

IMPLEMENTASI ALGORITMA C4.5 UNTUK MENENTUKAN PENERIMA BANTUAN SOSIAL (STUDI KASUS: KLASIFIKASI KEMISKINAN DESA LURAH)

Washilaturrizqi, Odi Nurdiawan

Program Studi Manajemen Informatika D3, STMIK IKMI Cirebon
Jalan Perjuangan No. 10B Majasem, Kota Cirebon, Indonesia
Chilajri3@gmail.com

ABSTRAK

Karena kemiskinan tidak dapat dengan mudah diberantas di Indonesia, pemerintah telah membuat beberapa program untuk mengatasi masalah ini, antara lain Program Bantuan Pangan Non Tunai, yaitu program bantuan pangan nontunai yang diberikan oleh pemerintah. kepada keluarga penerima manfaat setiap bulan. Persyaratan WNI antara lain fakir miskin, non ASN, TNI/Polri. Kriteria penerima bantuan sembako adalah tingkat pendidikan, jenis pekerjaan, jenis atap, jenis dinding, jenis lantai, tempat buang air besar dan jumlah anggota keluarga dalam satu rumah. Kesejahteraan keluarga di Desa Lurah merupakan permasalahan yang dialami oleh masyarakat, dimana tingkat kesejahteraan keluarga kurang baik sehingga mereka menyalurkan dana negara yang tidak tepat sasaran. Tujuan dari klasifikasi masalah ini adalah melakukan penelitian untuk mendapatkan hasil yang akurat tentang apakah penerima bantuan sembako nontunai dalam keadaan sejahtera keluarga atau tidak. model yang digunakan yaitu model algoritma klasifikasi C .5 menggunakan 11 atribut, pengujian 95 data yang dibagi menjadi dua bagian yaitu data pelatihan 80 data pengujian 20 memberikan nilai akurasi sebesar 58,18% pada pengujian ini

Kata kunci: Klasifikasi, kemiskinan, data mining dan C4.5

1. PENDAHULUAN

Program pemberdayaan pemerintah Indonesia saat ini belum mampu mendidik masyarakat untuk keluar dari kemiskinan, karena program bantuan tersebut tetap menjadi prioritas pemerintah. Program yang sedang berlangsung harus memberikan peluang untuk menciptakan manusia yang cerdas untuk memecahkan masalah mereka, terutama kemiskinan. Masalah kemiskinan merupakan salah satu masalah utama yang menjadi fokus perhatian pemerintah setiap negara. Kemiskinan juga merupakan salah satu masalah yang dihadapi oleh banyak negara berkembang, termasuk Indonesia. Banyak upaya telah dilakukan untuk mengatasi kemiskinan, termasuk program bantuan miskin ini. Pemerintah juga melakukan segalanya untuk menyelesaikan masalah yang muncul. Pemerintah mengeluarkan kebijakan untuk memberdayakan keluarga miskin, salah satu kebijakan pemerintah adalah bantuan sembako segera. (Sari & Saro, 2018)

Mengingat jumlah penduduk miskin yang banyak, pemerintah selalu berusaha mengatasi masalah kemiskinan melalui berbagai program bantuan seperti beras miskin (Raski), Program Keluarga Harapan (PKH), Bantuan Langsung Tunai (BLT) dan pencairan. bantuan pangan (BPNT). Bantuan sembako non tunai adalah bantuan sembako pemerintah menyalurkan gratifikasi kepada Keluarga Penerima Manfaat (KPM) setiap bulan. Program Pengolahan Bantuan Pendapatan memiliki syarat dan kriteria tersendiri untuk pembelian makanan dari e-warung melalui mekanisme e-money. Syaratnya, peserta yang menerima bantuan sembako nontunai adalah warga negara Indonesia termasuk fakir miskin,

