

Analisa Kesejahteraan Rumah Tangga Menggunakan Metode C5.0

Aisyah Wisri, Yuliana Susanti, Isnandar Slamet

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

aisyahwisri@gmail.com

Info Artikel

Riwayat Artikel

Diterima: 21 Oktober 2019

Direvisi: 18 November 2019

Diterbitkan: 15 Januari 2020

Kata Kunci

Kemiskinan
status rumah tangga
klasifikasi
pohon keputusan
c5.0

ABSTRAK

Kemiskinan merupakan hal lumrah yang dapat ditemui di negara berkembang. Hal ini terjadi ketika masyarakat tidak mampu memenuhi kebutuhan hidup dasar yang berdampak pada kesejahteraan rumah tangga. Penelitian ini dilakukan kepada setiap kepala rumah tangga untuk menentukan pola status rumah tangga dan dapat dijadikan bahan analisis pemerintah untuk menentukan status rumah tangga di masa yang akan datang. Klasifikasi berdasarkan pada status kesejahteraan, jenis kelamin, pendidikan terakhir, lapangan usaha, dan jumlah anggota keluarga. Teknik klasifikasi menggunakan pohon keputusan dapat digunakan untuk mengelompokkan data ke suatu kelas tertentu. Pohon keputusan yang digunakan dalam penelitian ini adalah algoritma C5.0.

*Copyright © 2019 SIMANIS.
All rights reserved.*

Korespondensi :

Aisyah Wisri
Prodi Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Sebelas Maret
aisyahwisri@gmail.com

1. PENDAHULUAN

Kemiskinan merupakan hal lumrah yang dapat ditemui di negara berkembang. Sebagai negara berkembang, Indonesia masih belum dapat mengatasi angka kemiskinan yang tinggi seutuhnya. Badan Pusat Statistik (BPS) merilis, meskipun telah menurun sebesar 9,82% ditahun 2018, namun angka kemiskinan penduduk Indonesia masih tergolong tinggi. Dari sekian banyak kebutuhan manusia, kebutuhan pangan, sandang, dan papan masih menjadi kebutuhan pokok yang mesti selalu menempati urutan atas dalam hal permintaan kebutuhan masyarakat [1].

Indonesia memiliki 34 provinsi, salah satunya provinsi Jawa tengah. Pada Provinsi Jawa Tengah, Kabupaten Wonosobo menempati urutan pertama dengan jumlah penduduk miskin terbanyak. Emil dalam [2] mengemukakan bahwa pendidikan yang rendah merupakan salah satu dari karakteristik yang dimiliki penduduk miskin. Seperti yang terdapat pada Indikator Pembangunan Daerah Kabupaten Wonosobo (2017), Angka Melek Huruf (AMH) dan Angka Melanjutkan (AM) pendidikan yang masih dibawah target pemerintah Kabupaten Wonosobo, menjadi salah satu faktor pendukung rendahnya tingkat pendidikan di kabupaten ini.

[3] mengatakan adanya standar dalam penetapan upah memberi pengaruh negatif terhadap penyerapan tenaga kerja masyarakat berpendidikan rendah. Pendidikan yang terhenti di jenjang menengah menyebabkan banyak masyarakat tidak masuk kualifikasi sebagai tenaga kerja dengan upah standar sehingga tidak sedikit yang memilih pekerjaan kasar, seperti buruh dan lainnya, namun masih tidak dapat menutupi biaya hidup yang cukup tinggi.

Menindak lanjutu rendahnya tingkat pendidikan yang menjadi faktor pendukung tingkat kemiskinan di Kabupaten Wonosobo, pemerintah telah menerapkan berbagai kebijakan sebagai upaya menanggulangi angka kemiskinan.

Dalam rangka menerapkan berbagai program penanggulangan kemiskinan, informasi mengenai siapa yang miskin dan dimana mereka berada menjadi modal penting bagi pemerintah, agar nantinya program penanggulangan dapat berhasil dan tepat sasaran.

2. KAJIAN PUSTAKA

Penelitian [4], membandingkan algoritme C5.0 dan algoritme CART untuk memprediksi konsumen dalam pembuatan kartu keanggotaan dan menyatakan bahwa metode C5.0 memiliki algoritme yang lebih akurat, mencapai 80% sedangkan metode CART hanya mencapai 70 %.

