

KAJIAN ALGORITMA C4.5 PADA SISTEM PENDUKUNG KEPUTUSAN DALAM PEMBERIAN DANA BANTUAN SISWA MISKIN (BSM)

Arif Kurniawan¹⁾ Sularso Budilaksono²⁾

Universitas Muhammadiyah Tangerang / Fakultas Teknik,
Program Studi Informatika

Jl. Perintis Kemerdekaan 1/33 Cikokol Kota Tangerang TLP. 55793251, 55772949, 55793802,
55736926

Email :arif.afk@bsi.ac.id

²⁾Email : sularso2007@gmail.com

ABSTRAK

Menurut Peraturan nomor 69 tahun 2009, biaya operasi standar nonpersonalia adalah standar biaya yang diperlukan untuk membiayai operasi nonpersonalia selama 1 (satu) tahun sebagai bagian dari satuan pendidikan dana pendidikan secara keseluruhan untuk melakukan kegiatan secara teratur dan melanjutkan pendidikan Nasional Tepat Guna Standar pendidikan. Bantuan Operasional Sekolah (BOS) adalah program pemerintah yang pada dasarnya adalah penyediaan dana untuk biaya operasional untuk unit nonpersonalia melaksanakan pendidikan dasar sebagai pendidikan wajib. Salah satu tujuannya adalah untuk menyediakan pendidikan gratis bagi siswa miskin dan meringankan siswa lain, sehingga mereka memperoleh layanan pendidikan dasar yang lebih memenuhi syarat untuk lulus. Meskipun dana operasional sekolah (BOS) diharapkan dapat meningkatkan jumlah siswa yang berpartisipasi, tetapi masih banyak anak - anak yang tidak bisa sekolah, putus sekolah dan tidak dapat melanjutkan Ulasan pendidikan mereka ke tingkat berikutnya pendidikan. Salah satu penyebab ini adalah sulitnya orang tua / keluarga dalam memenuhi kebutuhan pendidikan seperti seragam, buku, sepatu, biaya transportasi atau biaya pendidikan lainnya yang tidak tercakup oleh BOS. Ini adalah apa yang ada di balik pengembangan Program Bantuan Siswa Miskin (BSM). Tujuan dari penelitian ini adalah untuk melakukan studi dengan menggunakan algoritma C4.5 mendukung data dari data pada pembiayaan sekolah BSM MTs Attaqwa Kota Tangerang. Dari hasil pengujian yang dilakukan model atau metode yang digunakan dalam algoritma sistem ini pengambil keputusan C4.5 Muncul untuk memiliki akurasi 0.500 dan dengan bantuan alat pendukung Cepat Miner Studio v.6.0.002 untuk pengolahan data. Dengan penelitian ini diharapkan dapat memfasilitasi dan membantu sekolah dalam membuat keputusan dengan cepat intervensi penerima BSM, tepat dan akurat.

Keywords: C4.5 algorithm, Decision Tree, Poor Students Aid Fund, Rapid Miner.

ABSTRACT

According to Regulation number 69 of 2009, operating expenses nonpersonalia standard is the standard cost required to fund operations nonpersonalia for 1 (one) year as part of the overall education funding education units in order to perform activities on a regular basis and continuing education appropriate National Education Standards. School Operational Assistance (BOS) is a government program that is basically the provision of funding for operating costs for units nonpersonalia implementing basic education as compulsory education. One aim is to provide free education for poor students and relieve the other students, so that they acquire basic education services are more qualified to graduate. Although the school operational funds (BOS) is expected to increase the number of participating students, but there are still many children - children who can not go to school, drop out of school and can not continue their education to the next level of education. One cause of this is the difficulty of parents / families in meeting the educational needs such as uniforms, books, shoes, transportation costs or other educational expenses that are not covered by the BOS. This is what lies behind the development of Poor Students Assistance Program (BSM). The purpose of this study was to conduct a study using the C4.5 algorithm supporting data from the data on school funding BSM MTs Attaqwa Tangerang City. From the results of tests performed models or methods used in this system algorithm C4.5 decision makers appeared to have an accuracy of 0.500 and with the help of supporting tools Rapid Miner Studio v.6.0.002 for data processing. With the study is expected to facilitate and assist the school in making decisions BSM recipients quickly, precisely and accurately.

Keywords: C4.5 algorithm, Decision Tree, Poor Students Aid Fund, Rapid Miner.