dan bukan anggota ASN, TNI atau Polri, serta ada kriteria penerima bantuan sembako nontunai yaitu tipe atap, tipe tembok, jenis lantai, tempat buang air besar, pekerjaan, jumlah anggota keluarga dalam satu rumah dan pendidikan. Sebelum menerima bantuan KPM, KPM terlebih dahulu harus melakukan seleksi berupa verifikasi dan validasi data yang dilakukan oleh pendamping. Dengan demikian, diperlukan suatu sistem simulasi yang dapat mengklasifikasikan keluarga penerima secara cepat dan akurat. Penambahan data memudahkan lembaga untuk menentukan keluarga mana yang memenuhi syarat untuk mendapatkan bantuan pangan tanpa uang tunai dan mana yang tidak. (Nopi, 2021)

Banyak warga yang mengeluhkan mengapa mereka tidak mendapatkan bantuan dari pemerintah, padahal mereka sudah lama tinggal di Lurahkylä, dan kebanyakan dari mereka adalah orang asing yang tidak mau mengeluarkan surat pindah karena takut ketika pindah. bantuan yang mereka terima akan dianggap hilang karena mereka tidak lagi memiliki tempat tinggal sendiri. daerah di mana mereka awalnya menerima bantuan.

Tujuan dari tugas akhir ini adalah untuk memudahkan warga negara yang layak menerima dukungan pendapatan negara dan menghilangkan ketimpangan sosial antar warga.

Karena volume data yang besar, diperlukan teknologi untuk mengidentifikasi pola dan informasi dalam data secara cepat, akurat dan jelas. Dengan menggunakan algoritma C4.5 dapat membangun model pohon keputusan yang memprediksi keluarga yang memenuhi syarat dan dapat membantu pejabat pemerintah menentukan pohon keputusan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pada jurnal sistemasi penelitian terdahulu yang dilakukan oleh (Ermawati, 2019) dalam penelitian yang berjudul “algoritma klasifikasi C4.5 berbasis particle swarm optimization untuk memprediksi penerima bantuan pangan non tunai” salah satu kesulitan yang kadang dihadapi oleh pemerintah dalam menentukan kelayakan pada saat mengajukan pendataan, dari masalah yang ada digunakan metode klasifikasi untuk memprediksi kelayakan penerima manfaat bantuan pangan non tunai yaitu dengan menggunakan model algoritma klasifikasi C4.5, setelah dilakukan pengujian tersebut maka algoritma C4.5 menghasilkan nilai akurasi sebesar 98.56%

2.2. Kemiskinan

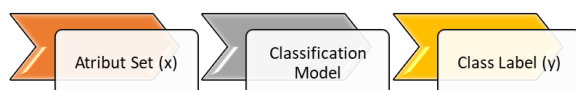
Kemiskinan merupakan masalah sosial yang masih belum terpecahkan di negara berkembang, terutama di Indonesia, karena kemiskinan tidak dapat dengan mudah dihilangkan. Karena masalah kemiskinan di negara berkembang tersebut, pemerintah menerapkan beberapa program untuk mengatasi masalah ini, antara lain Program Bantuan Pangan Non Tunai (BPNT), yaitu program bantuan pangan yang diberikan tanpa uang tunai. dari pemerintah kepada Keluarga Penerima Manfaat (KPM) setiap bulan melalui rekening elektronik yang hanya digunakan untuk membeli sembako di toko sembako dan untuk membeli sembako yang memang perlu dibeli yaitu. nasi, telur, tahu, tempe dan daging..(Fitriani, 2020)

2.3. Kriteria kemiskinan

- 1) Tingkat Pendidikan
- 2) Jenis lapangan usaha atau pekerjaan
- 3) Jenis atap
- 4) Jenis dinding
- 5) Jenis lantai
- 6) Fasilitas buang air besar
- 7) Jumlah anggota keluarga dalam satu rumah

2.4. Model klasifikasi

Klasifikasi adalah metode untuk mendefinisikan sekumpulan pola atau properti yang menggambarkan dan membedakan satu kelas data dari yang lain untuk menentukan apakah suatu objek perilaku dan atribut kelompok yang ditentukan Klasifikasi juga menunjukkan hubungan antar atribut (Agus Sugianto & Rizky Maulana, 2019)