[5] melakukan penelitian tentang prediksi kebangkrutan dari perusahaan elektronik di Taiwan. Penelitian tersebut menggunakan metode pohon keputusan *CART*, *C5.0*, dan *Genetic Programming*. dalam penelitian tersebut, diketahui bahwa *Return on Total Assets (ROA)*, rasio saat ini, rasio harga saham terhadap arus kas, rasio perputaran aktiva tetap, Rasio kepemilikan pemegang saham utama, penghasilan setelah pajak, Rasio cakupan, *Distance to Default (DD)*, mempengaruhi nilai finansial suatu perusahaan.

3. METODE PELAKSANAAN

Salah satu cara untuk mendapat informasi dari data penduduk miskin menggunakan *data mining*. Pada penelitian ini data diolah menggunakan teknik klasifikasi *data mining* yaitu pohon keputusan dengan metode *C5.0*.

A. Data Mining

Menurut [6] definisi *data mining* adalah proses pencarian pola-pola yang tersembunyi berupa pengetahuan yang tidak diketahui sebelumnya dari sekumpulan data yang mana data tersebut dapat berada didalam *database*, *data warehouse*, atau media penyimpanan lainnya.

B. Klasifikasi

Klasifikasi merupakan teknik untuk memberikan pengelompokan baru pada data yang ada dengan menggunakan sejumlah aturan. Aturan-aturan yang ada berasal dari data yang telah di klasifikasi sebelumnya. Berdasarkan metode pada data mining klasifikasi, penelitian ini menggunakan metode pohon keputusan.

C. Pohon Keputusan

Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan antara sejumlah variabel dependen dengan variabel independen.

Menurut [5] pohon keputusan adalah salah satu metode klasifikasi yang paling populer dan mudah untuk diinterpretasikan.

Sementara menurut [7] proses pada pohon keputusan adalah mengubah bentuk data awal menjadi model pohon keputusan kemudian mengubah model pohon keputusan menjadi aturan.

D. Algoritma C5.0

Algoritma *C5.0* merupakan algoritma klasifikasi yang menghasilkan pohon keputusan ditemukan oleh Ross Quinlan pada tahun 1987. Algoritma ini merupakan penyempurnaan dari algoritma *ID3* dan *C4.5*. Algoritma *C5.0* lebih baik daripada algoritma *C4.5* dalam hal memori dan akurasi [4].

Kecepatan dalam membuat pohon keputusan melalui *software* dinilai sangat cepat dibandingkan algoritma lainnya. Algoritma *C5.0* menerapkan *rule based model* (model berbasis aturan) sehingga memudahkan untuk melihat aturan pada pohon keputusan. Selain itu, algoritma *C5.0* dapat mengatasi *missing value* atau nilai yang hilang. Hal ini menjadi kelebihan *C5.0* yang dinilai lebih unggul dibanding algoritma lainnya.

Menurut [8], model pada algoritma *C5.0* bekerja dengan *split* (pemisahan) sampel berdasarkan atribut yang memiliki *information gain* tertinggi.

Menurut Patil *et.al.* (2012) untuk menghitung informasi himpunan kasus pada kelas *i* digunakan rumus

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i \log_2(p_i)$$

dengan $I(S_1, S_2, \dots, S_m)$ merupakan informasi dari himpunan kasus pada kelas *i*, *m* merupakan banyaknya kelas, p_i merupakan proporsi kelas *i* yang dirumuskan sebagai $p_i = S_i / S$. S_i , S_i merupakan jumlah sampel pada kelas *i*, dan *S* merupakan himpunan kasus. Langkah selanjutnya menghitung informasi himpunan kasus pada kelas *i* dan subset *j* sebagai berikut.

$$I(S_{1j}, S_{2j}, \dots, S_{mj}) = - \sum_{ij=1}^{mj} p_{ij} \log_2(p_{ij})$$

$(S_{1j}, S_{2j}, \dots, S_{mj})$ merupakan informasi dari himpunan kasus kelas *i* dan subset *j* dan p_{ij} merupakan proporsi kelas *i* dan subset *j*.