1. PENDAHULUAN

Latar Belakang Masalah

Dengan disetujuinya APBN-Perubahan tahun 2013, dan adanya kebijakan kenaikan harga bahan bakar minyak (BBM) bersubsidi yang diberlakukan mulai bulan Juni 2013, maka Pemerintah menetapkan program-program kompensasi terhadap masyarakat miskin dan rentan kemiskinan. Program kompensasi tersebut berupa Bantuan Siswa Miskin (BSM), Program Keluarga Harapan (PKH), program Raskin, Bantuan Langsung Sementara Masyarakat (BLSM), dan program infrastruktur dasar.

Dengan program Bantuan Siswa Miskin ini, diharapkan dapat membantu seluruh siswa yang berasal dari keluarga kurang mampu/miskin di lingkungan sekolah dalam membiayai sebagian kebutuhan pendidikannya, sehingga dapat menyelesaikan pendidikannya, bahkan dapat melanjutkan pendidikan ke jenjang berikutnya. Ada beberapa faktor yang menjadi pertimbangan pihak komite sekolah ketika akan memilih siswa yang akan dipilih sebagai penerima BSM. Faktor yang utama adalah mengenai kondisi keuangan orang tua sehingga dapat dikatakan miskin yang dibuktikan dengan KM (Kartu Multiguna). Dan faktor-faktor yang lain seperti prestasi.

Dari pengamatan penulis selama melakukan riset di MTs Attaqwa Blendung-Tangerang, penulis menganalisa kurang efisiennya proses pengambilan keputusan dalam menentukan siswa yang layak mendapatkan BSM dikarenakan sistem pengambilan keputusan yang berjalan sekarang masih dilakukan secara manual. Dari masalah yang berhasil penulis analisa, penulis dapat menyimpulkan bahwa metode *Algoritma C4.5* dan pohon keputusan dapat digunakan sebagai sistem pendukung keputusan dalam menentukan pemberian Dana BSM kepada siswa dengan menerapkan kriteria-kriteria penilaian yang jelas.

Ruang Lingkup Penelitian

Supaya masalah dalam penelitian ini tidak melebar, maka diperlukan batasan masalah. Untuk membatasi ruang lingkup penelitian, maka dalam penelitian ini hanya akan dibahas tentang :

1. Metode yang digunakan dalam penelitian sistem ini adalah *Algoritma C4.5* dan Pohon keputusan.
2. Sistem Pendukung Keputusan yang akan penulis buat merupakan sebagai alat bantu bagi pihak Sekolah, untuk menentukan

siswa yang layak menerima bantuan dana BSM. Berdasarkan kriteria-kriteria yang telah ditentukan oleh pihak sekolah .

Perumusan Masalah

Sejauh mana tingkat akurasi *Algoritma C4.5* dapat digunakan untuk memprediksi siswa yang berhak atau layak mendapatkan bantuan dana (BSM) di MTs Attaqwa Blendung-Tangerang?

Tujuan Penelitian

Tujuan penelitian ini adalah untuk menerapkan *Algoritma C4.5* dan Pohon Keputusan dalam menyelesaikan permasalahan pemberian dana BSM yang kurang tepat atau salah sasaran. Dengan adanya metode tersebut diharapkan dapat diterapkan sebagai sistem pendukung keputusan bagi para pengambil keputusan dalam hal ini adalah pihak sekolah serta pihak-pihak yang terkait.

2. LANDASAN TEORI DAN KERANGKA PEMIKIRAN

Faktor-Faktor Yang Mempengaruhi Pengambilan Keputusan

Dalam mengambil sebuah keputusan, terdapat dua kondisi yang mungkin akan dihadapi ketika keputusan tersebut telah dilaksanakan, yaitu keputusan dalam kondisi yang pasti dan keputusan dalam kondisi tidak pasti. Dalam kondisi yang pasti, terdapat tiga hal yang akan berpengaruh terhadap pengambilan keputusan (Prawirosentono, 2002), yaitu :

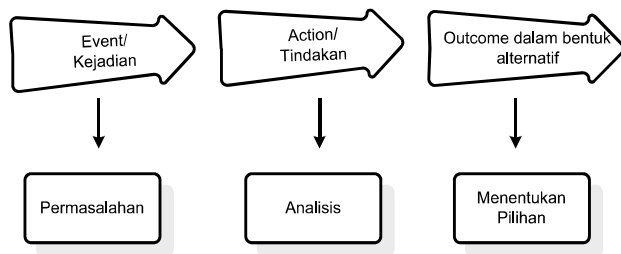
- a. *Available alternatives*: adalah suatu kondisi di mana pengambil kesimpulan dihadapkan pada beberapa alternatif. Keputusan dibuat dengan memilih salah satu pilihan yang menguntungkan institusi atau organisasi.
- b. *State of Nature*: adalah kondisi di luar kekuasaan pengambil keputusan.
- c. *Payoff*: merupakan kombinasi antara alternatif-alternatif keputusan yang ada dengan *State of Nature*

