



Studi Kasus

Klasifikasi Rumah Tangga Miskin Menggunakan *Ordinal Class Classifier*

Faried Effendy^a, Purbandini^b

^{a,b} Program Studi S1 Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Airlangga Surabaya

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 16 Januari 2018

Revisi Akhir: 23 April 2018

Diterbitkan Online: 01 Mei 2018

KATA KUNCI

Data Mining,
Ordinal Class Classifier (OCC),
Kemiskinan

KORESPONDENSI

Telepon: +62(087854864845)

E-mail: faried-e@fst.unair.ac.id

ABSTRACT

The Central Bureau of Statistics of Indonesia (BPS) classified the target households into three different categories which were very poor households (RSTM), poor households (RTM), and nearly-poor households (RTSM). BPS need some method that can accelerate the classification process to assist the performance of BPS in order to shorten the processing time. The data scale that used in the classification of poor households was ordinal. Generally, calculations of classification using ordinal ascales only can be found in the software WEKA *Ordinal Class Classifier (OCC)* that was one of the existing classification in WEKA. OCC could be resolve to attributes that are nominal, numerical, and ordinal. So in this research, OCC would be using to classify poor households. By comparing the algorithms performance there were several stages that need to be traversed. The first was the data collection stage, the second was the data processing stage and information by using preprocessing, the third was the analysis stage with tools WEKA. The fourth was a test stage by counting the value of *accuracy*, *precision*, and *recall*. The last stage was evaluation by comparing actual data with predictive data of the result of calculating system. From the classification process, it can be concluded that OCC has the highest *accuracy*, *precision*, and *recall* level which is 90% (3803) of training set and 10% (423) of testing set with *accuracy* of 90.5437%, *precision* 0.919, and *recall* 0.905.

1. PENDAHULUAN

Kemiskinan masih menjadi *trending topic* bagi pemerintah untuk ditangani dan selalu menjadi perhatian khusus sebab masalah kemiskinan seringkali dijadikan sebagai bahan sorotan untuk mengevaluasi kinerja pemerintah dalam meningkatkan taraf hidup rakyat [1].

Badan Pusat Statistik (BPS) mengklasifikasikan rumah tangga sasaran menjadi tiga kategori yaitu Rumah Tangga Sangat Miskin (RTSM), Rumah tangga miskin (RTM), dan Rumah Tangga Hampir Miskin (RTHM). Dalam pengklasifikasiannya BPS menggunakan sistem skoring yaitu setiap variabel diberi skor yang diberi bobot dan bobotnya didasarkan kepada besarnya pengaruh dari setiap variabel terhadap kemiskinan [2] Adapun skala data yang digunakan dalam klasifikasi rumah tangga miskin adalah ordinal. Pengertian dari skala ordinal adalah pengukuran yang mana skala yang digunakan disusun secara runtut dari yang rendah sampai yang tinggi [3].

Pada umumnya perhitungan klasifikasi menggunakan skala ordinal hanya ditemukan pada WEKA (Waikato Environment for Knowledge Analysis). WEKA adalah sebuah alat yang digunakan untuk membandingkan beberapa algoritma *machine learning* yang bisa diaplikasikan untuk permasalahan data mining [4]. Ini menunjukkan bahwa konsep klasifikasi ordinal belum

pernah terjadi. Akan tetapi dalam beberapa dekade terakhir ada kemajuan dalam pembelajaran *artificial* dalam konsep ordinal yaitu model *machine learning*, seperti *decision trees*, *neural network*, dan *support vector machines* telah mendukung klasifikasi ordinal [5].

Proses klasifikasi bertujuan untuk membantu kinerja dari BPS agar dapat mempersingkat waktu klasifikasi rumah tangga sasaran dan mengklasifikasi semua set data dengan benar. Akan tetapi tidak dapat dipungkiri bahwa kinerja suatu metode tidak bisa bekerja 100% benar. Salah satu metode yang bisa digunakan untuk klasifikasi rumah tangga sasaran tersebut adalah dengan menggunakan data mining.

Decision tree adalah salah satu metode data mining yang digunakan untuk klasifikasi. Klasifikasi pohon keputusan merupakan teknik klasifikasi yang sederhana yang banyak digunakan. Para peneliti telah mengembangkan berbagai algoritma pohon keputusan selama beberapa periode dengan meningkatkan kinerja dan kemampuan untuk menangani berbagai jenis data yaitu *Chi-squared Automatic Interaction Detector (CHAID)*, *Classification and Regression Tree (CHART)*, *Iterative Dichotomiser 3 (ID3)*, Algoritma C4.5, Algoritma C5.0, Algoritma Hunt's, dan *Ordinal Class Classifier (OCC)*. Algoritma tersebut mempunyai kelebihan masing-masing dan kinerja yang bagus dalam melakukan klasifikasi dibebberapa jenis data. Namun, peneliti saat ini memilih

menggunakan satu algoritma yang dapat menangani jenis data ordinal.