Informasi dari himpunan kasus pada kelas *i* dan subset *j* dapat digunakan untuk menghitung nilai *entropy*, sebagai berikut.

$$E(A) = \sum_{i=1}^m \frac{S_{1j} + \dots + S_{mj}}{S} (S_{1j}, \dots, S_{mj})$$

dengan $E(A)$ merupakan *entropy* atribut *A* dan S_{ij} merupakan sampel dari kelas *i* dan subset *j* dari atribut *A*.

Langkah terakhir menghitung *information gain* sebagai pemilihan atribut yang digunakan sebagai *node*.

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A)$$

Proses dilakukan sampai subset sampel tidak dapat dilakukan *split*

E. *Confusion Matrix*

Untuk mengevaluasi klasifikasi yang didasarkan pada pengujian untuk memprediksi obyek yang benar dan salah, urutan pengujian ditabulasikan didalam *confusion matrix*, dimana kelas yang diprediksi ditampilkan pada bagian atas matriks dan kelas yang diamati pada bagian kiri matriks.

Menurut [5], setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi. Berikut merupakan model *Confusion matrix* 2 kelas. Adapun *confusion matrix* terdapat pada tabel 1.

Tabel *Confusion Matrix*

| <i>Correct Classification</i> | <i>Classified Matrix</i> | |
|-------------------------------|--------------------------|-----------------|
| | + | - |
| + | benar, positif | salah, positif |
| - | salah, negatif | benar, negative |

4. METODE PENELITIAN

Data yang digunakan merupakan data SUSENAS kabupaten Wonosobo tahun 2018 sebanyak 796 data. Data diolah menggunakan *software R*. atribut yang digunakan terdapat pada Tabel 2.

Tabel 2 Atribut Data

| Variabel | Kategori | | | |
|----------|----------------------|---------------|--|--------------------------------------|
| Y | Status Kesejahteraan | Miskin | Pertambangan | |
| | | Tidak Miskin | | Industri pengolahan |
| X1 | Jenis Kelamin | Laki-laki | Pengadaan listrik, gas, uap panas dan udara dingin | |
| | | Perempuan | | pengelolaan air, limbah, daur ulang |
| X2 | Pendidikan terakhir | Tidak Sekolah | kontruksi | |
| | | Paket A | | perdagangan besar dan eceran |
| | | SD | | pengangkutan dan pergudangan |
| | | MI | | penyediaan akomodasi makan minum |
| | | Paket B | | informasi dan komunikasi |
| | | SMP | | keuangan dan asuransi |
| | | MTs | | ahli profesional, ilmiah, dan teknis |
| | | Paket C | | Ketenagakerjaan |
| | | SMA | | adm. Pemerintah |
| | | MA | | Pendidikan |
| | | SMK | | kesehatan manusia dan sosial |
| | | D1/D2 | | kesenian, liburan dan rekreasi |
| | | D3 | | jasa lainnya |
| S1 | rumah tangga | | | |
| S2 | | | | |
| S3 | | | | |
| X3 | Lapangan Usaha | Pertanian | X4 Anggota dalam rumah tangga | |
| | | Holtikultura | | 1 |
| | | Perkebunan | | 2 |
| | | Perikanan | | |
| | | Peternakan | | |
| | Kehutanan | | | |

| | |
|--|---|
| | 3 |
| | 4 |
| | 5 |

| | |
|--|----|
| | 6 |
| | 7+ |

Adapun langkah-langkah dalam penelitian ini yaitu, mendeskripsikan data SUSENAS tahun 2018. Selanjutnya dilakukan perhitungan *information gain* tertinggi untuk memperoleh *node* akar menggunakan persamaan 1. Perhitungan *information gain* dilakukan untuk memperoleh *node* cabang dan *node* terminal.

Berikutnya dilakukan perhitungan *confusion matrix* untuk mengetahui prediksi benar dan salah dari klasifikasi menggunakan persamaan 2 dan 3.

5. HASIL DAN PEMBAHASAN

Ringkasan jumlah kategori masing-masing atribut yang terdapat pada Tabel 3 digunakan untuk menentukan *node*.