Gambar 1. Klasifikasi sebagai pemetaan himpunan atribut x ke dalam label kategori kelasnya

Sebuah klasifikasi juga disebut sekumpulan record dan bisa juga disebut kasus atau sampel. Tubuh

(x,y) adalah pemetaan rekaman di mana x adalah array atribut dan y adalah label kelas. Sebuah tuple (x,y) adalah pemetaan record, dimana x sebagai himpunan atribut dan y sebagai label kelas, Klasifikasi yang digunakan adalah decision tree, rule-based classifier, neural network, support vector machine dan naïve bayes classifier (Junaedi et al., 2019).

2.5. Klasifikasi biasanya dibagi menjadi dua face:

- 1) Tahap pelatihan Bagian dari data yang kategori datanya diketahui dan dimasukkan untuk membentuk model estimasi
- 2) Tahap pengujian Bagian dari data yang terbentuk diuji dengan bagian data lainnya untuk mengetahui akurasi model

2.6. Algoritma C4.5

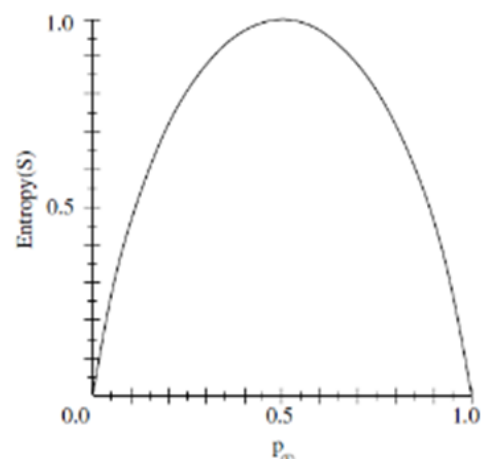
Pohon keputusan adalah metode yang umum digunakan dalam klasifikasi data mining. Metode algoritma C4.5 populer karena dapat mengklasifikasikan dan menunjukkan hubungan antar atribut. Algoritma juga dapat digunakan untuk membangun pohon keputusan, salah satunya adalah algoritma C4.5(Ginting et al., 2020)

2.7. Variasi algoritma C4.5

- a) Banyak algoritma seperti ID3, CART dan C4.5 dapat digunakan untuk menghasilkan pohon keputusan
- b) Algoritma C4.5 merupakan pengembangan lebih lanjut dari algoritma ID3
- c) Proses pohon keputusan terdiri dari mengubah bentuk tabel menjadi model pohon, mengubah model pohon menjadi aturan dan menyederhanakan aturan

Algoritma C4.5 dapat menangani data numerik dan disk. Algoritma ini juga menggunakan gain ratio, sebelum menghitung gain ratio perlu menghitung nilai informasi dalam satuan bit dari himpunan objek dengan menggunakan konsep entropy.

2.8. Konsep entropi



Gambar 2. Grafik Entropy

Entropi(S) adalah perkiraan jumlah bit yang diperlukan untuk membedakan suatu kelas (atau -) dari data acak dalam ruang sampel S. Entropi juga dianggap sebagai bit yang diperlukan untuk menunjukkan suatu kelas. semakin kecil nilai entropi, semakin banyak digunakan untuk pemisahan kelas.

Entropi digunakan untuk mengukur non-orisinalitas S.

Entropy pada ruang contoh S didefinisikan dengan:

$$Entropy(S) \equiv -p_{+} \log_2 p_{+} - p_{-} \log_2 p_{-}$$

Dimana:

- S : status (data) sampel yang digunakan untuk trainer
- P : jumlah solusi positif untuk kriteria tertentu
- P : jumlah solusi negatif untuk kriteria tertentu.
- Entropy(S): 0 jika semua contoh S termasuk dalam kelas yang sama
- Entropy(S) = 1 jika jumlah contoh positif dan negatif S sama
- $0 < Entropy(S) < 1$ jika jumlah contoh positif dan negatif pada S tidak sama

2.9. Konsep Gain

Gain (S,A) adalah perolehan informasi dari atribut A sehubungan dengan data keluaran S. perolehan data diperoleh data output atau variabel dependen S yang dikelompokkan dengan atribut A yang ditunjukkan penguatan (S ,A).

