



INFORMATION GAIN PADA ALGORITMA C4.5 UNTUK KLASIFIKASI PENERIMAAN BANTUAN PANGAN NON TUNAI (BPNT)

Rizal Amegia Saputra¹, Sri Wasiyanti², Denny Pribadi³

^{1,2}Sistem Informasi Akuntansi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

³Ilmu Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

rizal.rga@bsi.ac.id , sri.siw@bsi.ac.id , denny.dpi@bsi.ac.id

Jl. Cemerlang No. 8 Sukakarya Sukabumi

Keywords:

Classification, Non-Cash Assistance, C4.5 Algorithm, Information Gain

Abstract

The problem of poverty in various parts of the world, especially in Indonesia, has become a serious problem that has become the center of attention of the current government. The government has implemented various ways to increase the poverty rate, one of which is providing assistance to the poor through the Non-Cash Food Assistance Program (BPNT). However, stakeholders have problems that occur in the field where it is difficult to classify the data receiving assistance so that it results in less targeted assistance in providing assistance. In this study, using the C4.5 data classification algorithm with split information gain criteria, the C4.5 algorithm is one of the best algorithms in the classification algorithm, and 130 datasets obtained in 2018 in Nagrak Utara Village, Sukabumi. The decision tree model obtained from the C4.5 algorithm with an accuracy of 91.54% and an AUC value of 0.986, thus the C4.5 algorithm to determine the acceptance of BPNT includes a very good range of classification algorithms, this application used can help users.

Kata Kunci:

Klasifikasi, Bantuan Non Tunai, Algoritma C4.5, Information Gain

Abstrak

Masalah kemiskinan diberbagai belahan dunia khususnya indonesia menjadi persoalan serius yang menjadi pusat perhatian pemerintah saat ini. Pemerintah sudah melaksanakan dengan berbagai cara dalam menekan peingkatan angka kemiskinan, salah satunya memberikan bantuan kepada masyarakat miskin melalui perogram Bantuan Pangan Non Tunai (BPNT). Namun pemangku kepentingan memiliki permasalahan yang terjadi dilapangan yaitu sulitnya mengklasifikasikan data yang menerima bantuan sehingganya mengakibatkan kurang tepatnya sasaran dalam memberikan bantuan. Pada penelitian ini menggunakan algoritma klasifikasi datamining C4.5 dengan split kriteria information gain, algoritma C4.5 adalah salahsatu algoritma terbaik pada algoritma klasifikasi, dan dataset yang didapat pada tahun 2018 pada Desa Nagrak Utara Sukabumi sebanyak 130 data. Model pohon keputusan yang didapat dari pengujian algoritma C4.5 yaitu memiliki akurasi 91.54% dan nilai AUC 0.986, dengan demikian algoritma C4.5 untuk klasifikasi menentukan penerimaan BPNT itu termasuk range algoritma klasifikasi sangat baik, hal ini aplikasi yang digunakan dapat membantu pengguna.

Pendahuluan

Kemiskinan digambarkan sebagai ketidakmampuan untuk memenuhi kondisi pendapatan rendah atau kebutuhan dasar manusia (misalnya pangan, papan dan sandang) [1].

Terkadang kemiskinan harus terjadi dilingkungan masyarakat karena keadaan yang harus dijalani dalam kehidupan dilingkungan miskin seperti karena krisis ekonomi, gaya hidup dan budaya

dilingkurangan sekitar yang justru menyebabkan warga Indonesia menjadi warga yang termasuk masyarakat miskin[2].

Masalah kemiskinan diindonesia menjadi salah satu persoalan yang mendasar bagi pemerintah pusat untuk segera mengatasinya walau terbilang cukup sulit untuk mengatasinya [3]. Oleh karena itu, pemerintah membuat kebijakan dalam membantu masyarakat dengan program bantuan sosial seperti Bantuan Pangan Non Tunai (BPNT). Program bantuan ini diberikan kepada keluarga fakir miskin agar keluarga kategori fakir miskin dapat bertahan dan mengatasi krisis ekonomi[4], program BPNT diberikan kepada seluruh masyarakat miskin di Indonesia dengan kriteria yang telah di tuntukan oleh pemerintah pusat [5].

Desa Nagrak Utara Sukabumi merupakan desa yang cukup banyak menerima BPNT dari pemerintah, namun banyak data yang keliru sehingga sedikit kesulitan dalam menentukan layak dan belum layak penerimaan BPNT, sehingga ada beberapa yang belum tepat sasaran, oleh karena itu perlu sebuah aplikasi dalam penentuan pemilihan penerimaan BPNT agar tepat sasaran.