Tabel 3 Jumlah Kategori masing-masing atribut

| | | Miskin | Tidak miskin | total | | | | | |
|---------------------|----------------------------|--------|--------------|-------|----------------------------|-------------------------------------|-----|----|-----|
| | Jumlah | 598 | 197 | 796 | | panas dan udara dingin | | | |
| Jenis Kelamin | Laki-laki | 522 | 180 | 702 | | pengelolaan air, limbah, daur ulang | 0 | 2 | 2 |
| | Perempuan | 63 | 30 | 93 | | kontruksi | 51 | 9 | 60 |
| Pendidikan Terakhir | Tidak sekolah | 24 | 3 | 7 | | perdagangan besar dan eceran | 68 | 34 | 102 |
| | Paket A | 2 | 1 | 3 | | pengangkutan dan pergudangan | 26 | 7 | 33 |
| | SD | 380 | 88 | 468 | | penyediaan akomodasi makan minum | 5 | 10 | 15 |
| | MI | 27 | 7 | 34 | | informasi dan komunikasi | 2 | 0 | 2 |
| | Paket B | 13 | 3 | 16 | | keuangan dan asuransi | 4 | 5 | 9 |
| | SMP | 68 | 22 | 90 | | tenaga ahli dan profesional | 1 | 0 | 1 |
| | MTs | 13 | 5 | 18 | | ketenagakerjaan | 1 | 0 | 1 |
| | Paket C | 5 | 3 | 8 | | adm. Pemerintah | 9 | 11 | 20 |
| | SMA | 25 | 21 | 46 | | pendidikan | 6 | 8 | 14 |
| | MA | 2 | 1 | 3 | | kesehatan manusia dan sosial | 2 | 1 | 3 |
| | SMK | 20 | 16 | 36 | | kesenian, liburan dan rekreasi | 1 | 1 | 2 |
| | D1/D2 | 4 | 2 | 6 | | jasa lainnya | 24 | 11 | 35 |
| | D3 | 3 | 2 | 5 | | rumah tangga | 1 | 0 | 1 |
| | S1 | 11 | 18 | 29 | Anggota Dalam Rumah Tangga | | | | |
| | S2 | 1 | 4 | 5 | | | 1 | 5 | 50 |
| | S3 | 0 | 1 | 1 | | 2 | 80 | 70 | 150 |
| Lapangan Usaha | Pertanian | 108 | 15 | 123 | | 3 | 164 | 41 | 205 |
| | Holtikultura | 121 | 26 | 147 | | 4 | 182 | 28 | 210 |
| | Perkebunan | 13 | 5 | 18 | | | | | |
| | Perikanan | 0 | 1 | 1 | | | | | |
| | Peternakan | 19 | 5 | 24 | | | | | |
| | Kehutanan | 11 | 2 | 13 | | | | | |
| | Pertambangan | 4 | 1 | 5 | | | | | |
| | Industri pengolahan | 52 | 17 | 69 | | | | | |
| | Pengadaan listrik, gas,uap | 0 | 1 | 1 | | | | | |

| | | | | | | | | | |
|--|---|----|---|-----|--|----|----|---|----|
| | 5 | 97 | 7 | 104 | | 7+ | 25 | 0 | 25 |
| | 6 | 44 | 1 | 45 | | | | | |

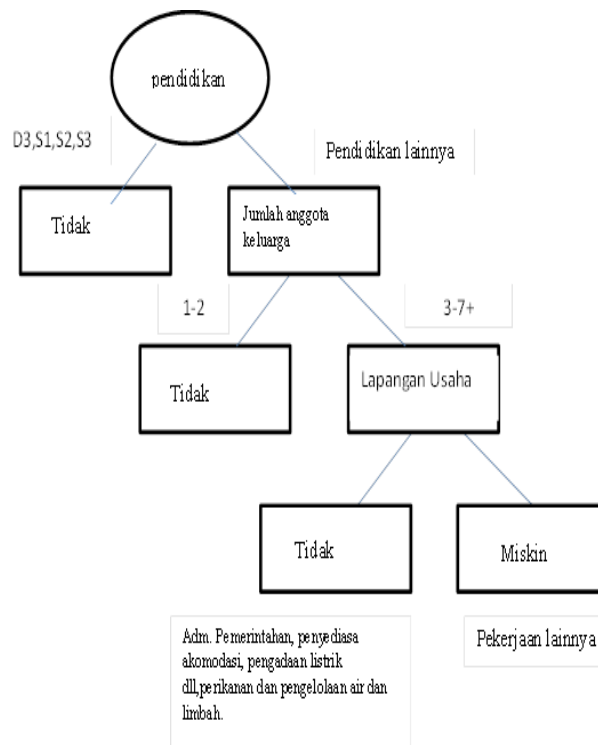
Tabel 4 menunjukkan hasil *information gain* pada masing-masing atribut.