Sedangkan untuk kondisi yang tidak pasti, juga terdapat tiga hal yang dapat mempengaruhi pengambilan keputusan. Ketiga hal tersebut adalah (Prawirosentono, 2002) :

- a. *Event*: adalah kejadian atau permasalahan yang harus diselesaikan dengan cara strategis tertentu, karena kejadian atau permasalahan tersebut akan menjadi masukan dalam proses pengambilan keputusan. Kejadian atau permasalahan yang menjadi masukan dalam proses pengambilan keputusan tersebut terbagi

dalam empat jenis, yaitu: permasalahan direktif, permasalahan strategis, permasalahan taktis dan permasalahan operasional.

- b. **Action:** adalah suatu tindakan dalam menyelesaikan permasalahan dengan langkah-langkah strategis tertentu. Langkah-langkah yang diambil tersebut merupakan suatu proses pengambilan keputusan dari beberapa alternatif langkah-langkah strategis dalam menyelesaikan permasalahan.
- c. **Outcome:** adalah hasil dari suatu langkah pengambilan keputusan dalam menyelesaikan permasalahan. Hasil dari outcome ini dapat berupa keuntungan, impas atau rugi.



Sumber : Prawirosentono, 2002
Gambar 2.1 Skema Pengambilan Keputusan dalam kondisi ketidakpastian

Jenis-Jenis Pengambilan Keputusan

Menurut *Simon* dalam *McLeod* (2004), membagi keputusan menjadi dua jenis, yaitu:

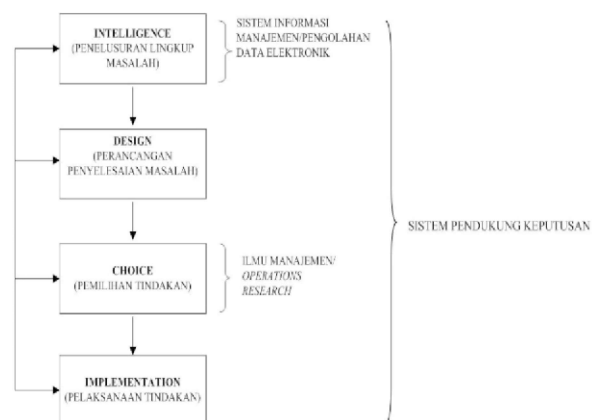
- a. Keputusan Tak Terprogram: adalah keputusan yang bersifat "baru", tidak terstruktur, dan jarang konsekuensi. Tidak ada metode yang pasti untuk menangani masalah ini karena belum pernah ada sebelumnya, atau karena sifat dan struktur persisnya tidak terlihat atau rumit, atau karena begitu pentingnya sehingga memerlukan perlakuan yang sangat khusus.
- b. Keputusan Terprogram: adalah keputusan yang bersifat berulang dan rutin sampai pada batas hingga suatu prosedur pasti telah dibuat untuk menanganinya sehingga keputusan tersebut tidak perlu diperlakukan *de novo* (sebagai sesuatu yang baru).

Tahapan Sistem Pendukung Keputusan

Simon dalam *McLeod* (2004), menjelaskan empat tahapan yang harus dilalui manajer dalam memecahkan sebuah masalah. Tahapan sistem pendukung keputusan menurut *Simon* adalah:

- a. Pertama adalah tahap **intelligence** yaitu tahap di mana permasalahan mulai ditemukan untuk selanjutnya diidentifikasi.

- b. Kedua adalah tahap **design** yaitu tahap di mana perancangan sudah mulai dilakukan berdasarkan permasalahan yang diidentifikasi. Bentuk-bentuk perancangan juga sudah mulai dipertimbangkan. Pada tahapan ini sudah mulai dapat dilakukan penggunaan *decision support system* yang memberikan alternatif untuk membantu memecahkan permasalahan.
- c. Ketiga adalah tahap **choice** yaitu tahap untuk menentukan alternatif apa yang akan digunakan di antara alternatif-alternatif yang telah ada. Pada tahap ini Sistem Pendukung Keputusan juga dapat dilakukan untuk menentukan alternatif-alternatif yang telah tersedia.
- d. Keempat adalah tahap **implementation** yaitu tahap di mana keputusan sudah dapat diimplementasikan/diterapkan dan selanjutnya hasilnya sudah dapat dilaporkan.



