
ANALISA KINERJA ALGORITMA C.45 DALAM MEMPREDIKSI HASIL BELAJAR

Muhammad Ardiansyah Sembiring¹, Mustika Fitri Larasati Sibuea², Andy Sapta³
^{1,2,3} Program Studi Sistem Informasi, STMIK Royal, Kisaran
e-mail: adinmantap88@gmail.com

Abstract: This research was done to predict the cumulative achievement index of students STMIK Royal Kisaran is based on students' academic history at the time of senior high school. This research will inform the student for early anticipation in following the lecture in order to get the maximum learning. The method used in solving this problem is by the method Decision tree C.45 algorithm starts from the formation of the decision system as preliminary data have attribute values and decision condition. Then calculate the entropy of each attribute. Calculating the highest gain value which would then be used as nodes. Determining Following the decision of the result tree Decision process using if-then logic begins from the highest to the root of the lowest roots. This research resulted in 7 pieces of pattern rules (rule) as a basis for predicting student achievement of learning outcomes.

Keywords: data mining, algorithm C.45, learning Outcomes.

Abstrak: Penelitian ini dilakukan untuk memprediksi hasil belajar mahasiswa sistem informasi STMIK Royal Kisaran berdasarkan riwayat akademik mahasiswa pada saat sekolah menengah atas. Penelitian ini akan menginformasikan kepada mahasiswa untukantisipasi dini dalam mengikuti perkuliahan agar mendapatkan hasil belajar yang maksimal. Metode yang digunakan dalam menyelesaikan permasalahan ini yaitu dengan metode *decision tree algoritma C.45* dimulai dari pembentukan *decision system* sebagai data awal yang memiliki nilai atribut kondisi dan keputusan. Kemudian menghitung nilai *entropy* dari masing-masing atribut. Menghitung nilai gain yang tertinggi yang selanjutnya akan digunakan menjadi *node*. Selanjutnya Menentukan keputusan dari hasil proses *decision tree* dengan menggunakan logika *if-then* dimulai dari akar tertinggi hingga akar terendah. Penelitian ini menghasilkan 7 buah pola aturan (*rule*) sebagai landasan dalam memprediksi pencapaian hasil belajar mahasiswa.

Kata kunci: *data mining*, algoritma C.45, hasil belajar.

PENDAHULUAN

Pendidikan merupakan salah satu faktor kemajuan dan kemandirian bangsa. Semakin maju pendidikan suatu bangsa,

maka akan semakin besar bangsa tersebut. Pendidikan perguruan tinggi merupakan pendidikan yang sangat penting dalam menggali segala potensi untuk siap terjun ke dunia profesional. Guna menghasilkan

lulusan yang berkompeten, maka kualitas dan manajemen pembelajaran di perguruan tinggi perlu ditingkatkan. Tujuannya agar hasil belajar setiap mahasiswa dapat terus ditingkatkan. Dalam usaha untuk meningkatkan kualitas lulusan, juga dibutuhkan kerjasama yang baik antara pemerintah, manajemen sekolah, kemitraan dengan dunia industri, kompetensi dosen, orang tua dan mahasiswa sendiri. Beberapa elemen tersebut sangat berpengaruh baik secara langsung maupun tidak langsung terhadap peningkatan mutu mahasiswa, yang pada akhirnya akan berimbas pada peningkatan kualitas pendidikan.

Rendahnya mutu pendidikan merupakan akibat dari rendahnya kualitas proses pembelajaran yang dilakukan di perguruan tinggi. Satu hal yang belum banyak dilakukan perguruan tinggi adalah melakukan antisipasi terhadap mahasiswa yang berpotensi mengalami hambatan dalam belajarnya. Hal ini dikarenakan riwayat akademik pada saat disekolah yang tidak sepenuhnya mendukung dengan hasil belajar yang dijalani pada perguruan tinggi. Hasil pembelajaran sangat penting dalam pendidikan dan dapat dilihat sebagai salah satu ukuran keberhasilan siswa dalam pendidikan di sekolah. Hal ini dianggap penting karena semakin awal mengetahui adanya potensi mahasiswa yang kemungkinan akan mengalami hambatan atau kesulitan dalam belajarnya, maka akan semakin cepat langkah-langkah antisipatif yang akan dilakukan.