Tabel 4 Nilai *information gain* per atribut

| | variabel | information gain |
|----|-------------------------|------------------|
| X1 | jenis kelamin | 0.00429 |
| X2 | pendidikan terakhir | 0.032848 |
| X3 | lapangan usaha | 0.01091 |
| X4 | jumlah anggota keluarga | 0.031423 |

Berdasarkan Tabel 4 *information gain* tertinggi adalah variabel pendidikan. Oleh karena itu, pendidikan dijadikan sebagai *node* akar. *Node* akar dipilih berdasarkan nilai *information gain* tertinggi setelah dilakukan penghapusan terhadap atribut yang menjadi *node* akar.

Nilai *information gain* pada iterasi mengklasifikasikan semua kasus ke dalam kelas tertentu. Pohon keputusan yang terbentuk berdasarkan *information gain* tertinggi ditunjukkan pada Gambar 1.



Gambar 1. Pohon algoritma C5.0

Dari pohon keputusan diatas didapatkan klasifikasi :

1. Masyarakat dengan pendidikan D3, S1,S2 dan S3 tidak tergolong masyarakat miskin
2. Masyarakat dengan pendidikan akhir lainnya dan memiliki anggota keluarga 1-2 tidak tergolong masyarakat miskin
3. Masyarakat dengan anggota keluarga 3- 7+ dan bekerja dibidang administrasi pemerintahan, penyedia jasa akomodasi, pengadaan listrik, perikanan dan pengelolaan air serta limbah tidak termasuk masyarakat miskin
4. Masyarakat dengan anggota keluarga 3-7+ orang serta memiliki pekerjaan lainnya termasuk kategori masyarakat miskin.

6. KESIMPULAN DAN SARAN

Dari data SUSENAS kabupaten Wonosobo ini didapatkan empat klasifikasi dengan menggunakan algoritma C5.0. hasil yang didapatkan diharapkan dapat berguna bagi pemerintah untuk menentukan status keluarga yang layak mendapat bantuan.

Untuk penelitian selanjutnya dapat menambah beberapa variabel untuk pembentukan pola status masyarakat yang lebih akurat.

DAFTAR PUSTAKA

- [1] Suryana, *Kewirausahaan*, Jakarta : Salemba Empat, 2008.
- [2] Supriatna, *Birokrasi Pemberdayaan dan Pengentasan Kemiskinan*, Bandung : Humaniora Utama Press, 1997.
- [3] Gianne. *Pengaruh Upah Minimum terhadap Penyerapan Tenaga Kerja Berpendidikan Rendah di Sektor Industri dan Perdagangan*, Thesis, Universitas Indonesia, 2009.
- [4] Patil, N., R. Lathi, and V. Chitre, *Customer Card Classification Based on CART and C5.0 Algorithms*. International Journal of Engineering Research and Applications (IJERA) 2 , 2012.
- [5] Han, J. and M. Kamber, *Data Mining Concept and Technique*. San Fransisco: Morgan Kauffman, 2006.
- [6] Kusriani dan Emha T., *Algoritma Data Mining*. Yogyakarta : ANDI, 2009.
- [7] Jayanti, N., S.Puspitodjati, dan T. Elida, *Teknik Klasifikasi Pohon Keputusan Untuk Memprediksi Kebangktan Bank Berdasarkan Rasio Keuangan Bank*, Universitas Gunadarma, 2008.
- [8] Elsayad, A.M., & Elsalamony, H.A. (2013). Diagnosis Breast Cancer using Decision Tree Models and SVM. *International Journal of Computer Applications*, Vol. 83, No. 5.