

Penerapan Teknik Data Mining Dengan Algoritma C4.5 Dalam Mengklasifikasikan Kelayakan Penerima Bantuan Sosial Di Desa Batuah

Wiwit Wahyudi

Universitas STEKOM, Jl.Majapahit 605 Semarang, Jawa Tengah, Indonesia

E-Mail: wiwit112@stekom.ac.id

Abstrak– Penyaluran bantuan sosial yang tepat sasaran merupakan tantangan krusial bagi Pemerintah Desa Batuah dalam upaya pengentasan kemiskinan. Masalah utama yang sering muncul adalah subjektivitas dalam proses seleksi yang menyebabkan ketidakakuratan data penerima. Penelitian ini bertujuan untuk menerapkan teknik data mining menggunakan algoritma C4.5 untuk mengklasifikasikan kelayakan penerima bantuan sosial berdasarkan kriteria seperti pendapatan, kondisi hunian, status pekerjaan, dan jumlah tanggungan. Metode penelitian dilakukan melalui tahapan Knowledge Discovery in Databases (KDD), yang meliputi pembersihan data, transformasi, hingga pemodelan menggunakan pohon keputusan. Data yang digunakan terdiri dari catatan historis penerima bantuan di Desa Batuah yang kemudian diuji menggunakan metode split data atau cross-validation. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu menghasilkan model klasifikasi dengan tingkat akurasi sebesar 87,50%. Berdasarkan pengujian menggunakan Confusion Matrix, diperoleh nilai precision sebesar 85,20% dan recall sebesar 90,15%. Model ini menghasilkan sejumlah aturan keputusan (rule) yang dapat digunakan oleh perangkat desa sebagai instrumen pendukung keputusan yang objektif. Dengan tingkat kesalahan yang rendah, penerapan algoritma ini terbukti efektif dalam meminimalisir risiko salah sasaran dalam distribusi bantuan sosial di Desa Batuah.

Keywords: Data Mining, Algoritma C4.5, Akurasi, Bantuan Sosial, Desa Batuah

Abstract– The distribution of targeted social assistance is a crucial challenge for the Batuah Village Government in efforts to alleviate poverty. The main problem that often arises is subjectivity in the selection process which leads to inaccuracies in the recipient data. This study aims to apply a data mining technique using the C4.5 algorithm to classify the eligibility of social assistance recipients based on criteria such as income, housing conditions, employment status, and number of dependents. The research method was carried out through the Knowledge Discovery in Databases (KDD) stage, which included data cleaning, transformation, and modeling using a decision tree. The data used consisted of historical records of aid recipients in Batuah Village which were then tested using the split data or cross-validation method. The results showed that the C4.5 algorithm was able to produce a classification model with an accuracy rate of 87.50%. Based on testing using the Confusion Matrix, a precision value of 85.20% and a recall of 90.15% were obtained. This model produces a number of decision rules that can be used by village officials as an objective decision support instrument. With a low error rate, the application of this algorithm has proven to be effective in minimizing the risk of mistargeting in the distribution of social assistance in Batuah Village.

Keywords: Data Mining, C4.5 Algorithm, Accuracy, Social Assistance, Batuah Village

1. PENDAHULUAN

Desa batuah ialah salah satu desa yang terletak di kecamatan seranau kabupaten kotawaringin timur. Desa ini mempunyai sebagian Rt serta Rw, serta cuma berjumlah 100 lebih jiwa yang masih tinggal di desa ini.



Desa batuah ini diharuskan buat mengikuti dalam program dari departemen sosial buat mengatasi aspek ekonomi yang susah serta mensejahterakan penduduk antara lain program dana desa (sembako), program ini ditujukan untuk kerabat yang miskin selaku upaya pemerintah dalam mengatasi kemiskinan terutama untuk desa batuah ini.

Kasus yang muncul yakni gimana barisan wilayah menetapkan kepala keluarga desa batuah yang memang betul- betul berhak menerima dorongan tersebut, mengenai ini biasanya hanya dicoba secara manual bersumber pada data yang telah terdapat dilapangan serta terkadang kerap memunculkan kecemburuan sosial, sebab yang menerima dorongan masih terdapat unsur- faktor keakraban ketetanggan dengan SLS setempat [1].

