu kelas atau konsep dari data, dengan tujuan untuk menggunakan model dan melakukan prediksi dari kelas suatu objek dimana tidak diketahui label dari kelas tersebut. Model yang ada berasal dari analisis pengumpulan training data (objek data dimana kelas label diketahui).

Desa Sinei terletak dikecamatan Tinombo Selatan kabupaten Parigi Moutong. Desa Sinei merupakan salah satu desa yang memiliki potensi alam yang luar biasa. Beras merupakan hasil pertanian yang selalu menjadi kebanggaan rakyat desa, selain itu desa Sinei juga merupakan daerah yang subur, letaknya yang strategis, berada dekat dengan garis equator menjadikan daerah ini termasuk beriklim stabil.

Berdasarkan data dari sekretaris kantor desa Siney hingga tahun 2021 terdapat 300 kepala keluarga yang masih tergolong miskin dari 700 kepala keluarga yang ada di desa Siney. Status rumah tangga miskin dikategorikan menjadi tiga yaitu rumah tangga sangat miskin (RTSM), rumah tangga miskin (RTM) dan rumah tangga hampir miskin (RTHM) (BPS, 2011). Berdasarkan penelitian terdahulu yang dilakukan, diketahui banyak faktor-faktor yang diduga dapat mempengaruhi tingkat kemiskinan rumah tangga miskin antara lain terkait kualitas kesehatan, kualitas ekonomi, SDM, dan fasilitas rumah tangga seperti luas lantai, luas kavling bangunan, dan sumber air minum (Faried Effendy, 2018).

Selama masa pandemik pemerintah banyak memberikan bantuan kepada masyarakat yang terdampak covid-19, salah satunya adalah bantuan sosial tunai (BST). BST merupakan keputusan yang dikeluarkan oleh Menteri Sosial Republik Indonesia Nomor 54/HUK/2020 tentang pelaksanaan bantuan sosial sembako dan bantuan sosial tunai dalam penanganan dampak Covid-19. Adapun syarat dari penerima BST adalah masyarakat yang masuk kedalam pendataan RT/RW, kehilangan

mata pencaharian selama pandemi Covid-19, tidak terdaftar dalam program bantuan sosial lainnya seperti PKH, BPNT, kartu sembako dan kartu prakerja (Nunik Dewi Pramanik, 2020).

Salah satu metode yang dapat digunakan untuk melakukan klasifikasi adalah *Ordinal Class Classifier* (OCC). OCC merupakan *meta classifier* yang bisa bekerja di atas setiap model yang dapat menghasilkan suatu kelas prediksi dan dapat mengasumsikan nilai kelas nominal. Jenis klasifikasi OCC ini dapat mengatasi atribut yang bersifat nominal, numerik, dan *ordinal*. Dalam melakukan proses klasifikasi OCC menggunakan algoritma C4.5 sebagai model dasar perhitungannya dan hasilnya nanti akan berupa aturan-aturan yang mudah diinterpretasikan (Faried Effendy, 2018). Berdasarkan pendahuluan di atas maka penelitian ini akan menggunakan OCC.

OCC merupakan algoritma klasifikasi data dan prediksi dengan teknik pohon keputusan yang terkenal dan banyak digunakan karena memiliki kelebihan-kelebihan sebagai berikut [1] Dapat mengolah data numerik, nominal, dan ordinal [2] Dapat menangani nilai atribut yang hilang [3] Menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang menggunakan memori utama dikomputer (Kantardzic, 2011).

II. METODE PENELITIAN

Pada penelitian ini adalah data sekunder yang diperoleh dari pemerintah desa Siney kecamatan Tinombo Selatan kabupaten Parigi Moutong. Adapun jenis data yang digunakan adalah data kuantitatif yang berupa data luas lantai, jenis atap rumah, jenis dinding, jenis lantai, fasilitas BAB, sumber penerangan utama, sumber air minum, bahan bakar memasak, ijazah, status lahan tempat tinggal, luas lahan tempat tinggal, fasilitas MCK, tempat pembuangan limbah, dan pendapatan kepala keluarga. Penelitian ini menggunakan metode OCC, adapun langkah-langkah dalam menyelesaikan penelitian ini yaitu [1] memulai penelitian dengan mengidentifikasi masalah atau mencari permasalahan ditempat yang ingin diteliti [2] studi literatur dengan mengumpulkan beberapa buku, artikel dan jurnal [3] mengidentifikasi masalah [4] pengumpulan data [5] transformasi data [6] pembagian data *training* dan data *testing* [7] pengolahan data menggunakan metode OCC [8] menyimpulkan hasil [9] selesai.