Berdasarkan penelitian yang dilakukan [6] tentang pengklasifikasian ordinal untuk mengevaluasi finansial pada perusahaan asuransi. Penelitian tersebut menggunakan *Ordinal Class Classifier* (OCC) untuk memecahkan masalahnya didapatkan hasil akurasi sebesar 98,7% dalam klasifikasi menggunakan bantuan test *10-fold cross-validation*. Selain itu pada penelitian [5] membandingkan keakuratan lima klasifikasi ordinal dan tiga non-ordinal pada tolak ukur lima belas dataset yang terjadi pada dunia nyata. Hasilnya menunjukkan bahwa klasifikasi ordinal yang diuji tidak memiliki keuntungan statistik yang bermakna dibandingkan pengklasifikasi non-ordinal. Dengan hasil yang didapatkan oleh penelitian tersebut peneliti menginginkan adanya penelitian lebih lanjut untuk mempelajari klasifikasi ordinal yang lebih akurat di masa depan.

OCC merupakan *meta classifier* yang bisa bekerja diatas setiap model yang menghasilkan kelas prediksi dan dapat mengasumsikan nilai kelas nominal. Jenis klasifikasi OCC 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. Sehingga dalam penelitian ini akan mencoba menggunakan OCC untuk mengklasifikasikan rumah tangga miskin.

2. TINJAUAN PUSTAKA

2.1. Teori Kemiskinan

Menurut Badan Pusat Statistik, 2011 mendefinisikan Garis Kemiskinan (GK) sebagai nilai rupiah yang harus dikeluarkan seseorang dalam sebulan agar dapat memenuhi kebutuhan dasar asupan kalori sebesar 2100 kkal/hari per kapita dikenal dengan Garis Kemiskinan Makanan (GKM) ditambah kebutuhan minimum non makanan yang merupakan kebutuhan seseorang yaitu papan, sandang, sekolah, trans-portasi dan kebutuhan individu rumah tangga dasar lainnya yaitu Garis Kemiskinan Non Makanan (GKNM).

Pada akhir tahun 2005 telah dilakukan pendataan khusus rumah tangga miskin dengan menggunakan 14 indikator kemiskinan. Untuk memenuhi kebutuhan berbagai program pelayanan dasar tersebut yang meliputi:

1. Luas lantai rumah
2. Jenis atap rumah terluas
3. Jenis dinding terluas
4. Jenis lantai terluas
5. Fasilitas tempat buang air besar
6. Sumber penerangan utama
7. Sumber air minum
8. Bahan bakar memasak
9. Intensitas konsumsi daging/ susu/ ayam per minggu
10. Intensitas membeli pakaian per tahun
11. Intensitas makan per hari
12. Pengobatan
13. Ijazah terakhir kepala keluarga
14. Kepemilikan Aset

Metode yang digunakan dalam penentuan kategori Rumah tangga miskin dengan menggunakan sistem skoring yaitu setiap variabel diberi skor yang diberi bobot dan bobotnya didasarkan kepada besarnya pengaruh dari setiap variabel terhadap kemiskinan. Jumlah variable dan besarnya bobot berbeda disetiap kabupaten. Dari bobot masing-masing variabel terpilih untuk tiap

kabupaten/kota selanjutnya dihitung indeks skor Rumah tangga miskin dengan formula:

$$I_{RTM} = \sum_i W_i X_i \quad (1)$$

Keterangan:

W_i = bobot variabel terpilih, dan $\sum W_i = 1$

X_i = nilai skor variabel terpilih (skor 1 untuk jawaban yang mengindikasikan miskin dan skor 0 untuk jawaban yang mengindikasikan tidak miskin).

I_{RTM} = indeks Rumah Tangga Miskin, dengan nilai antara 0 dan 1

Berdasarkan nilai I_{RTM} selanjutnya semua Rumah tangga dikategorikan:

1. RTSM ($I_{RTM} = 0,80-1,00$);
2. RTM ($I_{RTM} = 0,60-0,79$);
3. RTHM ($I_{RTM} < 0,60$)

2.2 Data Mining

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 [7]. Menurut [8] Secara garis besar data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu :

1. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.

2. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi. Secara sederhana data mining bisa dikatakan sebagai proses menyaring atau “menambang” pengetahuan dari sejumlah data yang besar. Istilah lain untuk data mining adalah *Knowledge Discovery in Database* (KDD). Walaupun sebenarnya data mining sendiri adalah bagian dari tahapan proses dalam KDD.