Buat mengatasi masalah yang ada hingga dibesarkan sesuatu pemecahan lewat suatu riset memakai Algoritma C4. 5 sebab algoritma C4. 5 mempunyai kelebihan ialah bisa mencerna informasi diskrit serta informasi numerik, dalam riset ini dengan prosedur algoritma C4. 5 buat melaksanakan penggalan data (data yang dimaksudkan buat mengetahui masyarakat yang benar-benar berhak menerima dorongan). eksploitasi data bersumber pada pengumpulan informasi dari sebagian Rt serta Rw memakai rapid miner C4. 5 serta decision tree buat mengenali jumlah masyarakat menerima atau tidaknya bantuan yang di selenggarakan di desa batuah tersebut, karena karena sekarang marak sekali adanya faktor kecurangan atau keakraban kekeluargaan.

Yang jadi pusat kepedulian nya buat desa batuah ini ialah aspek ekonomi yang lumayan memperhantikan, oleh sebab itu di desa batuah ini perlunya uluran ataupun dorongan pemerintah buat meringani beban warga batuah tersebut.

Program dorongan sosial (sembako) ialah sesuatu program penanggulangan kemiskinan. Buat Warga yang kurang sanggup semacam desa batuah, serta bansos ialah bagian dari program- program penanggulangan kemiskinan yang lain. Perogram Dorongan sosial ini merupakan sesuatu program yang membagikan dorongan sembako kepada rumah tangga yang miskin ataupun yang kurang sanggup.

Dengan program dana desa lanjutan yang masih hendak berlangsung sampai tahun (2022), hingga diharapkan perihal itu dapat memacu proses pemulihan ekonomi terutamanya buat desa- desa terpencil [2].

Di lain sisi, program tersebut pula diharapkan sanggup kurangi akibat keterpurukan akibat beban ekonomi yang lumayan menyulitkan warga setempat.

Decision Tree pada riset terdahulu ialah digunakan selaku klasifikasi buat menunjang dalam permasalahan pendataan informasi yang di maksud ialah semacam informasi attribut berbentuk nama, kategori kelamin, penghasilan, serta lain- lainnya.

Bersumber pada hasil informasi mining memakai prosedur klasifikasi decision tree yang di peroleh nilai akurasi ialah 93. 00% serta menciptakan AUC 0. 834 Dengan pengujian Rapid minner Campuran prosedur dicoba sebagian riset terdahulu memakai bermacam algoritma yang berbeda- beda, disini periset melaksanakan pengujian informasi penerima bansos berbentuk sembako yang akurat memakai decision tree. Predeksi hasil accuracy merupakan 93. 00% dengan menciptakan akurasi lumayan besar.

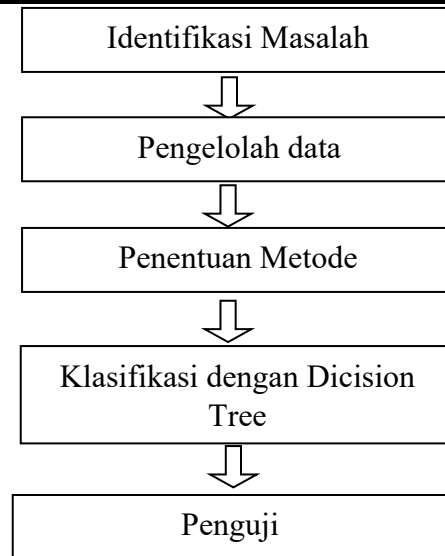
Dengan ini riset menunjang dalam pendataan anggota- anggota keluarga yang layak memperoleh dorongan ataupun tidaknya, dengan memanfaatkan software- software yang terdapat buat memudahkan dalam permasalahan pendataan pada desa batuah tersebut[3].

Tujuan Riset merupakan buat membongkar masalah- masalah yang terdapat dalam pendataan serta hasil akurasi yang diperoleh.

2. METODE PENELITIAN

Pada studi dibutuhkan perencanaan dan langkah- langkah yang terstruktur biar studi dapat berjalan dengan baik.