Adapun menurut Faried Effendy, 2018 ada beberapa tahapan menggunakan algoritma OCC adalah sebagai berikut :

- a. Mengelompokkan data training dan data testing, kemudian hitung masing-masing nilai atribut pada data training kemudian dilanjutkan dengan menghitung nilai Entropy dari masing-masing nilai atribut menggunakan persamaan (1) :

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

Keterangan :

S : himpunan kasus

n : Jumlah Partisi S

pi : Proporsi Dari Si Terhadap S

- b. Menghitung nilai gain menggunakan nilai entropy yang telah dihitung sebelumnya. Menghitung gain menggunakan persamaan (2) :

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

Keterangan :

S : himpunan kasus

A : Atribut

n : jumlah partisi atribut A

|Si| : Jumlah Kasus Pada Partisi Ke-i

|S| : Jumlah Kasus Dalam S

- c. Menghitung nilai Split Info dari tiap atribut menggunakan persamaan (3) :

$$SplitInfo(S, A) = \sum_{i=1}^c - \frac{|S_i|}{|S|} \times \log_2 \frac{|S_i|}{|S|}$$

Keterangan :

SS : himpunan kasus

A : atribut

Si : jumlah sampel untuk atribut A

- d. Menghitung gain ratio menggunakan nilai gain dan split info menggunakan persamaan (4):

$$GainRatio(S, A) = \frac{gain(S, A)}{SplitInfo(S, A)}$$

Keterangan :

S : himpunan kasus

A : atribut

Gain(S,A) : Information Gain Pada Atribut A

SplitInfo(S,A) : Split Information Pada Atribut A

- e. Selanjutnya mengambil nilai gain ratio terbesar untuk dijadikan simpul akar
- f. Atribut yang sudah terpilih kemudian ulangi perhitungan *entropy*, *information gain*, *split info* dan *gain ratio* dengan memilih gain ratio terbesar.
- g. Perhitungannya dilakukan sampai membentuk pohon kelasnya masing-masing.
- h. Setelah perhitungannya selesai maka akan membentuk sebuah pohon keputusan dan generate aturan keputusan awal.

III. HASIL DAN PEMBAHASAN

3.1. Transformasi Data

Sebelum menerapkan algoritma OCC, data ditransformasikan terlebih dahulu. Transformasi data yaitu dengan mengubah skala pengukuran data asli menjadi bentuk lain sehingga data tersebut dapat mudah dikenali polanya.

Tabel 1 : Transformasi Data

No	Indikator	Tranformasi Data
1	Pendapatan Kepala Keluarga	< 1000000 = Rendah
		1000000-2000000 = Sedang
		> 2000000 = Tinggi
2	Luas Lahan Tempat Tinggal	15-150 m ²
		151-286 m ²
		287-422 m ²
		423-558 m ²
		559-694 m ²
		695-830 m ²
		831-966 m ²
		967-1102 m ²
		1103-1238 m ²
		1239-1238 m ²
3	Luas Lantai	6-99 m ²
		100-193 m ²
		194-287 m ²
		288-381 m ²
		382-475 m ²
		476-569 m ²
		570-663 m ²
		664-757 m ²
		758-851 m ²
		852-945 ²

3.2. Pembobotan Indikator

Berikut hasil pembobotan dari desa Siney kecamatan tinombo selatan, pembobotan yang dilakukan tersebut sesuai dengan pembobotan yang dilakukan oleh desa Siney pada Tabel 2.