Beberapa penelitian mengenai bantuan BPNT yang sudah banyak dilakukan, seperti penelitian [5] dengan menggunakan metode Fuzzy pengujian dengan pengisian nilai bobot yang berubah, akan mempengaruhi hasil ranking dari alternatif yang akan didapatkan, dan terbukti metode fuzzy dapat menentukan siapa yang berhak menerima BPNT, Penelitian dari [6] menerapkan algoritma C5.0 dalam klasifikasi penerima BPNT, algoritma ini dapat membantu memberikan keputusan dalam menentukan pemberian bantuan kepada masyarakat miskin. Penelitian yang dilakukan oleh [7] menerapkan metode SAW untuk membantu dalam menentukan calon penerimaan BPNT dan terbukti bahwa metode SAW dapat di terapkan pada sistem pendukung keputusan untuk

merekomendasikan warga yang berhak menerima BPNT.

Metode C4.5 termasuk kedalam algoritma klasifikasi datamini, algoritma ini telah banyak digunakan dalam penentuan klasifikasi, dan algoritma ini memiliki kelebihan dibandingkan algoritma pohon keputusan lainnya. Salah satu kelebihan dari algoritma C4.5 yaitu dapat membuat rule yang mudah gambarkan dalam sebuah pohon keputusan[8]. Diharapkan dengan menggunakan algoritma C4.5 pada aplikasi penentuan penerimaan BPNT yang akan dibuat, itu dapat membantu pengguna dalam menentukan siapa yang berhak menerima BPNT.

Landasan Teori

Algoritma C4.5

Algoritma C4.5 termasuk algoritma yang dapat digunakan dalam proses klasifikasi data. Algoritma ini memiliki banyak keunggulan, sehingga banyak digunakan [9].

Dalam perhitungan algoritma C4.5 memiliki dua tahap yaitu pertama menghitung nilai *Entropy* dan yang kedua *Gain*.

Nilai *entropy* berfungsi untuk mengukur jumlah dari informasi yang ada pada atribut dengan persamaan berikut[10]:

$$Entropy(S) = - \sum_{i=1}^n Pi * LOG_2 Pi \quad (1)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi S

pi : proporsi dari Si terhadap S

sedangkan nilai *Gain* dapat digunakan dalam mengukur efektivitas atribut dalam proses klasifikasi[11], dengan persamaan sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \left| \frac{Si}{S} \right| * Entropy(Si) \quad (2)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Confusion Matrix

Model penilaian dari metode klasifikasi yang didasarkan dalam banyaknya data pengujian yang dapat diprediksi secara *True* dan *False* [12], *Confusion matrix* memiliki informasi yang dapat membandingkan hasil klasifikasi, berikut persamaan *Confusion Matrix*:

Tabel 1. Persamaan *Confusion matrix*[13]

Kelas Prediksi	Kelas Prediksi	
	Benar	Salah
Benar	TP	FN
Salah	FP	TN

Dari tabel 1 diatas dapat dijelaskan bahwa keakuratan model algoritma dapat dihitung menggunakan persamaan berikut [14]:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Keterangan:

TP: Jumlah Data positif yang diklasifikasikan dengan benar oleh komputer

TN: Jumlah Data negatif yang diklasifikasikan dengan benar oleh komputer

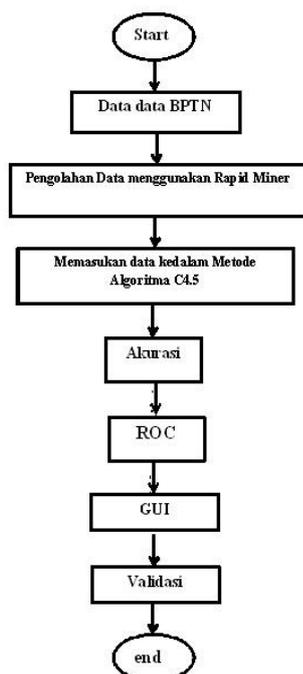
FP: Jumlah Data negatif yang diklasifikasikan dengan salah oleh komputer

FN: Jumlah Data positif yang diklasifikasikan dengan salah oleh komputer

Metode

Tahapan Penelitian

Pada tahap ini, dapat dijelaskan langkah-langkah penelitian yang akan dilakukan, dimulai pengumpulan data, mengolah data dan menguji data dengan algoritma yang akan diterapkan. Berikut tahapan penelitian dapat dilihat pada gambar 1 dibawah ini:



Gambar 1. Tahapan Penelitian

Tahapan penelitian di diatas hasil pengembangan penelitian dan dapat dijelaskan sebagai berikut :

1. Tahap awal dilakukan pengumpulan data penerima BPNT dan informasi terkait dengan judul penelitian ini berasal dari Desa Nagrak Utara yang bertujuan untuk menyediakan data mentah yang di olah terlebih dahulu, kemudian di bagi menjadi data latih dan data uji.
2. Data latih tersebut kemudian diuji coba dengan algoritma C4.5 pada Rapid Miner untuk mendapatkan model dan melihat hasil akurasi yang didapat.
3. Data latih di uji, dimana hasilnya dapat di validasi dari data tersebut.
4. Tahap ini data yang telah diolah menggunakan Rapid Miner dan Algoritma C4.5 akan menghasilkan keakuratan data prediksi yang berhak menerima BPNT kurva ROC yang menunjukkan hasil true positive rate & false positive rate.
5. Hasil dari algoritma tersebut diterapkan kedalam aplikasi yang akan dibangun.

Pada penelitian menggunakan metode eksperimen dalam uji coba data, dengan model eksperimen yang ada pada datamining yaitu CRISP-DM, dimana model ini memiliki 6 tahapan, yaitu [15]:

1. *Business/Research Understanding Phase*

Pada tahap ini dikumpulkan data penerima BPNT dalam 1 tahun terakhir dari Desa Nagrak Sukabumi yaitu sebanyak 130 data.

2. Fase Pemahaman Data

Pada tahap ini data yang berhasil dikumpulkan di pahami sampai dengan kriteria-kriteria yang akan digunakan, ada 9 kriteria sebagai atribut predictor dan 1 atribut hasil yang digunakan dengan mengacu kepada kriteria yang dikeluarkan KEMENSOS, didalam data yang sudah terkumpul terdapat data yang layak dan tidak layak menerima BPNT. 9 atribut yang menjadi parameter dapat dilihat pada tabel 2.

Table 2. Atribut Penerimaan BPNT

Atribut	Kategori
Luas rumah	>30 m
	<30 m
Penghasilan	>600000
	<600000
Jenis lantai	Keramik
Jenis dinding	Tembok
	Bambu

Sumber penerangan	Tembok Listik numpang Listik
Jenis air	Mata air Sumur
Membayar biaya pengobatan	Sanggup Tidak sanggup
Bahan bakar	Gas lpg Kayu bakar
Tabungan	< 200.000 > 200.000

3. Fase data persiapan

Berdasarkan data training yang diperoleh, maka langkah selanjutnya dilakukan teknik pre-processing, dimana teknik ini berguna untuk menghapus data yang selalu berubah, data tidak lengkap, mengkategorikan data, dan penghapusan data yang ganda.

4. Tahapan pemodelan

Pada tahap ini pembuatan model, dimana model yang akan digunakan yaitu model hasil dari algoritma C4.5 dengan kriteria yang digunakan yaitu *information gain*.

5. Tahap evaluasi

Tahap evaluasi merupakan tahap pengujian penilaian dengan menggunakan *Confusion Matrix* dan Kurva ROC, algoritma akan dapat dikategorikan berdasarkan range klasifikasi.

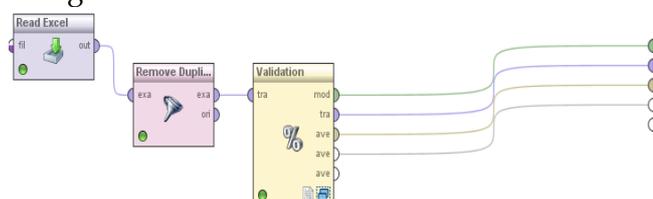
6. Tahap Deployment

Setelah rule dan model diketahui pada tahap terakhir ini diterapkan model algoritma C4.5 kedalam aplikasi yang akan di bangun.