2.3 Ordinal Class Classifier

Ordinal Class Classifier merupakan algoritma klasifikasi data dan prediksi dengan teknik pohon keputusan yang terkenal dan banyak digunakan karena memiliki kelebihan-kelebihan yaitu [9]:

- a. Dapat mengolah data numerik, nominal, dan ordinal
- b. Dapat menangani nilai atribut yang hilang
- c. Menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat di antara algoritma-algoritma yang menggunakan memori utama di komputer.

Ada beberapa tahapan dalam membangun sebuah pohon keputusan dengan *Ordinal Class Classifier*, yaitu :

1. Mempersiapkan data training.
2. Menghitung *entropy*
3. Untuk memilih kriteria sebagai akar, didasarkan pada nilai gain tertinggi dari kriteria kriteria yang ada
4. Melakukan perhitungan Split Info dan *Gain ratio*.
5. Ulangi langkah ke-2 hingga semua *record* terpartisi.
6. Maka akan terbentuk pohon keputusan.
7. Pruning

2.4 WEKA

WEKA adalah sebuah paket tools machine 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*. Perangkat lunak ini ditulis dalam hirarki class Java dengan metode berorientasi objek dan dapat berjalan hampir di semua *platform* [10].

3. METODOLOGI

3.1. Pengumpulan Data

Teknik pengumpulan data yang digunakan pada penelitian ini dengan 1. Studi literatur dilakukan dengan cara memahami dan mempelajari lebih mendalam bagaimana pengukuran efektivitas dari *Ordinal Class Classifier* dalam klasifikasi rumah tangga miskin dengan bantuan buku, jurnal, dan sumber lain yang terkait. Data berasal dari jurnal [4].

3.2. Pengelolaan Data dan Informasi

Data yang telah terkumpul selanjutnya diproses melalui tahap pengolahan. Adapun yang termasuk dalam lingkup pengolahan data, yaitu:

1. Berdasarkan dari studi literatur akan digunakan dalam langkah-langkah perhitungan dan pengolahan data masukan untuk proses klasifikasi rumah tangga miskin dengan *Ordinal Class Classifier*.
2. Data yang didapatkan dari jurnal yang berjudul “Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang Berdasarkan Faktor-faktor yang Mempengaruhi dengan Pendekatan CART (*Classification and Regression Trees*)” berjumlah 4735 data yang merupakan data *dummy*.
3. Tahap pra-pemrosesan ada beberapa tahapan yang harus dilakukan dalam penelitian ini, yaitu:
 - a. *Data Cleaning*
Melakukan *data cleaning* atau pembersihan data dengan tujuan untuk menghilangkan data missing value dan menghilangkan atribut yang tidak diperlukan sehingga tidak mengganggu proses selanjutnya.
 - b. *Data Transformation*
Transformasi data bertujuan untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis ragam. Beberapa teknik data mining membutuhkan format data khusus sebelum bisa diaplikasikan. Data yang didapatkan dari jurnal berupa data *dummy* dibutuhkan transformasi data karena klasifikasi hanya menerima input data ordinal. Perubahan data dilakukan secara manual dengan mengurutkan kelas atribut dari yang terkecil sampai dengan yang terbesar.
4. Data rumah tangga miskin akan dibagi menjadi data *training* dan data *testing*. Data akan dikelompokkan menjadi beberapa bagian tujuannya adalah untuk melakukan pengukuran efektivitas berdasarkan jumlah data. Pengelompokan data sebagai berikut:
 1. Data *training* dan data *testing* 50%
 2. Data *training* 60% dan data *testing* 40%
 3. Data *training* 70% dan data *testing* 30%
 4. Data *training* 80% dan data *testing* 20%

5. Data *training* 90% dan data *testing* 10%

3.3. Analisis menggunakan Ordinal Class Classifier

Metode yang digunakan dalam penelitian ini adalah *Ordinal Class Classifier*.