Sumber : McLeod, 2004
Gambar 2.2 Tahapan Sistem Pendukung Keputusan

Algoritma C4.5

Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Konsep dasar algoritma *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*).

Pembangunan *tree* dimulai dengan data pada simpul akar (*root node*) yang dilanjutkan dengan pemilihan sebuah atribut, formulasi sebuah *logical test* pada atribut tersebut dan pencabangan pada setiap hasil dari test. Langkah ini terus bergerak ke *subset* ke contoh yang memenuhi hasil dari simpul anak cabang (*internal node*) yang sesuai melalui proses rekursif pada setiap simpul anak cabang. Langkah-langkah tersebut diulangi hingga dahan-dahan dari *tree* memiliki contoh dari satu kelas tertentu. Gambar 1 memuat contoh dari sebuah *decision tree*. Beberapa model

decision tree yang sudah dikembangkan antara lain adalah *IDS*, *ID3*, *C4.5*, *CHAID* dan *CART*.

Algoritma C4.5 adalah pengembangan dari algoritma *ID3*. Oleh karena pengembangan tersebut *algoritma C4.5* mempunyai prinsip dasar kerja yang sama dengan algoritma *ID3*. Hanya saja dalam *algoritma C4.5* pemilihan atribut dilakukan dengan menggunakan *Gain Ratio* dengan rumus :

$$\text{GainRatio}(S,A) = \frac{\text{Gain}(S,A)}{\text{SplitInfo}(S,A)}$$

Atribut dengan nilai *Gain Ratio* tertinggi dipilih sebagai atribut test untuk simpul. Dengan *gain* adalah *information gain*. *Split Info* menyatakan entropi atau informasi potensial dengan rumus :

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

Sedangkan fungsi *data mining* yang digunakan dalam penelitian ini adalah fungsi klasifikasi dengan *algoritma C4.5*. Dalam melakukan klasifikasi akan digunakan data latihan atau data pengalaman, mengapa disebut demikian? Data histori disebut data latihan (*training data*) karena sesuai dengan pengertiannya yaitu akan menghasilkan pengetahuan. Pengetahuan dalam menggolongkan resiko penentuan penerimaan BSM yang tepat dimasa yang akan mendatang berdasarkan atribut-atribut yang ditentukan.

Terdapat 4 langkah dalam menentukan pohon keputusan menggunakan *algoritma C4.5* seperti disampaikan "*Crows* dalam Abidin (2010)" :

1. Memilih atribut sebagai akar (*root*).
2. Membuat cabang untuk tiap-tiap nilai.
3. Membagi kasus dalam cabang
4. Mengulangi lagi proses dalam cabang, sehingga semua kasus dalam cabang selesai.

Bantuan Beasiswa Siswa Miskin (BSM)

Setelah Pemerintah menetapkan kebijakan Bantuan Operasional Sekolah (BOS) sebagai salah satu pendukung untuk menuntaskan program Wajar Dikdas 9 Tahun, ternyata kebijakan BOS tersebut belum mampu menjamin seluruh masyarakat untuk dapat sekolah, terutama bagi anak-anak usia sekolah yang berasal dari keluarga miskin. Bantuan Operasional Sekolah (BOS) yang diberikan kepada MI dan MTs maupun lembaga penyelenggara Wajar Dikdas setara MI dan MTs, hanya mampu mengurangi beban biaya pendidikan yang harus dikeluarkan masyarakat, tetapi tidak mampu untuk membebaskan seluruh biaya pendidikan, sehingga banyak siswa miskin yang tidak sanggup atau

melanjutkan pendidikannya karena harus mengeluarkan biaya untuk buku, transportasi, seragam madrasah, sepatu, buku tulis atau biaya lainnya yang tidak dapat dipenuhi dari dana BOS. Dengan adanya hal diatas, Kementerian Agama melalui APBN menetapkan program "Bantuan Beasiswa Siswa Miskin" yang diberikan kepada sebagian siswa miskin di tingkat MI/MTs/MA, walaupun jumlah siswa yang mendapatkan bantuan ini masih sangat terbatas. Hal tersebut bertujuan untuk membiayai sebagian kebutuhan pendidikannya, sehingga dapat menyelesaikan pendidikannya, bahkan dapat melanjutkan pendidikan ke jenjang berikutnya, (www.madrasah.kemendag.go.id).