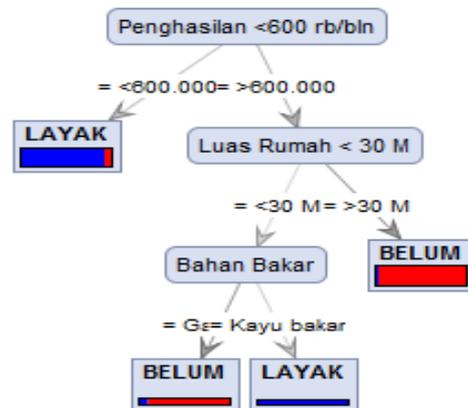
Hasil dan Pembahasan

Hasil Eksperimen dan Pengujian Model

Saat membuat pohon keputusan, pertama-tama menghitung jumlah kategori yaitu yang layak menerima dan yang belum layak menerima dan menghitung nilai entropi setiap kategori berdasarkan atribut yang ditentukan dari data latih. Terlihat dari jumlah data latih sebanyak 130 record dengan jumlah data yang layak menerima BPNT sebanyak 58 record dan 72 record yang belum layak menerima pada tahun 2018, mengolah data melalui Framework Rapid Miner dengan model algoritma C4.5, yaitu sebagai berikut:



Gambar 2. Algoritma C4.5 pada Framework Rapid Miner Sehingga didapat hasil model pohon keputusan seperti gambar sebagai berikut:



Gambar 3. Pohon Keputusan Penentuan Penerima BPNT

Pada hasil gambar 3 diatas, maka dapat diperoleh rule pohon keputusan seperti dibawah ini:

- Penghasilan <600 rb/bln = <600.000: LAYAK {LAYAK=51, BELUM=4}
- Penghasilan <600 rb/bln = >600.000
 - | Luas Rumah < 30 M = <30 M
 - | | Bahan Bakar = Gas LPG: BELUM {LAYAK=1, BELUM=8}
 - | | Bahan Bakar = Kayu bakar: LAYAK {LAYAK=3, BELUM=0}
 - | Luas Rumah < 30 M = >30 M: BELUM {LAYAK=3, BELUM=60}

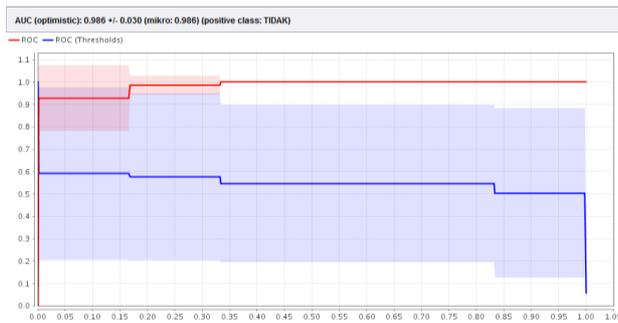
Pengujian Model

Hasil pengujian model bahwa nilai akurasi yang didapat yaitu sebesar 91.54%, dan dapat dilihat pada gambar 4 dari hasil pengolahan data menggunakan rapidminer:

accuracy: 91.54% +/- 7.26% (mikro: 91.54%)			
	true YA	true TIDAK	class precision
pred. YA	52	5	91.23%
pred. TIDAK	6	67	91.78%
class recall	89.66%	93.06%	

Gambar 4. Model Confusion Matrix

Dan model Kurva ROC yang dihasilkan pada penelitian ini yaitu menghasilkan nilai 0.986, yang berarti bawah algoritma C4.5 dengan kriteria Information Gain termasuk kedalam kategori klasifikasi Sangat Baik, dikarenakan memiliki nilai 0.90-1.00. Berikut gambar 5 Kurva ROC



Gambar 5. Kurva ROC

Deploment

Setelah model didapat, maka rule model diterapkan pada aplikasi berbasis desktop, aplikasi yang dibangun ini diharapkan dapat membantu dalam memprediksi penentuan penerimaan BPNT, berikut tampilan aplikasi yang dibuat:



Gambar 6. Hasil Tes Uji

Kesimpulan dan Saran

Dari hasil eksperimen menggunakan model algoritma C4.5 dengan kriteria Information Gain untuk menentukan penerimaan BPNT ini didapat nilai akurasi yaitu 91,54 % dan nilai AUC yaitu 0.986, hal ini terbukti bahwa penerapan Algoritma C4.5 dalam menentukan penerimaan BPNT sangat tepat digunakan untuk membantu pihak-pihak yang berkepentingan dalam pengambilan keputusan karena memiliki akurasi yang tinggi.