1) Tahapan Ordinal Class Classifier

- a. Memasukkan data *training*. Data *training* dibagi menjadi beberapa kelompok yaitu 50%, 60%, 70%, 80%, dan 90%. Sehingga data *training* yang dimasukan dan dihitung sesuai dengan kelompoknya masing-masing.
- b. Hitung frekuensi kemunculan masing-masing nilai atribut pada data *training*. Kemudian hitung nilai *Entropy* dari masing-masing nilai atribut menggunakan persamaan 2.

$$Entropy(S) = \sum_{i=1}^c - p_i * \log_2 p_i \quad (2)$$

- c. Hitung nilai *Information Gain* menggunakan nilai *Entropy* yang telah dihitung sebelumnya. Perhitungan *information gain* menggunakan persamaan 3.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} * Entropy(S_v) \quad (3)$$

- d. Hitung nilai *Split Info* dari tiap atribut menggunakan persamaan 4.

$$SplitInfo(S, A) = \sum_{i=1}^c - \frac{|S_i|}{|S|} \times \log_2 \frac{|S_i|}{|S|} \quad (4)$$

- e. Hitung nilai *Gain ratio* menggunakan nilai *Information gain* dan *Split Info*. Menghitung *Gain ratio* menggunakan persamaan 5.

$$GainRatio(S, A) = \frac{gain(S, A)}{SplitInfo(S, A)} \quad (5)$$

- f. Ambil nilai *Gain ratio* terbesar dan jadikan simpul akar.
- g. Hilangkan atribut yang dipilih sebelumnya dan ulangi perhitungan nilai *Entropy*, *Information gain*, *Split Info*, dan *Gain ratio* dengan memilih *Gain ratio* terbesar.
- h. Ulangi perhitungan tersebut hingga semua atribut pohon memiliki kelas.
- i. Jika semua pohon sudah memiliki kelas, maka tampilkan pohon keputusan awal dan generate aturan keputusan awal.
- j. Tahap Pruning.
Pruning (pemangkasan) dilakukan saat pohon keputusan selesai dibuat. Dengan menggunakan pendekatan post pruning. Dengan menggunakan Persamaan 6.

$$e = \frac{\left(f + \frac{z^2}{2N} + z \sqrt{\frac{f^2}{N} + \frac{z^2}{4N^2}} \right)}{\left(1 + \frac{z^2}{N} \right)} \quad (6)$$

3.4. Pengujian

Proses pengujian dilakukan menggunakan *contingency table*. Jumlah data yang dilakukan untuk pengujian ada 5 kelompok yaitu 10%, 20%, 30%, 40%, dan 50%.

3.5. Evaluasi

Evaluasi dilakukan dengan memasukan data rumah tangga miskin di Kecamatan Jombang kemudian membandingkan dan menghitung selisih antara hasil data *testing* klasifikasi rumah tangga miskin yang dihasilkan oleh sistem dengan data aktual klasifikasi rumah tangga miskin.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data.

Teknik pengumpulan data yang digunakan pada penelitian ini yaitu sebagai berikut:

1. Studi literatur dilakukan dengan cara memahami dan mempelajari lebih mendalam bagaimana pengukuran kinerja dari *Ordinal Class Classifier* dalam klasifikasi rumah tangga miskin dengan bantuan buku, jurnal, dan sumber lain yang terkait.
2. Data berasal dari rumah tangga miskin di wilayah Kecamatan Jombang, Kabupaten Jombang, Provinsi Jawa Timur periode 2013-2015.

4.2 Pengolahan data dan Informasi

Data yang telah terkumpul selanjutnya diproses melalui tahap pengolahan. Adapun yang termasuk dalam lingkup pengolahan data, yaitu:

1. Tahap pra-pemrosesan ada beberapa tahapan yang harus

dilakukan dalam penelitian ini, yaitu:

a. *Data Cleaning*

Tujuan dari data cleaning adalah menghilangkan nilai-nilai data yang salah, memperbaiki kecacauan data dan memeriksa data yang tidak konsisten. Pada data klasifikasi rumah tangga miskin terdapat banyak sekali data yang *missing value*. Jika data *missing value* tidak dihilangkan maka akan mengganggu proses berikutnya. Adapun beberapa teknik yang digunakan dalam mengatasi *missing value*, yaitu mengabaikan atau menghapus data yang *missing value*, mengisi data yang *missing value* secara manual, menggunakan nilai yang paling mungkin untuk mengisi *missing value* tersebut. Pada penelitian ini menggunakan teknik menghapus data yang *missing value* salah satunya dengan mengabaikan atau menghapus data tersebut Cara ini biasa dilakukan pada data yang digunakan untuk klasifikasi data mining. Tabel 1 merupakan beberapa contoh data *dummy* klasifikasi rumah tangga miskin dan berikut merupakan contoh menghapus data yang *missing value*:

Tabel 1. Sebelum Dilakukan *Cleaning*

No .	Luas Lantai1	Luas Lantai2	JenisAta p1	JenisAta p2	JenisAta p3	JenisAta p4	JenisAta p5	JenisAta p6	Jenis Dinding1	Jenis Dinding2	Jenis Dinding3
1	1	0	0	1	0	0	0	0	1	0	0
2	0	0	0	1	0	0	0	0	1	0	0
3	0	0	0	1	0	0	0	0	0	0	1
4	1	0	0	1	0	0	0	0	0	0	0
5	1	0	0	1	0	0	0	0	1	0	0
6	1	0	0	1	0	0	0	0	0	0	1

Pada data nomor 2, 3, dan 4 terdapat atribut Luas Lantai dan Jenis Dinding yang *missing value* dibuktikan dengan tidak adanya nilai

1 diantara atribut tersebut. Sehingga data harus dihapuskan dan hasilnya seperti pada Tabel 2.