3. METODOLOGI PENELITIAN

Penggunaan Metode Eksperimen

Penelitian ini digunakan untuk menentukan prediksi siswa yang layak untuk mendapatkan BSM. Ada beberapa tahap yang dilakukan dalam melakukan penelitian ini adalah metode eksperimen. Metode ini menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian (C.R.Kothari, 2004). Misalnya, menguji apakah tinggi rendahnya honor seorang guru mempengaruhi kinerja proses belajar mengajar. Metode eksperimen juga dibagi dua, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen, misalnya pengaruh honor guru terhadap kinerja mengajar. Sedangkan eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik (C.R.Kothari, 2004). Dalam penelitian ini, penulis mengambil model eksperimen absolut.

Ada beberapa tahap yang dilakukan dalam melakukan eksperimen ini, penulis menggunakan model *Cross-Standard Industry for Data Mining (CRISPDM)* (Larose, 2005) yang terdiri dari 6 tahap, yaitu :

1. Tahap *business understanding*.

Tahap dimana peneliti mencari permasalahan yang terdapat di lokasi/tempat yang sedang diteliti dan mencari solusi dalam permasalahan tersebut. Permasalahan dalam studi kasus Dana BSM ini ialah kurangnya pemberian dana BSM kepada siswa-siswa yang memang sekiranya layak untuk mendapatkannya. Agar pemberian dana BSM ini akurat dan tepat sasaran, maka dikembangkan model klasifikasi algoritma *C4.5*

untuk menentukan kelayakan pemberian dana BSM dengan tujuan analisa yang dilakukan lebih akurat.

2. Tahap *data understanding*.

Data BSM diambil dari MTs Attaqwa Blendung-Tangerang pada tahun 2011, dimana jumlah datanya mencapai 150 siswa gabungan dari kelas 1 sampai dengan kelas 3. Ada beberapa atribut yang digunakan di dalam data BSM yaitu nama siswa, peringkat kelas, penghasilan orang tua, telat pembayaran SPP, jumlah anggungan (jumlah anak), anak yatim atau tidak, memiliki kartu multiguna, dan *class* (label yang diberikan kepada siswa, label baik untuk yang mendapatkan BSM ataupun yang tidak mendapatkan BSM).

Nilai dari semua atribut yang ada di tabel, merupakan nilai kategorikal dan bukan nilai angka, misalnya seperti atribut Jumlah Tanggungan yang mempunyai nilai sedikit merupakan kategori jika debitur mempunyai jumlah anak 1 sampai 2. Gambar 3.1 di bawah ini ditampilkan nama atribut, nilai angka dan kategorinya. Berikut rule nilainya :

Tabel 3. 1 Kategori atribut

Attribut	Nilai Angka	Kategori
Gaji	< 900.000	Kurang
	900.000 – 1.200.000	Sedang
	> 1.200.000	Cukup
Peringkat	1 - 5	Sangat Baik
	6 - 10	Baik
	> 10	Kurang Baik
Saudara	0 - 1	Sedikit
	2 - 4	Banyak
	>4	Sangat Banyak

3. Tahap *data preparation*.

Tabel di bawah ini menunjukkan data BSM tahun 2011 baik yang mendapatkan BSM maupun yang tidak mendapatkan BSM.

Tabel 3.2 Data Siswa yang diajukan BSM di MTs Attaqwa Blendung-Tangerang, tahun 2011 (Terlampir)

Tabel 3.2 merupakan data murni yang belum diolah menjadi data training. Jika kita lihat secara seksama pada tabel 3.2, maka akan didapati beberapa atribut yang tidak digunakan dan record yang duplikasi. Untuk itu maka diperlukan tehnik dalam *preprocessing* (Han & Kamber, 2006) yaitu :

- Data cleaning* bekerja untuk membersihkan nilai yang kosong ,tidak konsisten atau mungkin tupel yang kosong (*missing values* dan *noisy*).
- Data integration* berfungsi menyatukan tempat penyimpanan (arsip) yang berbeda

ke dalam satu data. Dalam hal ini, ada dua arsip yang diambil sebagai data *warehouse* yaitu data siswa dan data BSM.