Gambar 1. Tahapan Penelitian

a. Identifikasi Masalah

Riset ini diawali dengan mengumpulkan informasi lewat wawancara dengan warga desa batuah tentang ini di butuhkan buat mengenali data serta kasus yang hendak dituntaskan, kasus utama pada riset ini merupakan gimana membangun system yang terdapat bisa mengklasifikasikan kepala keluarga yang layak menerima dorongan sembako.

b. Pengelolaan Data

Pengelolaan informasi pada riset ini hendak mencakup seluruh aktivitas buat mempersiapkan informasi saat sebelum masuk proses pemodelan di aplikasi rapidminer. Dalam melaksanakan pengolahan informasi dini, hendak dicoba sebagian tahapan supaya pada kesimpulannya hendak didapatkan informasi yang dapat digunakan pada sesi selanjutnya. Setelah itu dicoba pembersihan informasi, melenyapkan menghapus informasi yang tidak lengkap (outer/ noise) buat membagikan hasil informasi yang bermutu serta informatif. Tahapan tersebut antara lain: cleaning, reduction, transformation dan hendak dicoba split validation[4].

c. Penentuan Metode

Penentuan Prosedur Pada sesi ini menarangkan tentang prosedur yang diseleksi serta hendak digunakan pada riset ini ialah prosedur Algoritma C4. 5.

d. Klasifikasi dengan decision tree

Ialah melaksanakan pengelolah informai memakai prosedur decision tree buat mengenali nilai akurasinya.

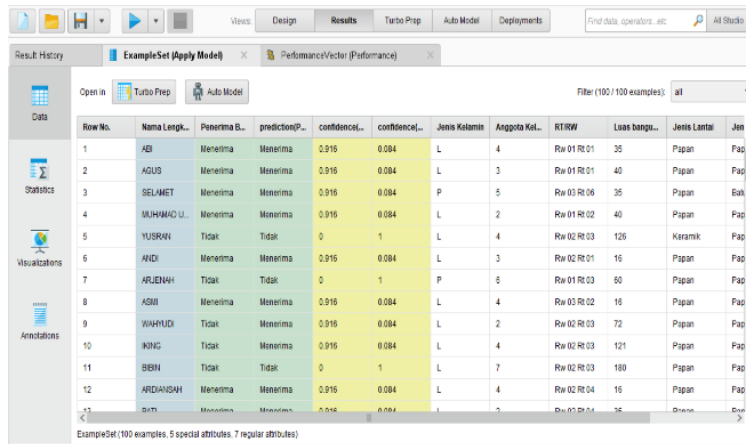
e. Pengujian

Proses pengujian pada studi ini memakai confusion matrix, yang hendak menghitung nilai precision, recall, serta accuracy. Confusion matrix terdiri dari true positive, false positive, true negative serta false negative buat menghitung presisi, recall serta akurasi.

Precision ialah tingkatan ketepatan antara data yang dimohon oleh pengguna dengan jawaban yang diberikan oleh keberhasilan sistem dalam menciptakan kembali suatu data.

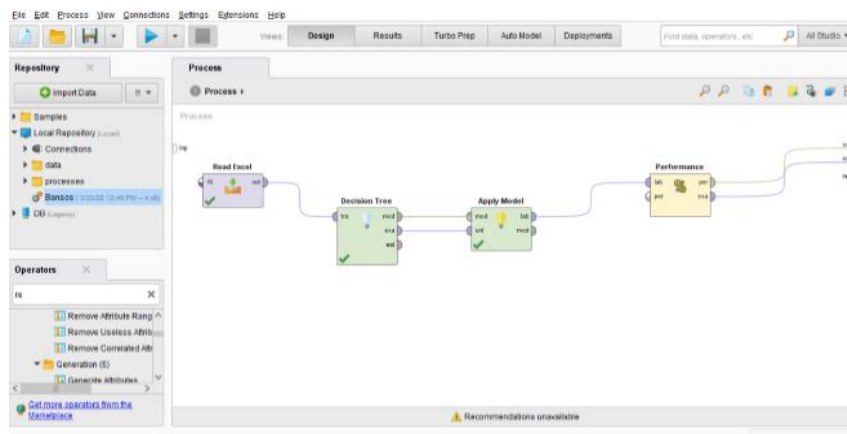
3. HASIL DAN PEMBAHASAN

Hasil pengukuran informasi accuracy yang diperoleh dari informasi testing menggapai 93. 00%. Dari informasi tersebut dikenal prediksi Dapat dengan true Dapat menggapai 80 penduduk serta true Tidak Dapat sebanyak 7 penduduk, dengan pencapaian class precision 100. 00%. Sebaliknya buat prediksi Tidak Dapat dengan true Dapat sebanyak 0 penduduk serta buat true Tidak Dapat menggapai 13 penduduk dengan pencapaian class precision 100%. Buat class recall dengan true Dapat menggapai 100% sebaliknya buat class recall dengan true tidak Dapat menggapai 100. 00%. Hasil analisa informasi testing dengan informasi training pada rapidminer bisa dilihat pada foto di dasar ini.