Tabel 2. Sesudah Dilakukan *Cleaning*

N o .	Luas Lantai1	Luas Lantai2	JenisAta p1	JenisAta p2	JenisAta p3	JenisAta p4	JenisAta p5	JenisAta p6	Jenis Dinding 1	Jenis Dinding 2	Jenis Dinding 3
1	1	0	0	1	0	0	0	0	1	0	0
5	1	0	0	1	0	0	0	0	1	0	0
6	1	0	0	1	0	0	0	0	0	0	1

b. *Data Transformation*

Transformasi data bertujuan untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis ragam. Adapun beberapa teknik transformasi data, yaitu: *smoothing*, agregasi, generalisasi, normalisasi, dan *attribute constraction*. Beberapa teknik data mining membutuhkan format data khusus sebelum bisa diaplikasikan. Data yang didapatkan dari jurnal berupa data yang sudah dilakukan peng-codingan yaitu data *dummy*, sehingga dibutuhkan transformasi data untuk mengubahnya dalam bentuk data yang sebenarnya karena klasifikasi hanya menerima input data ordinal. Pada penelitian ini menggunakan teknik generalisasi karena mengubah data tingkat rendah digantikan dengan konsep tingkat

yang lebih tinggi. Perubahan data dilakukan secara manual dengan mengurutkan kelas atribut dari yang terkecil sampai dengan yang terbesar.

Tabel 3. Klasifikasi Rtm Data *Dummy*

Luas Lantai1	Luas Lantai2	JenisAtap1	JenisAtap2	JenisAtap3	JenisAtap4	JenisAtap5	JenisAtap6	Jenis Dinding 1	Jenis Dinding 2	Jenis Dinding 3
0	1	0	1	0	0	0	0	1	0	0
1	0	1	0	0	0	0	0	1	0	0
1	0	0	1	0	0	0	0	0	0	1
1	0	0	1	0	0	0	0	1	0	0
1	0	0	1	0	0	0	0	1	0	0

Data pada tabel 3 merupakan data *dummy* yang kemudian akan ditransformasikan menjadi data ordinal dengan cara manual. Keterangan kategori pada data *dummy* dapat dilihat pada tabel 4.

Tabel 4. Keterangan Data *Dummy*

Variabel	Kategori
X ₁ Luas Lantai	Luas Lantai1= <8m ² Luas Lantai2= ≥8m ²
X ₂ Jenis Atap	Jenis Atap1= Beton Jenis Atap2= Genteng Jenis Atap3= Kayu Sirap Jenis Atap4= Seng Jenis Atap5= Asbes Jenis Atap6= Ijuk/rumbia
X ₃ Jenis Dinding	Jenis Dinding1= Tembok Jenis Dinding2= Kayu Jenis Dinding3= Bambu

Tabel 5. Hasil Transformasi Menjadi Data Ordinal

Luas Lantai	Jenis Atap	Jenis Dinding
0	2	1
1	1	1
1	2	3
1	2	1
1	2	1

Tabel 5 merupakan tabel hasil transformasi data *dummy* menjadi ordinal. Berikut adalah urutan data ordinal mulai dari yang terkecil hingga yang terbesar yang sudah ditunjukkan pada tabel 6.

Tabel 6. Keterangan Data Ordinal

Variabel	Skala	Kategori
X ₁ Luas Lantai	Ordinal	0 = <8m ² 1 = ≥8m ²
X ₂ Jenis Atap	Ordinal	1 = Beton 2 = Genteng 3 = Kayu Sirap 4 = Seng 5 = Asbes 6 = Ijuk/rumbia
X ₃ Jenis Dinding	Ordinal	1 = Tembok 2 = Kayu 3 = Bambu

1. Data rumah tangga miskin yang didapatkan pada Kecamatan Jombang, Kabupaten Jombang, Jawa Timur berjumlah 4735 dataset karena dilakukan pembersihan data maka data menjadi berjumlah 4226 dataset.
2. Data akan dikelompokkan menjadi beberapa bagian tujuannya adalah untuk melakukan pengukuran kinerja

berdasarkan jumlah data. Pengelompokan data sebagai berikut:

- a. Data training dan data *testing* 50%.
- b. Data training 60% dan data *testing* 40%.
- c. Data training 70% dan data *testing* 30%.
- d. Data training 80% dan data *testing* 20%.
- e. Data training 90% dan data *testing* 10%.