- Data reduction*, jumlah atribut dan *tupel* yang digunakan untuk *data training* mungkin terlalu besar, hanya beberapa atribut yang diperlukan sehingga atribut yang tidak diperlukan akan dihapus. Tupel dalam *data set* mungkin terjadi duplikasi atau terdapat tupel yang sama, sehingga untuk memperkecil jumlah tupel, tupel yang sama dijadikan akan dijadikan dalam satu tupel untuk mewakili tupel tersebut sehingga akan terlihat pada tabel 3.3 di bawah :

Tabel 3.3 Data Training BSM (Terlampir)

4. Tahap *modelling (modeling)*.

Tahap ini juga dapat disebut tahap *learning* karena pada tahap ini *data training* diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Model yang digunakan dalam tahap ini menggunakan *algoritma C4.5*. Seperti yang telah dijelaskan sebelumnya, ada beberapa tahap yang harus dilalui dalam membentuk pohon keputusan, tentunya *algoritma C4.5* digunakan untuk membuat pohon keputusan. Untuk membuat pohon keputusan, dibutuhkan *data training* yang diambil dari tabel 3.3.

- Menghitung jumlah kasus *class* baik dan jumlah kelas ya dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut Penghasilan, Peringkat, Saudara, Terlambat_SPP, Yatim, dan Kartu_Multiguna. Kemudian, lakukan penyeleksian atribut dengan menghitung *Gain* tertinggi. *Entropy* dihitung dengan persamaan 2 sebagai berikut :

$$Entropy(S) = \sum_{i=1}^n - pi * \log_2 pi$$

Keterangan :

- S : himpunan kasus
- A : atribut
- n : jumlah partisi S
- pi : proporsi dari Siterhadap S

Kemudian hitung nilai *Gain* dengan metode *information gain*:

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

Keterangan :

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

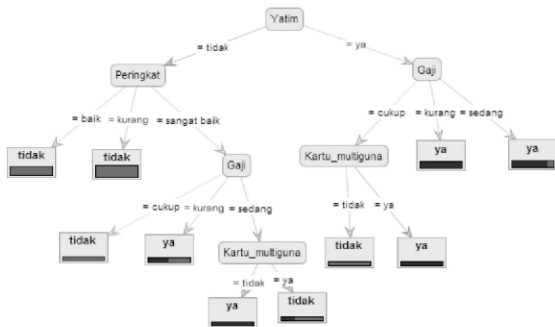
4. PEMBAHASAN

Perhitungan Gain dan Entropy

Dengan menggunakan persamaan Gain, hitung jumlah kasus dari masing-masing atribut dan hitung Gain. Hasil perhitungan dapat dilihat pada tabel di bawah ini :

Tabel 3.4 Tabel Perhitungan Node 1 (Terlampir)

Setelah dilakukan perhitungan untuk membentuk pohon keputusan, maka hasil pohon yang terbentuk akan terlihat seperti gambar 3.1. Dapat dilihat bahwa atribut Yatim menjadi akar dari pohon.



Gambar 3.1 Pohon Keputusan

Pohon keputusan pada gambar 3. Untuk menginterpretasikan pohon keputusan di atas menjadi aturan-aturan yaitu :

1. R1: IF Yatim=Tidak AND Peringkat=Baik THEN class=Tidak
2. R2: IF Yatim=Tidak AND Peringkat=Kurang THEN class=Tidak
3. R3: IF Yatim=Tidak AND Peringkat=Sangat Baik AND Gaji=Cukup THEN class=Tidak
4. R4: IF Yatim=Tidak AND Peringkat=Sangat Baik AND Gaji=Kurang THEN class=Ya
5. R5: IF Yatim=Tidak AND Peringkat=Sangat Baik AND Gaji=Sedang AND Kartu_Multiguna=Tidak THEN class=Ya
6. R6: IF Yatim=Tidak AND Peringkat=Sangat Baik AND Gaji=Sedang AND Kartu_Multiguna=Ya THEN class=Tidak
7. R7: IF Yatim=Ya AND Gaji=Cukup AND Kartu_Multiguna=Tidak THEN class=Tidak
8. R8: IF Yatim=Ya AND Gaji=Cukup AND Kartu_Multiguna=Ya THEN class=Ya
9. R9: IF Yatim=Ya AND Gaji=Kurang THEN class=Ya
10. R10: IF Yatim=Ya AND Gaji=Sedang THEN class=Ya

Pengujian Model

Nilai accuracy, precision, dan recall dari data training dapat dihitung dengan menggunakan Rapid Miner. Setelah diuji coba dengan metode cross-validation dengan 10 kali pengulangan (10-fold), didapatkan hasil pengukuran terhadap data training yaitu accuracy= 80%, precision= 72.73% dan recall=

88.89%. Model yang telah dikembangkan akan diuji keakuratannya.

Model yang telah dikembangkan akan diuji keakuratannya dengan memasukkan sejumlah data uji (DATA_TEST) ke dalam model. Untuk mengukur keakuratan model dengan baik, data uji seharusnya bukan data yang berasal dari data training (Han & Kamber, 2006). Data uji diambil dari Laporan Penerima Dana BSM tahun 2012. Ada 20 sampel yang akan dilakukan pengujian data dengan data training.