Tabel 2 : Pembobotan Indikator Desa Siney

Urutan	Nama Indikator	Bobot
W1	Pendapatan Kepala Keluarga	0,1
W2	Status Lahan Tempat Tinggal	0,095
W3	Luas Lahan Tempat Tinggal	0,091
W4	Luas Lantai	0,087

W5	Jenis Atap Rumah	0,082
W6	Jenis Dinding	0,078
W7	Jenis Lantai	0,074
W8	Ijazah	0,069
W9	Fasilitas MCK	0,065
W10	Fasilitas BAB	0,06
W11	Sumber Penerangan Utama	0,056
W12	Sumber Air Minum	0,052
W13	Bahan Bakar Memasak	0,047
W14	Tempat Pembuangan Limbah cair	0,043

3.3. Pemberian Skor Pada Atribut

Setiap indikator memiliki variabel dimana skor diberikan untuk setiap variabel. Skor variabel didasarkan pada besar kecilnya pengaruh masing-masing variabel terhadap kemiskinan. Skor untuk setiap variabel berkisar antara 1 dan 0, dimana skor untuk jawaban 1 menunjukkan tidak miskin dan skor 0 menunjukkan tidak miskin (Faried Effendy, 2018).

Berikut adalah perhitungan indeks rumah tangga miskin (IRTM) untuk menentukan klasifikasi RTSM, RTM dan RTHM. Proses penentuan IRTMnya dengan cara mengalikan nilai bobot dan skor sehingga dapat ditentukan kategori dari RTSM, RTM dan RTHM. Berikut perhitungannya :

Untuk indikator pendapatan kepala keluarga (X_1) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,1 \times 1 = 0,1$$

Untuk indikator status lahan tempat tinggal (X_2) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,095 \times 0 = 0$$

Untuk indikator luas lahan tempat tinggal (X_3) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,091 \times 1 = 0,091$$

Untuk indikator luas lantai (X_4) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,087 \times 1 = 0,087$$

Untuk indikator jenis atap (X_5) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,082 \times 1 = 0,082$$

Untuk indikator jenis dinding (X_6) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,078 \times 1 = 0,078$$

Untuk indikator jenis lantai (X_7) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,074 \times 1 = 0,074$$

Untuk indikator ijazah (X8) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,069 \times 1 = 0,069$$

Untuk indikator fasilitas mck (X9) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,065 \times 1 = 0,065$$

Untuk indikator fasilitas bab (X10) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,06 \times 1 = 0,06$$

Untuk indikator sumber penerangan utama (X11) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,056 \times 1 = 0,056$$

Untuk indikator sumber air minum (X12) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,052 \times 0 = 0$$

Untuk indikator bahan bakar memasak (X13) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,047 \times 1 = 0,047$$

Untuk indikator tempat pembuangan limbah cair (X14) kolom pertama

$$I_{RTM} = \sum_i W_i X_i \\ = 0,043 \times 1 = 0,043 \\ I_{RTM} = 0,10 + 0 + 0,09 + 0,09 + 0,08 + 0,08 + 0,07 + 0,07 + 0,07 + 0,06 + 0,06 + 0 + 0,05 \\ + 0,04 = 0,9$$

Maka I_{RTM} yang diperoleh dari baris pertama untuk X1 hingga X14 dengan total sebesar 0,9 yang termasuk dalam kategori RTSM.

3.4. Pelatihan OCC

Pada penerapan *decision tree* dengan menggunakan algoritma OCC, peneliti ingin melihat rumah tangga yang telah dikategorikan dalam kategori RTSM, RTM, dan RTHM dengan melihat dari atribut yang sudah ditentukan sebelumnya. Adapun yang menjadi konsep pohon keputusan tersebut adalah dengan mengubah data yang ada menjadi sebuah pohon keputusan, serta setelah pohon keputusan selesai selanjutnya diubah kedalam bentuk aturan-aturan keputusan (*Rule*).