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

Dimana:

- A : Atribut
- S : Sample
- n : Banyaknya bagian dari himpunan atribut A
- |S_i| : jumlah sampel di -i
- |S| : jumlah spesimen di S

2.10. Cara membangun pohon keputusan menggunakan algoritma:

- Pilih atribut sebagai akar. Sebuah akar didapat dari nilai gain tertinggi dari atribut yang ada
- Membuat cabang untuk masing-masing nilai
- Bagi kasus dalam sebuah cabang
- Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Kelebihan pohon keputusan:

- Daerah keputusan yang sederhana dan spesifik
- Tidak perlu menghilangkan perhitungan, karena dengan menggunakan metode pohon keputusan, pengujian sederhana hanya didasarkan pada kriteria atau kelas tertentu
- Mudah diadaptasi untuk memilih properti dari node internal yang berbeda
- Penguji tidak perlu mengevaluasi aspek tertentu dari distribusi tinggi atau distribusi kelas pada parameter.

Kekurangan pohon keputusan:

- Sulitnya merancang pohon keputusan yang optimal
- Hasil kualitas keputusan yang diperoleh tergantung bagaimana pohon itu dirancang
- Ketika jumlah kelas dan kriteria yang digunakan sangat banyak, apa yang menyebabkan keputusan tersebut. waktu dan kapasitas memori yang dibutuhkan
- Akumulasi kesalahan pembacaan yang diperlukan dalam pohon keputusan besar di setiap level.(Kusrini, 2009)

3. METODE PENELITIAN

3.1. Tahap Observasi langsung

pemantauan langsung status masyarakat sesuai dengan keluhan dan permasalahan yang muncul.

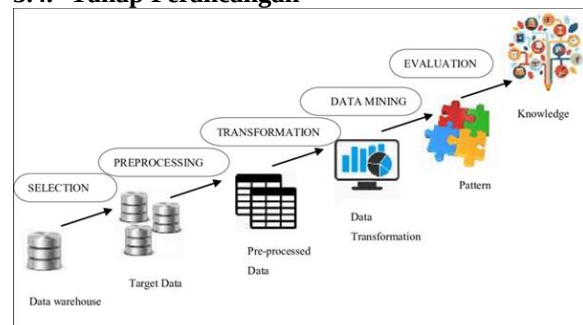
3.2. Tahap Wawancara

mengirimkan tanya jawab kepada RT dan orang-orang yang berhubungan langsung dengan perangkat desa untuk mendapatkan kriteria keluarga miskin

3.3. Studi perpustakaan

Peneliti mencari dukungan di internet, seperti buku, jurnal dan literatur lainnya.

3.4. Tahap Perancangan



Gambar 3. Tahapan Knowledge Discovery In Database

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Pada proyek tahap akhir ini, data diambil dari desa lurah, kemudian data tersebut dimasukkan ke dalam proses pemilahan atau seleksi untuk menghilangkan data yang bertentangan atau duplikat dan memperbaiki kesalahan pada data yang ada. Kemudian setelah memilih data 275 data digunakan.

4.2. Proses Data

Memilih sendiri atau mengolah data yang layak dianalisis, dilanjutkan dengan memilih data yang tidak memenuhi kriteria. Kriterianya sendiri adalah jenis atap, jenis dinding, jenis lantai, tempat buang air besar, pekerjaan, jumlah anggota keluarga dalam satu rumah dan pendidikan.