Row No.	Nama Lengkap	Penerima B...	prediction(P...	confidence...	confidence...	Jenis Kelamin	Anggota Kat...	RTRW	Luas bang...	Jenis Lantai	Jen
1	ABI	Menerima	Menerima	0.916	0.084	L	4	Rw 01 Rr 01	35	Papan	Fap
2	AGUS	Menerima	Menerima	0.916	0.084	L	3	Rw 01 Rr 01	40	Papan	Fap
3	SELAMET	Menerima	Menerima	0.916	0.084	P	5	Rw 03 Rr 06	35	Papan	Esk
4	MUHAMAD U.	Menerima	Menerima	0.916	0.084	L	2	Rw 01 Rr 02	40	Papan	Fap
5	YUSRIAN	Tidak	Tidak	0	1	L	4	Rw 02 Rr 03	126	Keramik	Fap
6	ANDI	Menerima	Menerima	0.916	0.084	L	3	Rw 02 Rr 01	16	Papan	Fap
7	ARJENAH	Tidak	Tidak	0	1	P	6	Rw 01 Rr 03	60	Papan	Fap
8	ASMI	Menerima	Menerima	0.916	0.084	L	4	Rw 03 Rr 02	16	Papan	Fap
9	WAHYUDI	Tidak	Menerima	0.916	0.084	L	2	Rw 02 Rr 03	72	Papan	Fap
10	KING	Tidak	Menerima	0.916	0.084	L	4	Rw 02 Rr 03	121	Papan	Fap
11	SENI	Tidak	Tidak	0	1	L	7	Rw 02 Rr 03	180	Papan	Fap
12	ARDIANSAH	Menerima	Menerima	0.916	0.084	L	4	Rw 02 Rr 04	16	Papan	Fap
13	DITI	Menerima	Menerima	0.916	0.084	L	5	Rw 03 Rr 01	36	Stone	Fap

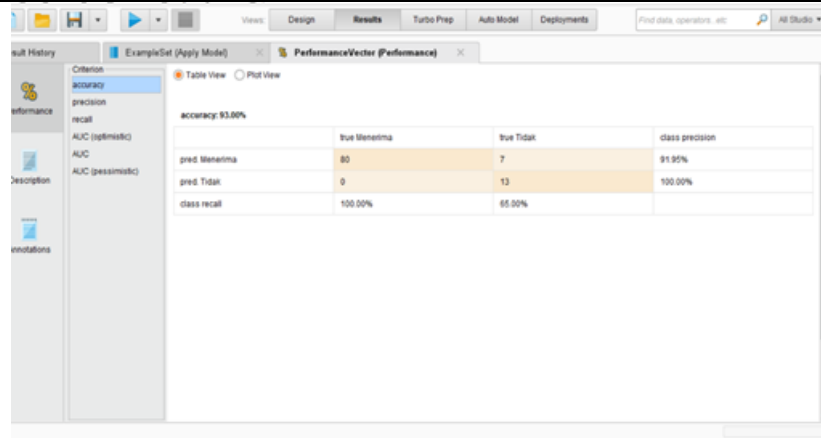
Gambar 2. Hasil analisa informasi testing



Gambar 3. Hasil analisa informasi testing 2

Proses informasi Read excel serta Memakai decision tree ialah melaksanakan analisis bermacam model serta memilah model performance buat menciptakan nilai akurasi prediksi yang baik. Pada foto ialah proses model, sehabis pembacaan file informasi, blok read excel setelah itu dihubungkan dengan decision tree. Proses informasi training serta informasi testing terletak di dalam blok proses decision tree. Dari proses decision tree tersebut dicoba pengolahan informasi pada rapidminer dengan main process[5].