3.4.1. Membuat Pohon Keputusan Algoritma OCC Secara Manual

Untuk dapat membuat pohon keputusan, pertama-tama harus menghitung entropy dari jumlah kasus terlebih dahulu. Adapun perhitungannya sebagai berikut:

$$\begin{aligned} Entropy (S) &= \sum_{i=1}^n p_i * \log_2 \frac{1}{p_i} = - \sum_{i=1}^n p_i * \log_2 p_i \\ &= \left(-\frac{16}{150} \times \log_2 \frac{16}{150} \right) + \left(-\frac{64}{150} \times \log_2 \frac{64}{150} \right) + \left(-\frac{70}{150} \times \log_2 \frac{70}{150} \right) \\ &= (-0,106 \times -3,228) + (-0,426 \times -1,228) + (-0,466 \times -1,099) \\ &= 0,342 + 0,523 + 0,512 \\ &= 1,377 \approx 1,38 \end{aligned}$$

Didapatkan nilai entropy total adalah 1,38, selanjutnya untuk menghitung gain dari kriteria pendapatan kepala keluarga maka perlu menghitung entropy dari subkriteria pendapatan kepala keluarga. Adapun perhitungannya sebagai berikut :

Entropy X1 untuk subkriteria “rendah < 1.000.000”

$$\begin{aligned} Entropy (S1) &= \sum_{i=1}^n p_i * \log_2 \frac{1}{p_i} = - \sum_{i=1}^n p_i * \log_2 p_i \\ &= \left(-\frac{10}{84} \times \log_2 \frac{10}{84} \right) + \left(-\frac{43}{84} \times \log_2 \frac{43}{84} \right) + \left(-\frac{31}{84} \times \log_2 \frac{31}{84} \right) \\ &= (-0,119 \times -3,070) + (-0,511 \times -0,966) + (-0,369 \times -1,438) \\ &= 0,365 + 0,493 + 0,530 = 1,388 \approx 1,39 \end{aligned}$$

Entropy X1 untuk subkriteria “sedang 1.000.000 – 2.000.000”

$$\begin{aligned} Entropy (S2) &= \sum_{i=1}^n p_i * \log_2 \frac{1}{p_i} = - \sum_{i=1}^n p_i * \log_2 p_i \\ &= \left(-\frac{6}{41} \times \log_2 \frac{6}{41} \right) + \left(-\frac{15}{41} \times \log_2 \frac{15}{41} \right) + \left(-\frac{20}{41} \times \log_2 \frac{20}{41} \right) \\ &= (-0,146 \times -2,772) + (-0,365 \times -1,450) + (-0,487 \times -1,035) \\ &= 0,404 + 0,529 + 0,504 \\ &= 1,437 \approx 1,44 \end{aligned}$$

Entropy X1 untuk subkriteria “tinggi > 2.000.000”

$$\begin{aligned} Entropy (S2) &= \sum_{i=1}^n p_i * \log_2 \frac{1}{p_i} = - \sum_{i=1}^n p_i * \log_2 p_i \\ &= \left(-\frac{0}{25} \times \log_2 \frac{0}{25} \right) + \left(-\frac{6}{25} \times \log_2 \frac{6}{25} \right) + \left(-\frac{19}{25} \times \log_2 \frac{19}{25} \right) \\ &= 0 \end{aligned}$$

Setelah semua entropy subkriteria sudah diketahui, maka gain dari pendapatan kepala keluarga sudah bisa dihitung. Adapun perhitungannya sebagai berikut:

$$\begin{aligned} gain (S, X1) &= entropy (S) - \sum_{i \in values(A)} \frac{|S_i|}{|S|} \times entropy (S_i) \\ &= 1,38 - \left(\frac{84}{150} \times 1,39 \right) + \left(\frac{41}{150} \times 1,44 \right) + \left(\frac{25}{150} \times 0 \right) \\ &= 1,38 - (0,7784 + 0,3888 + 0) \\ &= 1,38 - 1,1672 \\ &= 0,2128 \end{aligned}$$

Setelah nilai gain diketahui, hitung nilai splitinfo dan gain ratio dari indikator pendapatan kepala keluarga. Adapun perhitungannya sebagai berikut:

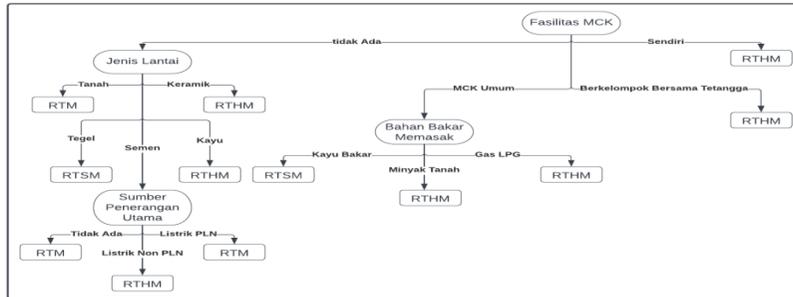
$$\begin{aligned} splitinfo(S, A) &= \sum_{i=1}^c - \frac{|S_i|}{|S|} \times \log_2 \frac{|S_i|}{|S|} \\ &= \frac{(-84) \times \log_2 \left(\frac{84}{150} \right) + (-41) \times \log_2 \left(\frac{41}{150} \right) + (-25) \times \log_2 \left(\frac{25}{150} \right)}{150 \times \log_2 (2)} \\ &= \frac{(-84) \times (-0,25181) + (-41) \times (-0,56330) + (-25) \times (-0,77815)}{150 \times 0,30} \\ &= \frac{21,15204 + 23,0953 + 19,45375}{45} \\ &= \frac{63,70109}{45} = 1,41557 \end{aligned}$$

Perhitungan gain ratio sebagai berikut:

$$\begin{aligned} gainratio(S, A) &= \frac{gain(S, A)}{splitinfo(S, A)} \\ &= \frac{0,2128}{1,41557} = 0,15 \end{aligned}$$

Setelah perhitungan selesai selanjutnya, pilih gainratio tertinggi untuk dijadikan *root node*, kemudian perhitungan diulangi sampai semua atribut memiliki kelasnya masing-masing.

Pohon keputusan akhir yang terbentuk dapat dilihat pada Gambar 1.



Gambar 1 : Node Cabang 3

Karena proses pembentukan pohon keputusan menggunakan algoritma OCC diatas telah selesai, dari Gambar 1 diperoleh aturan (*rule*) untuk mengklasifikasikan rumah tangga miskin di desa Siney kecamatan tinombo selatan. Berikut adalah aturan (*rule*) klasifikasi rumah tangga miskin menggunakan OCC.

1. Jika fasilitas MCK “tidak ada” dan jenis lantai “tanah” maka RTM
2. Jika fasilitas MCK “tidak ada” dan jenis lantai “kayu” maka RTSM
3. Jika fasilitas MCK “tidak ada” dan jenis lantai “semen” dan sumber penerangan utama “tidak ada” maka RTM
4. Jika fasilitas MCK “tidak ada” dan jenis lantai “semen” dan sumber penerangan utama “listrik non PLN” maka RTHM
5. Jika fasilitas MCK “tidak ada” dan jenis lantai “semen” dan sumber penerangan utama “listrik PLN” maka RTM
6. Jika fasilitas MCK “tidak ada” dan jenis lantai “tegel” maka RTHM
7. Jika fasilitas MCK “tidak ada” dan jenis lantai “keramik” maka RTHM
8. Jika fasilitas MCK “MCK umum” dan bahan bakar memasak “kayu bakar” maka RTSM
9. Jika fasilitas MCK “MCK umum” dan bahan bakar memasak “minyak tanah” maka RTHM
10. Jika fasilitas MCK “MCK umum” dan bahan bakar memasak “gas lpg” maka RTHM
11. Jika fasilitas MCK “berkelompok bersama tetangga” maka RTHM
12. Jika fasilitas MCK sendiri maka RTHM