4.3. Transformasi Data

Proses transformasi data terpilih dalam bentuk mining procedure, Lalu memisahkan menghilangkan kolom yang tidak digunakan (exclude kolom). proses ini menyesuaikan type data atau change type seperti:

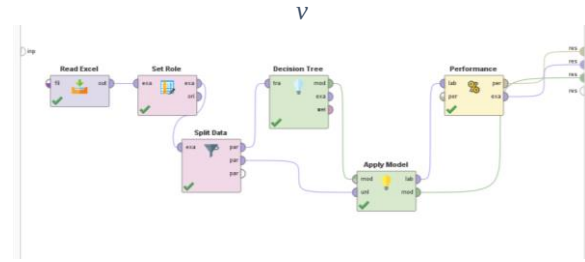
- Polynomial yaitu untuk atribut yang memiliki lebih dari 2 kategori
- Binominal yaitu untuk atribut yang memiliki kategori 2
- Real yaitu untuk tipe data yang memiliki nilai koma
- Integer yaitu untuk bilangan tanpa koma

4.4. Data Mining

Data mining bisa untuk melakukan klasifikasi, klusterisasi, dan menemukan pola asosiasi sehingga bisa melakukan peralaman (*predicting*).

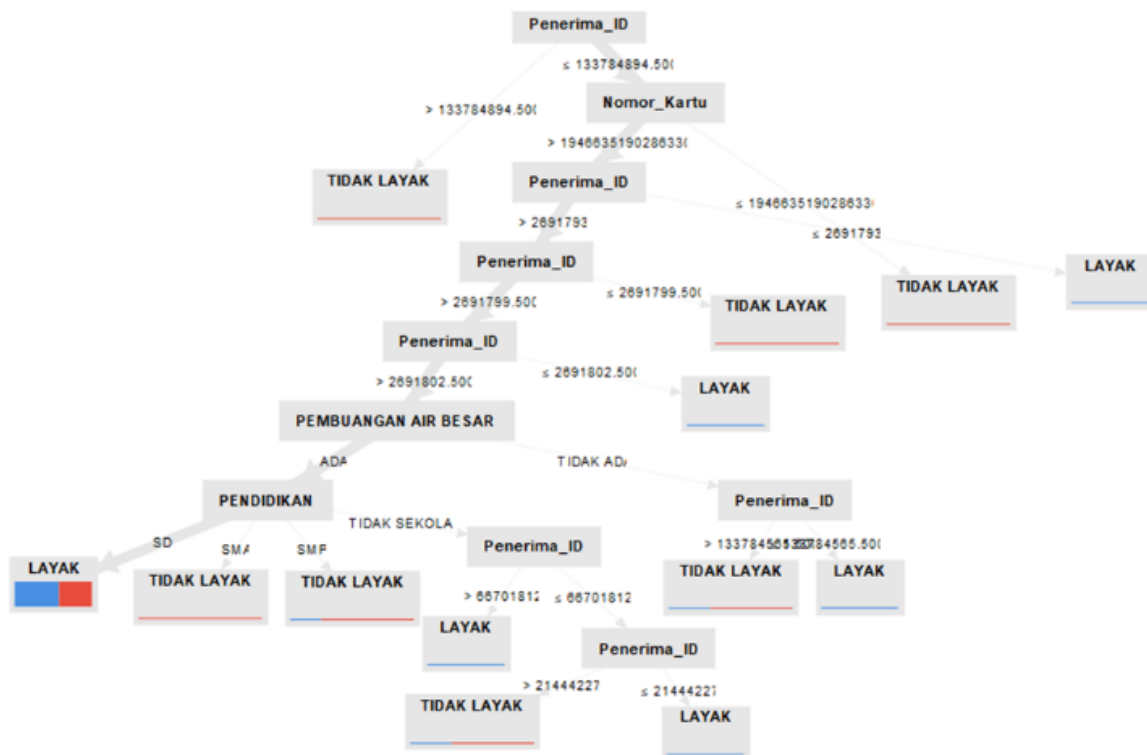
4.6. Decision Tree

4.5. Model Algoritma C4.5



Gambar 4. Model Algoritma C4.5

Gambar 4 merupakan pengujian model algoritma C4.5 menggunakan software Rapidminer. Read Excel pada gambar merupakan tools untuk mengambil data training yang akan dibuat kemudian dihubungkan ke set role dan sampai ke performace.