Confusion matrix confusion matrix merupakan sesuatu prosedur yang umumnya digunakan buat melaksanakan perhitungan akurasi pada konsep informasi mining. Penilaian memakai confusion matrix ini menciptakan 3 keluaran. Ialah accuracy, precision serta recall.



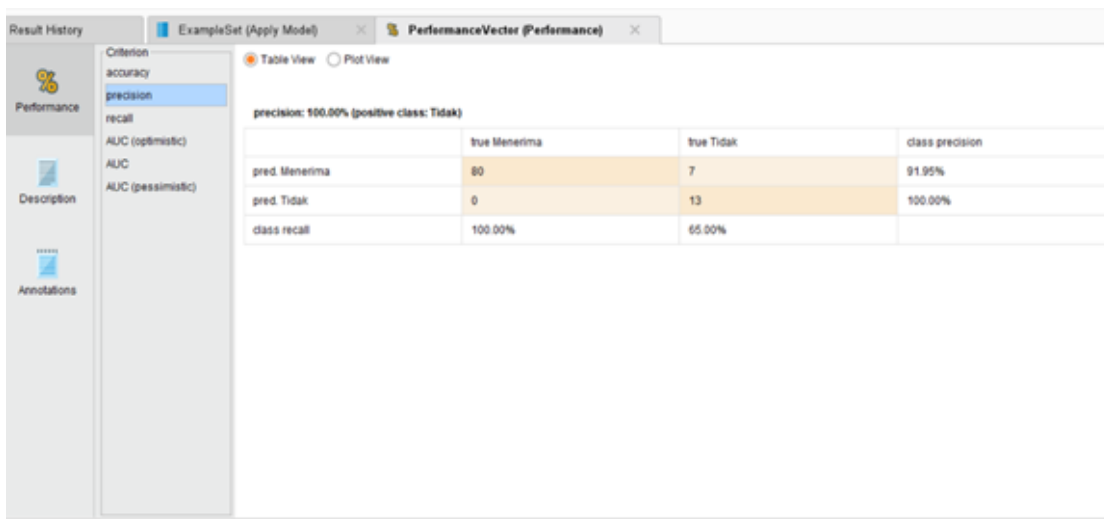
The screenshot shows the PerformanceVector (Performance) window in RapidMiner. The 'Table View' is selected, displaying a confusion matrix and performance metrics. The overall accuracy is 93.00%.

	true Menerima	true Tidak	class precision
pred Menerima	80	7	91.95%
pred Tidak	0	13	100.00%
class recall	100.00%	65.00%	

Gambar 4. Tingkat Acuracy

Hasil tingkatan akurasi serta nilai AUC dapat kita amati serta tahu bersama dengan meng-klik tab PerformanceVector (Performance), sehingga bakal kelihatan hasil akurasinya semacam foto diatas. Dimana model yang sudah dibangun diuji tingkatan akurasinya dengan memasukan ataupun uji yang berasal dari informasi training dengan memakai performance pada software rapidminer studio 9. 10 buat menguji tingkatan accuracy.

Hasil pengukuran informasi accuracy yang diperoleh dari informasi training menggapai 93. 00% buat nilai class recall serta nilai class precision dipaparkan kalau jenis “Menerima” menciptakan class recall sebesar 93. 00% serta class precision sebesar 91.95% dan kategori “Tidak menerima” true tidak sebanyak 7, dengan pencapaian class precision 100.00%, menghasilkan class recall sebesar true menerima 100.00% dan untuk true tidak 65.00%.