3.4.2. Membuat Pohon Keputusan Algoritma OCC Secara Program

	keterangan	jumlah kasus	RTSM	RTM	RTHM	entropy
TOTAL		150	16	64	70	1.38
PENDAPATAN KEPALA KELUARGA	rendah	84	10	43	31	1.39
PENDAPATAN KEPALA KELUARGA	sedang	41	6	15	20	1.44
PENDAPATAN KEPALA KELUARGA	tinggi	25	0	6	19	0.00
STATUS LAHAN TEMPAT TINGGAL	milik orang lain	23	6	15	2	1.21
STATUS LAHAN TEMPAT TINGGAL	milik sendiri	127	10	49	68	1.39
LUAS LAHAN TEMPAT TINGGAL (m2)	k1	35	7	20	8	1.41
LUAS LAHAN TEMPAT TINGGAL (m2)	k2	51	8	25	18	1.45
LUAS LAHAN TEMPAT TINGGAL (m2)	k3	21	1	8	12	1.20
LUAS LAHAN TEMPAT TINGGAL (m2)	k4	4	0	3	1	0.00
LUAS LAHAN TEMPAT TINGGAL (m2)	k5	17	0	7	10	0.00
LUAS LAHAN TEMPAT TINGGAL (m2)	k6	3	0	0	3	0.00
LUAS LAHAN TEMPAT TINGGAL (m2)	k7	2	0	0	2	0.00
LUAS LAHAN TEMPAT TINGGAL (m2)	k8	1	0	0	1	0.00
LUAS LAHAN TEMPAT TINGGAL (m2)	k9	9	0	1	8	0.00
LUAS LAHAN TEMPAT TINGGAL (m2)	k10	7	0	0	7	0.00
LUAS LANTAI (m2)	k1	39	6	24	9	1.33
LUAS LANTAI (m2)	k2	51	6	22	23	1.40
LUAS LANTAI (m2)	k3	41	4	16	21	1.35
LUAS LANTAI (m2)	k4	1	0	0	1	0.00
LUAS LANTAI (m2)	k5	0	0	0	0	0.00
LUAS LANTAI (m2)	k6	0	0	0	0	0.00
LUAS LANTAI (m2)	k7	0	0	0	0	0.00
LUAS LANTAI (m2)	k8	0	0	0	0	0.00
LUAS LANTAI (m2)	k9	18	0	2	16	0.00
LUAS LANTAI (m2)	k10	0	0	0	0	0.00
JENIS ATAP RUMAH	kayu jerami	3	0	3	0	0.00
JENIS ATAP RUMAH	seng	142	16	59	67	1.39
JENIS ATAP RUMAH	genteng	5	0	2	3	0.00
JENIS DINDING	kayu kualitas rendah	35	12	21	2	1.21
JENIS DINDING	kayu kualitas tinggi	16	2	8	6	1.41
JENIS DINDING	beton	99	2	35	62	1.07
JENIS LANTAI	tanah	8	1	3	4	1.41
JENIS LANTAI	kayu	18	7	8	3	1.48
JENIS LANTAI	semen	98	8	51	39	1.31
JENIS LANTAI	tegel	8	0	0	8	0.00
JENIS LANTAI	keramik	18	0	2	16	0.00
IJAZAH TERAKHIR KEPALA KELUARGA	SD	23	2	11	10	1.34
IJAZAH TERAKHIR KEPALA KELUARGA	SMP	58	9	28	21	1.45
IJAZAH TERAKHIR KEPALA KELUARGA	SMA	65	5	25	35	1.30
IJAZAH TERAKHIR KEPALA KELUARGA	S1	4	0	0	4	0.00
FASILITAS MCK	tidak ada	64	10	36	18	1.40
FASILITAS MCK	MCK umum	9	3	4	2	1.53
FASILITAS MCK	berkelompok bersama tetangga	8	3	5	0	0.00
FASILITAS MCK	sendiri	69	0	19	50	0.00
FASILITAS TEMPAT BUANG AIR BESAR	sungai	19	4	15	0	0.00
FASILITAS TEMPAT BUANG AIR BESAR	jamban umum	19	6	9	4	1.51
FASILITAS TEMPAT BUANG AIR BESAR	jamban bersama tetangga	13	5	8	0	0.00
FASILITAS TEMPAT BUANG AIR BESAR	jamban sendiri	99	1	32	66	0.98
SUMBER PENERANGAN UTAMA	tidak ada	14	7	6	1	1.30
SUMBER PENERANGAN UTAMA	listrik non PLN	6	1	4	1	1.25
SUMBER PENERANGAN UTAMA	listrik PLN	130	8	54	68	1.26
SUMBER AIR MINUM	air isi ulang	13	4	3	6	1.53
SUMBER AIR MINUM	mata air	137	12	61	64	1.34
BAHAN BAKAR UNTUK MEMASAK	kayu bakar	22	7	11	4	1.47
BAHAN BAKAR UNTUK MEMASAK	minyak tanah	2	0	1	1	0.00
BAHAN BAKAR UNTUK MEMASAK	gas LPG	126	9	52	65	1.29
TEMPAT PEMBUANGAN LIMBAH	sawah	78	9	38	31	1.39
TEMPAT PEMBUANGAN LIMBAH	lubang ditanah	41	6	18	17	1.45
TEMPAT PEMBUANGAN LIMBAH	instalasi pengelolaan limbah	31	1	8	22	1.02