Setelah data uji dimasukkan ke dalam confusion matrix, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah sensitivity, specificity, precision dan accuracy. Sensitivity digunakan untuk membandingkan jumlah true positives terhadap jumlah tupel yang positives sedangkan specificity adalah perbandingan jumlah true negatives terhadap jumlah tupel yang negatives. Untuk menghitung digunakan persamaan di bawah ini :

$$sensitivity = \frac{t_pos}{pos}$$

$$specificity = \frac{t_neg}{neg}$$

$$precision = \frac{t_pos}{(t_pos + f_pos)}$$

$$accuracy = sensitivity \frac{pos}{(pos + neg)} + specificity \frac{neg}{(pos + neg)}$$

Dimana :

- t_pos = jumlah true positives
- t_neg = jumlah true negatives
- pos = jumlah tupel positif
- neg = jumlah tupel negatif
- f_pos = jumlah false positives

Kemudian masukkan nilai yang ada di dalam confusion matrix ke dalam persamaan di atas, sehingga akan menghasilkan nilai seperti di bawah ini :

Tabel 4.4. Jumlah Sensitivity, Specificity, Precision, Accuracy and Recall dalam presentasi.

	Nilai (%)
Sensitivity	73
Specificity	89
Precision	72.73
Accuracy	80
Recall	88.89

accuracy: 80.00% +/- 24.49% (mikro: 80.00%)			
	true tidak	true ya	class precision
prod. tidak	8	1	88.89%
prod. ya	3	8	72.73%
class recall	72.73%	88.89%	

Gambar 4.1. Jumlah Sensitivity, Specificity, Precision, Accuracy and Recall menggunakan Tools Rapid Miner Studio V.6 dengan model Confusion Matrix.

```

PerformanceVector
PerformanceVector:
accuracy: 80.00% +/- 24.49% (mikro: 80.00%)
ConfusionMatrix:
True:   tidak   ya
tidak:  8       1
ya:     3       8
precision: 72.73% (positive class: ya)
ConfusionMatrix:
True:   tidak   ya
tidak:  8       1
ya:     3       8
recall: 88.89% (positive class: ya)
ConfusionMatrix:
True:   tidak   ya
tidak:  8       1
ya:     3       8
AUC (optimistic): unknown (positive class: ya)
AUC: unknown (positive class: ya)
AUC (pessimistic): unknown (positive class: ya)
    
```

Gambar 4.2. Jumlah Deskripsi *Precision*, *Accuracy* and *Racall* menggunakan *Tools Rapid Miner Studio V.6* dengan model *Confusion Matrix*.



Gambar 4.3 Nilai AUC dalam grafik ROC dengan metode *information gain*

Perancangan Sistem Aplikasi dan GUI

Untuk melakukan pengujian rule yang telah dibuat diatas, maka disini akan diuji dengan menggunakan aplikasi yang dirancang menggunakan *Microsoft Visual Basic 6.0*. Akan dilakukan pengujian data sebanyak 20 data untuk mengevaluasi apakah rule yang dibuat sesuai atau tidak.

Dari Hasil Pengujian menggunakan Aplikasi Pengujian Rule dengan bantuan *Microsoft Visual Basic 6.0*, dari 20 sample yang diuji 85% sesuai dengan rule dan 15% tidak sesuai dengan *rule* yang telah dibentuk oleh *algoritma C4.5*.

5. PENUTUP

Kesimpulan

Dari hasil penelitian terbukti bahwa *algoritma C4.5* memiliki tingkat akurasi sebesar 0.80 dengan menggunakan sampel data penerima dana BSM (Bantuan Siswa Miskin). Hal ini dibuktikan dengan hasil evaluasi penelitian bahwa *algoritma C4.5* mampu menganalisa penerima dana BSM dengan akurasi sebesar 80%.