Adapun saran dalam penelitian selanjutnya yaitu:

1. Mengumpulkan data yang lebih banyak lagi dari beberapa tahun dan beberapa desa, sehingga data akan lebih besar dan hasil pengukuran yang akan didapat bisa lebih baik lagi dari penelitian sekarang.
2. Menggunakan algoritma optimasi untuk mengoptimalkan akurasi yang didapat

3. Aplikasi yang sudah dibuat dikembangkan ke aplikasi mobile, agar dapat digunakan oleh masyarakat luas

Referensi

- [1] A. Bastian and N. P. A. Bakhtiar, "Rancang Bangun Aplikasi Sistem Pendukung Keputusan Penentuan Masyarakat Miskin Menurut Kriteria Kemensos," in *Industrial Research Workshop and National Seminar (IRWNS)*, 2018, no. 1, pp. 280–287.
- [2] D. N. Wijaya, Z. Arifin, and S. Hadi, "Pengaruh Dana Desa, Alokasi Dana Desa dan Dana Alokasi Umum Terhadap Kemiskinan di Provinsi Jawa Timur Tahun 2015-2016," *J. Ilmu Ekon.*, vol. 2, no. 1, pp. 156–166, 2018.
- [3] S. Kawulur, R. A. Koleangan, and P. C. Wauran, "Analisa Pengaruh Pendapatan Asli Daerah Dan Dana Desa Dalam Menurunkan Tingkat Kemiskinan Di 11 Kabupaten Provinsi Sulawesi Utara," *J. Berk. Ilm. Efisiensi*, vol. 19, no. 03, pp. 107–117, 2019.
- [4] L. Van Gobel, "Efektivitas Pemberian Bantuan Sosial Kepada Kelompok Usaha Bersama Di Kecamatan Bolaang Mongondow Selatan," *J. Manaj. Sumber Daya Manusia, Adm. dan Pelayanan Publik*, vol. II, p. 98, 2015.
- [5] R. L. Fuady and A. M. Abadi, "Penentuan Penerimaan Bantuan Pangan Non Tunai (BPNT) Dengan Menggunakan Fuzzy Multiple Attribute Descission Making," *Semin. Mat. Dan Pendidik. Mat.*, pp. 203–210, 2017.
- [6] I. Kurniawan and R. A. Saputra, "Penerapan Algoritma C5.0 Pada Sistem Pendukung Keputusan KelayakanPenerimaan BerasMasyarakat Miskin," *J. Inform.*, vol. 4, no. 2, pp. 236–240, 2017.
- [7] A. Andreas, A. Wiryadinata, and H. Agung, "Penerapan Algoritma Simple Additive Weighting untuk Membantu Dalam Menentukan Calon Penerima Bantuan Pangan Non Tunai," *Kalbiscentia*, vol. 6, no. 1, pp. 37–43, 2019.

- [8] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS*, vol. VI, no. 1, pp. 15-16, 2014, doi: 10.1109/EPEPEMC.2016.7752007.
- [9] F. F. Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *Tek. Inform. Dan Sist. Inf.*, vol. 3, no. 2, pp. 95-103, 2017.
- [10] A. R. Febie Elfaladonna, "Analisa Metode Classification-Decission Tree Dan Algoritma," *Sci. Inf. Technol.*, vol. 2, no. 1, pp. 10-17, 2019.
- [11] M. Ridwan, "Sistem Rekomendasi Proses Kelulusan Mahasiswaberbasis Algoritma Klasifikasi C4.5," *J. Ilm. Inform.*, vol. 2, no. 1, pp. 105-111, 2017, doi: 10.35316/jimi.v2i1.460.
- [12] W. F. Pattipeilohy, A. Wibowo, and D. R. Utari, "Pemodelan Dan Prototipe Sistem Informasi Untuk Prediksi Pembaharuan Polis Asuransi Mobil Menggunakan Algoritma C.45," *Pros. SNATIF ke-6 Tahun 2019*, no. 2007, pp. 96-101, 2019.
- [13] F. Gorunescu, *Data Mining "Concepts, Models and Techniques"*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [14] S. Phadikar, "Classification of Rice Leaf Diseases Based onMorphological Changes," *Int. J. Inf. Electron. Eng.*, vol. 2, no. 3, pp. 460-463, 2012, doi: 10.7763/ijiee.2012.v2.137.
- [15] R. A. Saputra, L. S. Ramdhani, and S. Supriatman, "Penerapan Metode Iterative Dichotomizer 3 (Id 3) Untuk Menentukan Beasiswa Berprestasi Pada Smp Pgri Caringin Sukabumi," *J. Pilar Nusa Mandiri*, vol. 15, no. 1, p. 36, 2019.