Gambar 2 : Data Hasil *Entropy* Untuk Simpul Akar

Seperti terlihat pada Gambar 2, data tersebut merupakan hasil perhitungan awal untuk membangun pohon keputusan dengan python, dimana jumlah kasus adalah 150 data, yang selanjutnya dibagi menjadi 3 klasifikasi yaitu 16 jumlah kasus RTSM, 64 jumlah kasus RTM, dan 70 jumlah kasus RTHM dan nilai entropy yang diperoleh, total 1,38. Selain itu, perhitungan dilanjutkan dengan menghitung jumlah total kasus dan menghitung nilai *entropy*, *gain*, *split info*, dan *gain ratio* sehingga mendapatkan hasil yang ditunjukkan pada Gambar 3.

	kriteria	gain	split	info	gain ratio
0	PENDAPATAN KEPALA KELUARGA	0.21	1.41	0.15	
1	STATUS LAHAN TEMPAT TINGGAL	0.09	0.62	0.15	
2	LUAS LAHAN TEMPAT TINGGAL (m2)	0.39	2.61	0.15	
3	LUAS LANTAI (m2)	0.19	1.96	0.10	
4	JENIS ATAP RUMAH	0.06	0.35	0.17	
5	JENIS DINDING	0.24	1.23	0.20	
6	JENIS LANTAI	0.27	1.59	0.17	
7	IJAZAH TERAKHIR KEPALA KELUARGA	0.05	1.61	0.03	
8	FASILITAS MCK	0.69	1.51	0.46	
9	FASILITAS TEMPAT BUANG AIR BESAR	0.54	1.46	0.37	
10	SUMBER PENERANGAN UTAMA	0.12	0.68	0.18	
11	SUMBER AIR MINUM	0.02	0.43	0.05	
12	BAHAN BAKAR UNTUK MEMASAK	0.08	0.70	0.11	
13	TEMPAT PEMBUANGAN LIMBAH	0.05	1.47	0.03	

Gambar 3 : Data Hasil *Gain ratio* Untuk Simpul Akar

Berikut adalah *output* dari pohon keputusan menggunakan python:

```
FASILITAS MCK --> MCK umum --> BAHAN BAKAR UNTUK MEMASAK --> kayu bakar --> RTSM
FASILITAS MCK --> MCK umum --> BAHAN BAKAR UNTUK MEMASAK --> minyak tanah --> RTHM
FASILITAS MCK --> MCK umum --> BAHAN BAKAR UNTUK MEMASAK --> gas LPG --> RTHM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> semen --> SUMBER PENERANGAN UTAMA --> tidak ada --> RTM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> semen --> SUMBER PENERANGAN UTAMA --> listrik non PLN --> RTHM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> semen --> SUMBER PENERANGAN UTAMA --> listrik PLN --> RTM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> tanah --> RTM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> kayu --> RTSM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> tegel --> RTHM
FASILITAS MCK --> tidak ada --> JENIS LANTAI --> keramik --> RTHM
FASILITAS MCK --> berkelompok bersama tetangga --> RTM
FASILITAS MCK --> sendiri --> RTHM
```

Gambar 4 : Pohon Keputusan Menggunakan Python

3.5. Pengujian Menggunakan OCC

Pada proses pengujian jumlah data yang dilakukan terdapat 5 kelompok yaitu 10%, 20%, 30%, 40%, dan 50%. Selama pengujian ini menginterpretasikan hasil dari algoritma OCC, untuk data pengujian 10% (30) terdapat 29 data diprediksi benar dengan akurasi 96,6%, untuk data uji 20% (60) 56 data diprediksi benar dengan akurasi 93,3%, 30% (90) dari data uji memiliki 81 data yang diprediksi dengan benar dengan akurasi 90%, 40% (120) dari data pengujian terdapat 103 data yang diprediksi benar sehingga diperoleh akurasi sebesar 85,8%, dan pengujian 50% (150) terdapat 122 data yang diprediksi dengan benar sehingga diperoleh akurasi sebesar 81,3%. Berikut adalah hasil *test* yang diperoleh :