4.3 Analisa menggunakan Ordinal Class Classifier

Pada penerapan decision tree dengan algoritma *Ordinal Class Classifier* disini, kita ingin melihat Rumah Tangga yang dikategorikan dalam kelas Rumah tangga sangat miskin, Rumah tangga miskin, dan Rumah tangga hampir miskin dengan melihat dari atribut yang telah ditentukan sebelumnya. Adapun yang menjadi konsep pohon keputusan adalah mengubah data yang ada menjadi pohon keputusan, serta dari pohon keputusan akan diubah kedalam bentuk aturan-aturan keputusan (Rule).

- 1) Membuat Pohon Keputusan Algoritma *Ordinal Class Classifier*

Berdasarkan kriteria-kriteria yang digunakan, maka data rumah tangga miskin dikelompokkan kedalam kelas sangat miskin, miskin, dan hampir miskin untuk klasifikasi rumah tangga miskin pada Kecamatan Jombang, Kabupaten Jombang.

Untuk menentukan atribut yang merupakan the best classifier dan diletakkan sebagai root perlu menghitung *entropy* total untuk keseluruhan atribut dan gain untuk semua atribut tersebut.

1. Data training dimasukkan. Data training telah dibagi menjadi beberapa kelompok yaitu 50%, 60%, 70%, 80%, dan 90%.
2. Maka *entropy* total untuk keseluruhan atribut dapat dihitung dengan menggunakan Persamaan (1).
- 3.

$$Entropy(S) = \left(\left(-\frac{4}{15} \right) \times \log_2 \left(\frac{4}{15} \right) + \left(-\frac{6}{15} \right) \times \log_2 \left(\frac{6}{15} \right) + \left(-\frac{5}{15} \right) \times \log_2 \left(\frac{5}{15} \right) \right)$$

4. Untuk menghitung gain kriteria luas lantai adalah dengan menghitung *entropy* subkriteria yang ada pada kriteria Luas Lantai seperti berikut ini.

S₁ : dataset kriteria ≥8m²

$$Entropy(S_1) = \left(\left(-\frac{0}{2} \right) \times \log_2 \left(\frac{0}{2} \right) + \left(-\frac{2}{2} \right) \times \log_2 \left(\frac{2}{2} \right) + \left(-\frac{0}{2} \right) \times \log_2 \left(\frac{0}{2} \right) \right) = 0$$

S₂ : dataset kriteria luas lantai <8m²

$$Entropy (S_2) = \left(\left(-\frac{4}{13} \right) \times \log_2 \left(\frac{4}{13} \right) + \left(-\frac{4}{13} \right) \times \log_2 \left(\frac{4}{13} \right) + \left(-\frac{5}{13} \right) \times \log_2 \left(\frac{5}{13} \right) \right) = 1,58$$

Setelah semua nilai *entropy* diketahui maka nilai Gain untuk kriteria luas lantai bisa dihitung dengan menggunakan Persamaan (2).

$$Gain(S, luas_lantai) = 1,57 - \left(\left(\frac{2}{15} \times 0 \right) + \left(\frac{13}{15} \times 1,58 \right) \right) = 0,2$$

a. *Gain ratio* dan *Split Info* Luas Lantai

Untuk menghitung *Gain ratio* kriteria luas lantai adalah dengan menghitung *Split Information* terlebih dahulu menggunakan persamaan (3) dan *Gain ratio* menggunakan persamaan (4), setelah itu dilanjutkan dengan melakukan perhitungan gain baru seperti berikut:

$$SplitInfo(S, luas_lantai) = - \left(\left(\frac{2}{15} \log \frac{2}{15} \right) + \left(\frac{13}{15} \log \frac{13}{15} \right) \right) = 0,57$$

$$GainRatio(S, luas_lantai) = \frac{0,20}{0,57} = 0,35$$

Tabel 7 merupakan hasil perhitungan *Gain ratio* dari iterasi 1 sampai iterasi 3

Tabel 7. Hasil Perhitungan *Gain ratio* Dari Iterasi 1 Sampai Iterasi 3

No.	Cabang	Iterasi 1	Iterasi 2	Iterasi 3
1	Luas Lantai	0.35	0.27	1
2	Jenis Atap	0	0.00	0
3	Jenis Dinding	1	-	-
4	Jenis Lantai	0.17	0.12	0.09
5	Tempat BAB	0.34	0.25	0.07
6	Sumber Penerangan	0.63	0.64	-
7	Sumber air minum	0.40	0.20	0.45
8	Bahan Bakar	0.07	0.08	0.34
9	Konsumsi daging	0	0.00	0
10	Konsumsi pakaian	0.39	0.00	0
11	Konsumsi Makanan perhari	0.39	0.00	0
12	Pengobatan	0.32	0.25	0.07
13	Ijazah terakhir	0.16	0.33	0.21
14	Kepemilikan Aset	0.25	0.05	0.19

b. Tahapan Pruning

Pada tahap ini bagian pohon yang kurang penting dilakukan pruning (pemangkasan) sehingga didapatkan pohon kalsifikasi yang optimal. Untuk melakukan pruning digunakan persamaan (5).