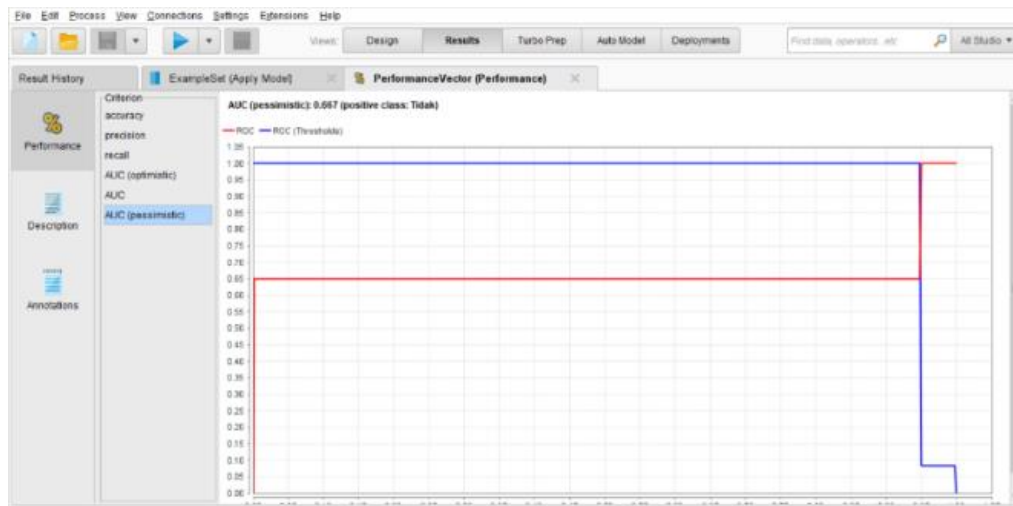


The screenshot shows the PerformanceVector (Performance) window in RapidMiner. The 'Table View' is selected, displaying a confusion matrix and performance metrics. The overall precision is 100.00% for the positive class (Tidak).

	true Menerima	true Tidak	class precision
pred Menerima	80	7	91.95%
pred Tidak	0	13	100.00%
class recall	100.00%	65.00%	

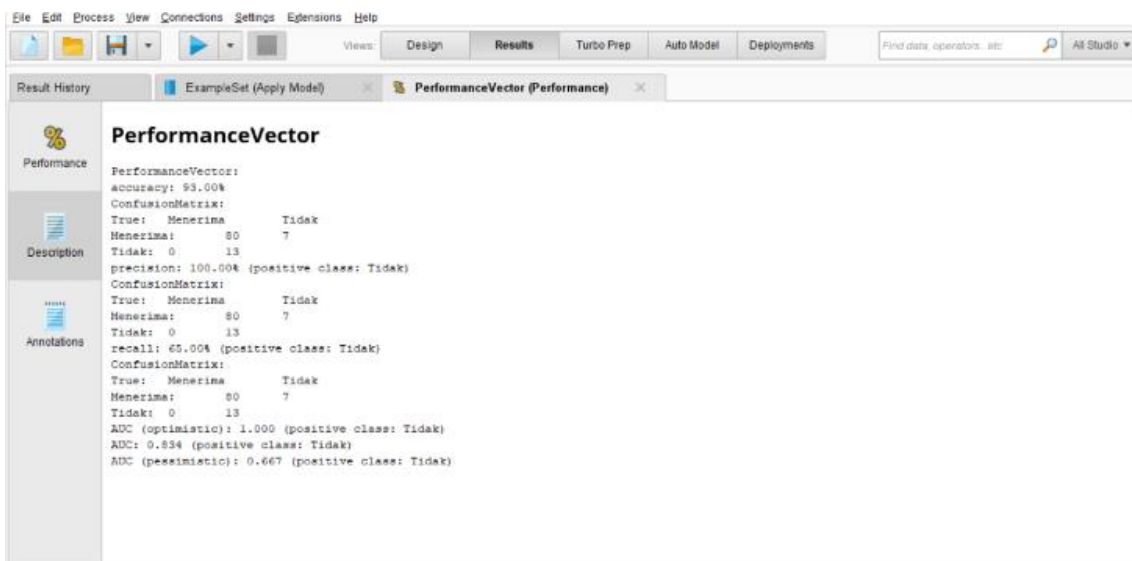
Gambar 5. Hasil pengukuran informasi accuracy

Hasil pengukuran informasi precision yang diperoleh dari informasi training menggapai 100. 00% dari informasi tersebut dikenal prediksi menerima dengan true menerima menggapai 80 serta true Tidak menerima sebanyak 7, dengan pencapaian class precision 91. 95%. Sebaliknya buat prediksi Tidak menerima dengan true menerima sebanyak 0 serta buat true Tidak berhak menggapai 13 dengan pencapaian class precision 100. 00%. Buat class recall dengan true Berhak menggapai 100. 00% sebaliknya buat class recall dengan true Tidak Berhak menggapai 65. 00%.