Gambar 5. Pohon Keputusan

Berdasarkan Gambar 5 maka rule tersebut dimanfaatkan untuk mengambil keputusan pada data baru.

4.7. Nilai Akurasi Algoritma C4.5

accuracy: 58.18%			
	true LAYAK	true TIDAK LAYAK	class prediction
pred. LAYAK	157	113	58.15%
pred. TIDAK LAYAK	2	3	60.00%
class recall	98.74%	2.59%	

Gambar 6. Nilai Akurasi Algoritma C4.5

Gambar 6 diketahui dari 275 data, 157 diklasifikasikan layak sesuai dengan prediksi yang telah dilakukan dengan metode algoritma C4.5, 113 data diprediksi layak tapi ternyata hasilnya tidak layak, 3 data class tidak diprediksi sesuai, dan 2 data diprediksi tidak ternyata layak.

Gambar 6 menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 sebesar 58.18% dan bisa dihitung untuk mencari nilai accuracy adalah tingkat kedekatan antara nilai yang didapat terhadap nilai sebenarnya, sensitivity adalah tingkatan positif yang sebenarnya, specificity, ppv dan npv.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat ditarik beberapa kesimpulan mengenai penerapan algoritma C4.5 dalam klasifikasi penerima bantuan sembako nontunai sebagai berikut: Penerapan algoritma C4.5 dalam klasifikasi penerima bantuan sembako nontunai penerima bantuan sembako tunai. -penerima bantuan sembako tunai. Pengujian algoritma C4.5 dengan akurasi perolehan 58.18. pohon keputusan dibuat dengan algoritma C4.5.

Saran untuk pengembangan penelitian ini lebih lanjut terkait dengan penelitian adalah sebagai berikut: menggunakan lebih banyak atribut yang termasuk dalam kriteria kelayakan bantuan sembako non tunai untuk mencapai akurasi yang lebih besar. menambahkan informasi sedemikian rupa sehingga dapat mempengaruhi nilai akurasi. penelitian ini dilakukan untuk menentukan kelayakan penerima bansos sembako nontunai. menggunakan metode klasifikasi data mining untuk menentukan kelayakan penerima bansos sembako nontunai

DAFTAR PUSTAKA

- [1] Agus Sugianto, C., & Rizky Maulana, F. (2019). Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama). *Techno.COM*, 18(4), 321–331.
- [2] Ermawati, E. (2019). Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai. *Sistemasi*, 8(3), 513. <https://doi.org/10.32520/stmsi.v8i3.576>
- [3] Fitriani, E. (2020). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*, 9(1), 103. <https://doi.org/10.32520/stmsi.v9i1.596>
- [4] Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 36–44. <https://doi.org/10.35585/inspir.v10i1.2535>
- [5] Junaedi, I., Nuswantari, N., & Yasin, V. (2019). Perancangan Dan Implementasi Algoritma C4.5 Untuk Data Mining. *Journal of Information System, Informatics and Computing*, 3(1), 29–44.
- [6] Kusriani, K. & emha taufiq luthfi. (2009). *ALGORITMA DATA MINING*.
- [7] Nopi, N. P. (2021). Klasifikasi Penerima Bantuan Pangan Non Tunai Menggunakan Metode Decision Tree. *Jurnal Informasi Dan Teknologi*, 3, 167–173. <https://doi.org/10.37034/jidt.v3i4.148>
- [8] Sari, F., & Saro, D. (2018). Implementasi Algoritma C4.5 Dalam Menentukan Lokasi Prioritas Penyuluhan Program Keluarga berencana di kecamatan dumai timur. *Jurnal Penelitian Pos Dan Informatika*, 8(1), 63. <https://doi.org/10.17933/jppi.2018.080105>