4.4 Pengujian

Hasil analisis akan di hadirkan dalam bentuk *contingency table*. Berdasarkan pengujian yang telah dilakukan terdiri dari pengujian training dan *testing* dengan menggunakan *contingency table*. Berikut ini adalah hasil pengujian ketiga algoritma tersebut:

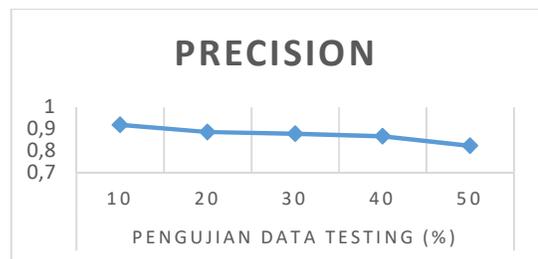
Tabel 8. Perbandingan Hasil Klasifikasi Data Training

Meto de	Pengujian Data Training (%)				
	50	60	70	80	90
OCC Accuracy	90.8%	92.1 %	92 %	92.2 %	93.1%
Precison	0.909	0.923	0.922	0.923	0.931
Recall	0.908	0.921	0.920	0.922	0.931

Tabel 9. Perbandingan Hasil Klasifikasi Data *Testing*

Metode	Pengujian Data Training (%)				
	50	40	30	20	10
OCC Accuracy	82.4 %	86.8%	87.9 %	87.9 %	90.5 %
Precision	0.823	0.868	0.878	0.887	0.919
Recall	0.823	0.867	0.879	0.879	0.905

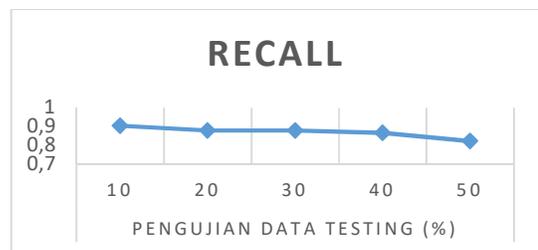
Pengujian dilakukan berdasarkan jumlah data yang dibagi kedalam kelompok pengujian, yaitu kelompok yang memuat 2113 (50%), 2536 (60%), 2958 (70%), 3381 (80%), dan 3803 (90%) data. Gambar 1 merupakan grafik hasil nilai *precision* dari *Ordinal Class Classifier*.



Gambar 1. Grafik Precision

Hasil pengujian menunjukkan bahwa nilai tertinggi diperoleh pada data *testing* yang berjumlah 423 (10%). *Ordinal Class Classifier* memiliki nilai *precision* tertinggi yaitu 0,919.

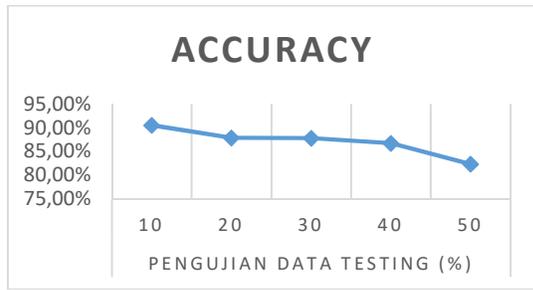
Sedangkan untuk grafik dari hasil *recall* oleh *Ordinal Class Classifier* dapat dilihat pada Gambar 2.



Gambar 2. Grafik Recall

Hasil pengujian menunjukkan bahwa nilai tertinggi diperoleh pada data *testing* yang berjumlah 423 (10%). *Ordinal Class Classifier* memiliki nilai *recall* tertinggi yaitu 0,905.

Gambar 3 merupakan grafik tingkat *accuracy* dari kelompok dari yang diperoleh menggunakan perhitungan *Ordinal Class Classifier*.



Gambar 3 Grafik Accuracy

Hasil pengujian menunjukkan bahwa nilai tertinggi diperoleh pada data *testing* yang berjumlah 423 (10%). *Ordinal Class Classifier* memiliki nilai *accuracy* tertinggi yaitu 90.5437 %.