Saran

Walaupun model *algoritma C4.5* sudah diterapkan dan berjalan dengan tingkat akurasi 80%, namun ada beberapa hal yang harus ditambahkan untuk menambah akurasi *algoritma C4.5*, yaitu :

1. Melakukan pruning terhadap *algoritma C4.5* sehingga pohon keputusan yang terbentuk tidak terlalu besar bahkan mungkin untuk jumlah data yang besar sekalipun. Ini dilakukan untuk mengefisienkan kinerja dari *algoritma C4.5* tanpa mengurangi keakuratannya.
2. Untuk melihat tingkat akurasi dari algoritma, akan lebih baik *algoritma C4.5* dibandingkan atau dikomparasi model *algoritma* lain seperti *Naive Bayes*, *Support Vector Machine* atau *Neural Network*.
3. Menerapkan *algoritma C4.5* ke dalam data yang lebih besar untuk melakukan pengujian akurasi dari *algoritma* tersebut.
4. Dimasa yang akan datang, penelitian ini dapat digunakan dan dikembangkan oleh peneliti yang lain sebagai model untuk kegiatan penelitian sejenis ataupun bidang penelitian yang lain dengan tema dan metode yang sama dengan yang penulis gunakan.

Tentu saja penelitian ini masih jauh dari kata sempurna, oleh karena itu saran dan kritik yang bersifat membangun sangat diperlukan untuk memperbaiki dan mengembangkan penelitian ini di masa yang akan datang. Semoga penelitian ini dapat bermanfaat bagi kemajuan ilmu pengetahuan dan semua pihak.

DAFTAR REFERENSI

- [1] Arikunto, Suharsimi. 2007. *Manajemen Penelitian*. Jakarta : Rineka Cipta.
- [2] C.R.Kothari. 2004. *Research Methology Methods and Techniques*. India: New Age International Limited.

- [3] Han, J., & Kamber, M. 2006. Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman.
- [4] Isaac, S., & Michael, W. B. 1981. Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences. San Diego: EDITS Publisher.
- [5] Jiang, Y. 2009. Credit Scoring Model Based on Decision Tree and the Simulated Annealing Algorithm. 2009 World Congress on Computer Science and Information Engineering (hal. 18 - 22). Los Angeles: IEEE Computer Society.
- [6] Kotsiantis, S., Kanellopoulos, D., Karioti, V., & Tampakas, V. 2009. An ontology-based portal for credit risk analysis. 2009 2nd IEEE International Conference on Computer Science and Information Technology, (hal. 165-169). Beijing.
- [7] Kusriani, & Luthfi, E. T. 2009. Algoritma Data Mining. Yogyakarta: Andi Publishing.
- [8] Lai, K. K., Yu, L., Zhou, L., & Wang, S. 2006. Credit Risk Evaluation with Least Square Support Vector Machine. Springer-Verlag, 490-495.
- [9] Larose, D. T. 2005. Discovering Knowledge in Data. New Jersey: John Willey & Sons, Inc.
- [10] Marimin. 2004. *Teknik dan aplikasi pengambilan keputusan kriteria majemuk*. Jakarta : Grasindo.
- [11] McLeod Jr, Raymond. 2004. Sistem Informasi Manajemen. Jakarta : Prenhallindo.
- [12] Odeh, O. O., Featherstone, A. M., & Das, S. 2010. Predicting Credit Default: Comparative Results from an Artificial Neural Network, Logistic Regression and Adaptive Neuro-Fuzzy Inference System. EuroJournals Publishing, Inc. 2010, 7-17.
- [13] Prawirosentono, Suyadi. 2002. Strategi Pengambilan Keputusan Bisnis. Jakarta : Bumi Aksara.
- [14] Satchidananda, S. S., & Simha, J. B. 2010. Comparing Decision Trees With Logistic Regression for Credit Risk Analysis. Management of e-Commerce and e-Government (ICMeCG), 2010 Fourth International Conference, (hal. 35 - 38). Chengdu.
- [15] Sekaran, U. 1992. Research Metthods for Business : A Skill Building Approach, 2th Edition. New York: John Willey & Son.
- [16] Sugiyono. 2009. Metode Penelitian Kuantitatif Kualitatif dan R & D. Bandung: Alfabeta.
- [17] Surbakti, Irfan, 2002. Sistem Pendukung Keputusan. Surabaya : Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Surabaya
- [18] Triwahyuniati, N. 2008. Pelaksanaan Analisis Kredit Pada di PT Bank Huga Semarang. Thesis Magister Kenotariatan, 38.
- [19] Turban, Efraim., Aronson, Jay., dan Ting Peng Liang. 2005. Decision Support System and Intelligent System. Yogyakarta : Andi Offset.
- [20] Witten, I. H., Frank, E., & Hall, M. A. 2011. Data Mining : Practical Machine Learning and Tools. Burlington: Morgan Kaufmann Publisher.
- [21] Xhemali, D., Hinde, C. J., & Stone, R. G. 2009. Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. IJCSI International Journal of Computer Science Issues, 16-23.
- [22] Zurada, J. 2010. Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions. HICSS '10 Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, (hal. 1-9). Koloa.