Tabel 3 : Hasil Klasifikasi Data *Testing*

Metode OCC	Pengujian (%)				
	10	20	30	40	50
<i>Accuracy</i>	96,60%	93,30%	90%	85,83%	81,30%
<i>Precision</i>	98,10%	80,90%	67,70%	75,30%	68,80%
<i>Recall</i>	87,50%	85,10%	68,10%	82,60%	69,20%

3.6. Evaluasi

Tujuan dari evaluasi adalah untuk menguji apakah algoritma OCC itu sendiri memberikan klasifikasi yang sesuai untuk rumah tangga miskin. Analisis dilakukan dengan

memasukkan data rumah tangga miskin kemudian membandingkannya dengan menghitung selisih antara hasil uji rumah tangga miskin yang dihasilkan sistem dengan data aktual rumah tangga miskin.

Adapun perhitungan akurasi sebagai berikut:

```
Prediksi Benar : 29 data
Prediksi Salah : 1 data
Akurasi : 96.6666666666667 %
Error : 3.333333333333286 %
[[26 0]
 [ 1 3]]
```

Gambar 4.18. Hasil Pengujian data 10%

$$\begin{aligned} \text{Tingkat Akurasi} &= \frac{\text{jumlah prediksi benar}}{\text{jumlah data testing}} \times 100\% \\ &= \frac{29}{30} \times 100\% = 96,6\% \end{aligned}$$

IV. KESIMPULAN

Berdasarkan hasil penelitian diatas, maka dapat disimpulkan bahwa metode *OCC* untuk mengklasifikasikan rumah tangga miskin dapat diimplementasikan dengan baik. Adapun klasifikasi hasil dari RTSM yaitu 105 KK, RTM 175 KK dan RTHM 20 KK keakuratannya diuji dengan mengambil data sebesar 270 data *training* dan 30 data *testing* dari jumlah data sebanyak 300. Pengujiannya dengan menghitung nilai *precision*, *recall*, dan *accuracy*. Berdasarkan pengujian data nilai akurasi yang tertinggi yaitu sebesar 96,6% diperoleh pada pengujian data latih 90% dan data uji 10%.

DAFTAR PUSTAKA

- [1]. Badan Pusat Statistik. (2011). Analisis Data Kemiskinan Berdasarkan Data Pendataan Program Perlindungan Sosial 2011. Diakses dari <http://www.bps.go.id> pada tanggal 29 Desember 2013.
- [2]. Effendy, F., & Purbandini. (2018). Klasifikasi Rumah Tangga Miskin Menggunakan OCC. Jurnal Nasional Teknologi dan Sistem Informasi, Vol. 04(01), 030-036.
- [3]. Pratama, Y. C. (2014). Analisis Faktor-faktor yang Memengaruhi Kemiskinan di Indonesia. UIN Syarif Hidayatullah Jakarta.
- [4]. Kantardzic, M. (2001). Data Mining, Concepts, Models, Methods, and Algorithms. IEEE Press, A John Wiley & Sons, Inc, Publication.
- [5]. Turban, et al., & dkk. (2007). Decision Support System and Intelligent System Seventh Edition. New Delhi: Prentice Hall.
- [6]. Setyawan, & Aripin. (2016). Analisis Algoritma Naive Bayes untuk Mengklasifikasi Penduduk Ke Dalam Status Tahapan Keluarga Sejahtera Berbasis Forward Selection, Skripsi.

- [7]. Pramanik, N. D. (2020). Dampak Bantuan Paket Sembako dan Bantuan Langsung Tunai Terhadap Kelangsungan Hidup Masyarakat Padalarang Pada Masa Pandemi Covid-19. *Intelekva: Jurnal Ekonomi, Sosial & Humaniora*, Vol. 1(2).