Dari pengujian yang telah dilakukan berdasarkan kelompok data yang sudah ditentukan dapat disimpulkan bahwa *Ordinal Class Classifier* memiliki tingkat *accuracy*, *precision*, dan *recall* pada pengujian jumlah data 423 (data *testing* 10% dan data training 90%).

4.5 Evaluasi

Evaluasi dilakukan untuk meninjau apakah *Ordinal Class Classifier* sudah menghasilkan sebuah hasil klasifikasi rumah tangga miskin yang sesuai. Evaluasi dilakukan dengan memasukan data rumah tangga miskin di Kecamatan Jombang kemudian membandingkan dan menghitung selisih antara hasil data *testing* klasifikasi rumah tangga miskin yang dihasilkan oleh sistem dengan data rumah tangga miskin sebenarnya.

Berdasarkan tabel hasil evaluasi maka didapatkan hasil prediksi, sehingga dapat dihitung akurasi untuk masing-masing algoritma. Adapun perhitungan akurasi sebagai berikut:

1. *Ordinal Class Classifier* pada data training 90% dan data *testing* 10%

=== Confusion Matrix ===

a	b	c	<-- classified as
96	3	0	a = 1
6	255	23	b = 2
0	8	32	c = 3

Prediksi yang benar = 96+255+32 = 382

Jumlah data *testing* = 423

Sehingga,

$$\text{akurasi} = \frac{382}{423} \times 100\% = 90,5437\%$$

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil pengujian dan pembahasan dapat disimpulkan sebagai berikut:

Dalam penelitian ini dilakukan perhitungan menggunakan model *Ordinal Class Classifier*. Keakuratannya diuji dengan mengambil data sebesar 10% hingga 90% dari jumlah data 4226. Pengujian dilakukan dengan menghitung nilai *precision*, *recall*, dan *accuracy*. Berdasarkan pengujian data nilai akurasi tertinggi yaitu 90,5437% diperoleh pada pengujian data training 90% dan data *testing* 10%.

5.2. Saran

Beberapa saran yang penulis sampaikan bagi pengembang penelitian ini antara lain adalah:

1. Dapat dilakukan dengan menggunakan metode data mining klasifikasi lainnya yang mendukung pengujian data yang ada.
2. Penelitian ini dapat dikembangkan dengan menambahkan kategori pada klasifikasi rumah tangga miskin, misalnya pendapatan kepala keluarga.

DAFTAR PUSTAKA

- [1] Inayah, Riza., & dkk. (2013). Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang Berdasarkan Faktor-faktor yang Mempengaruhi dengan Pendekatan CART (Classification and Regression Trees). *Jurnal Sains dan Seni Pomits*, Vol.3, No.2, 2337-3520.
- [2] World Bank Institute. 2002. *Dasar-dasar Analisis Kemiskinan*. Edisi Terjemahan. Badan Pusat Statistik, Jakarta.
- [3] Sugiyono. 2012. *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- [4] Defiyanti S & dkk. (2016). Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi Spam-Mail.
- [5] Ben David, A., Sterling, Leon., & Tran TriDat. (2009) Adding monotonicity to learning algorithms may impair their *accuracy*. *Expert Systems with Application*, 36, 6627-5534.
- [6] Jhonpita, Phaiboon., & dkk. (2009). *Ordinal Classification Method for the Evaluation of Thai Non-life Insurance Companies*. Chulalongkorn University, Bangkok, Thailand
- [7] Turban, et al., & dkk. (2007). *Decision Support System and Intelligent System Seventh Edition*. New Delhi: Prentice Hal
- [8] Han J, Kamber M. (2001). *Data Mining : Concepts and Techniques*. Simon Fraser University, Morgan Kaufmann Publishers.
- [9] Kantardzic, M. (2011). *Data Mining, Concepts, Models, Methods, and Algorithms*. IEEE Press, A John Wiley & Sons, Inc, Publication.
- [10] Nugraha C, & dkk. (2016). Penerapan Metode Decision Tree(Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa Smpn1 Kintamani. *Seminar Nasional Vokasi dan Teknologi (SEMNASVOKTEK)*. Denpasar-Bali, 22 Oktober 2016.

BIODATA PENULIS



Faried Effendy, M.Kom dilahirkan di Surabaya, 06 Juni 1982. Menyelesaikan pendidikan S-1 (Matematika Unair) lulus Tahun 2004 dan S-2 (Teknik Informatika ITS) lulus tahun 2013.



Purbandini, M.Kom dilahirkan di Surabaya 12 Juli 1971. Menyelesaikan pendidikan S1 Matematika di UNPAD dan S2 Teknik Informatika di ITS.