Gambar 6. Hasil Pengukuran

Serta buat model Kurva ROC yang dihasilkan pada studi ini ialah menciptakan tingkat 0.667, yang berarti dasar algoritma C4.5 dengan kriteria Information Gain tercantum kedalam jenis klasifikasi lumayan efektif, disebabkan mempunyai tingkat 0.65- 1.00



Gambar 7. Information Gain

Bersumber pada pengolahan informasi memakai perangkat lunak rapidminer didapat tingkat accuarsy sistem sebesar 93.00% artinya kalau rule yang dihasilkan tingkatan kebenaran nyaris mendekati 100%. Dimana model yang sudah dibangun diuji tingkatan akurasi dengan memasukan ataupun uji yang berasal dari informasi training dengan memakai performance pada aplikasi rapidminer 9.10 buat pengujian nilai akurasi.

3.2. Rumus Algoritma C4.5

Algoritma C4. 5 ialah algoritma yang digunakan buat membentuk tumbuhan keputusan. Tumbuhan keputusan ialah prosedur klasifikasi serta prediksi yang sangat kokoh serta populer. Prosedur tumbuhan keputusan mengganti kenyataan yang sangat besar jadi tumbuhan keputusan yang merepresentasikan ketentuan. Ketentuan bisa dengan gampang dimengerti dengan bahasa natural [3] di jelaskan kalau algoritma C4. 5 ialah salah satu algoritma yang digunakan buat membentuk decision tree bersumber pada training informasi.

Dalam perhitungan algoritma C4. 5 mempunyai 2 sesi ialah pertama menghitung nilai Entropy serta yang kedua Gain. Nilai entropy berperan buat mengukur jumlah dari data yang terdapat pada atribut dengan persamaan berikut[5]:

$$Entrop(S)=-\sum_{i=1}^n P_i \cdot \log_2 P_i$$

Sebaliknya nilai Gain bisa digunakan dalam mengukur daya guna atribut dalam proses klasifikasi dengan persamaan selaku berikut:

$$Gai(S.A)=Entro(S)-\sum_{i=1}^n S_i \cdot Entropy(S_i)$$

4. KESIMPULAN

Pada riset ini melaksanakan prediksi klasifikasi penentuan penerima dorongan bantuan sembako dengan memakai algoritma C4. 5, algoritma C4. 5 ialah salah satu prosedur yang bisa digunakan buat mengklasifikasikan informasi. Decesion tree ialah pengklasifikasian statistik yang bisa digunakan buat memprediski probabilitas keanggotaan sesuatu class. Algoritma C4.5 mempunyai guna buat menciptakan pengetahuan ataupun pola- pola kesamaan ciri dalam sesuatu kelompok ataupun kelas tertentu. Prediksi tingkatan penerimaan dorongan sembako yang digunakan ada 2 kelas, ialah layak serta tidak layak. Informasi yang digunakan buat prediksi ialah informasi yang diambil dari ilustrasi informasi masyarakat di desa batuah. Dari hasil penilaian mengenai performance didapatkan akurasi yang dihasilkan buat 100 informasi training yang digunakan menciptakan akurasi sebesar 93. 00%, recall 100. 00%, dan precision 91. 95%. Akurasi bisa dipengaruhi oleh sebagian aspek, antara lain: jumlah informasi training, informasi testing serta atribut yang digunakan.

REFERENSI

- [1] A. F. Vista Anestiviya1, in maret). ANALISIS POLA MENGGUNAKAN METODE C4.5 UNTUK PEMINATAN JURUSAN SISWA BERDASARKAN KURIKULUM (STUDI KASUS : SMAN 1 NATAR). Jurnal Teknologi dan Sistem Informasi (JTSI, vol. 2, 2021, pp. 80–85.
- [2] S. D. Ringga Sentagi Asa1, “IDENTIFIKASI PENYALURAN ZAKAT MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS DI BAZNAS KABUPATEN AGAM,,” J. Sains Dan Inform., vol. 5, pp. 44–53, 2019.
- [3] E. E, “PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA PENERIMA BANTUAN SOSIAL COVID-19,,” in Jurnal Ilmu Komputer dan Sistem Informasi, XXXX, pp. 1–8.
- [4] L. Swastina, “Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa,,” J. GEMA Aktual., vol. 2, pp. 93–98, Jun. 2013.
- [5] E. Fitriani, “PERBANDINGAN ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN,,” J. Sist. Inf., vol. 9, pp. 103–115, 2020.