Dampak dari kurangnya antisipasi atau pencegahan sejak dini terhadap mahasiswa yang berpotensi terhambat pada saat perkuliahan tampak dari hasil belajar berupa indeks prestasi yang rendah. Berdasarkan uraian di atas, maka penelitian ini bertujuan untuk membuat prediksi sejak dini terhadap mahasiswa yang berpotensi tidak berprestasi atau mengalami kegagalan pada sejumlah mata kuliah hingga sulit menyelesaikan studi tepat waktu.

Penelitian ini dilakukan di STMIK Royal Kisaran Prodi Sistem Informasi. Berdasarkan observasi yang telah dilakukan manajemen STMIK Royal belum memanfaatkan kumpulan data riwayat akademik dan hasil belajar mahasiswa dengan maksimal. Maka dengan menggunakan *data mining* metode *decision tree*, kumpulan riwayat akademik mahasiswa tercatat di akademik kampus akan diproses untuk mendapatkan pola berupa *rule* yang akan menjadi landasan dalam melakukan prediksi hasil belajar mahasiswa. Adapun atribut dari data yang akan dikelola meliputi data asal sekolah, jurusan sekolah dan peringkat kelas pada saat di sekolah dan Indeks Prestasi Kumulatif. Data - data tersebut belum dimanfaatkan dengan baik sebagai bahan pertimbangan, kajian dan penelitian untuk membuat suatu informasi yang berguna bagi sekolah. Berdasarkan uraian di atas, penelitian ini menggunakan metode *data mining* untuk membuat suatu prediksi mengenai hasil belajar mahasiswa dengan memanfaatkan tumpukan data yang telah dimiliki kampus tersebut.

Data mining sebenarnya merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Adapun proses KDD sebagai berikut:

1. *Data Selection*: pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
2. *Preprocessing*: sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk

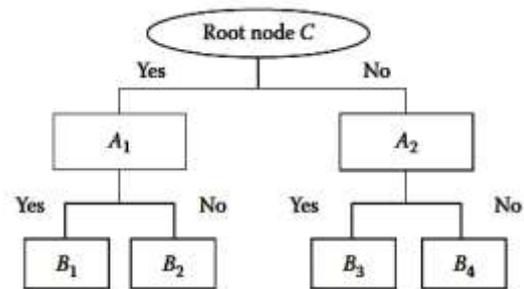
KDD, seperti data atau informasi eksternal.

3. *Transformation*: yaitu proses *coding* pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.
4. *Data mining*: proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
5. *Interpretation/Evaluation*: pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

METODE

Decision tree dengan algoritma C.45 merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root*. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

1. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.



Gambar 1. Node pada Decision Tree

Decision tree tergantung pada aturan *if-then*, tetapi tidak membutuhkan parameter dan metrik. Struktur sederhana dan dapat ditafsirkan memungkinkan *decision tree* untuk memecahkan masalah atribut *multi-type*. *Decision tree* juga dapat mengelola nilai-nilai yang hilang atau data *noise*.

Adapun langkah-langkah metode *decision tree* dengan algoritma C.45 dalam membangun sebuah pohon keputusan adalah sebagai berikut:

1. Membentuk *decision system* yang terdiri dari atribut kondisi dan atribut keputusan. Memperlihatkan sebuah contoh *decision system* yang pada penelitian ini. Ia hanyaterdiri dari n objek, $E_1, E_2, E_3, E_4, \dots, E_n$ dan *attribute* kondisi yaitu penjualan, pembelian, stok gudang, serta beban usaha. Sedangkan profit adalah *decision attribute*.
2. Hitung jumlah data kolom, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
3. Pilih atribut sebagai *Node*.
4. Buat cabang untuk tiap-tiap anggota dari *Node*.
5. Periksa apakah nilai *entropy* dari anggota *Node* ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai *entropy*

anggota *Node* adalah nol, maka proses pun berhenti.

6. Jika ada anggota *Node* yang memiliki nilai *entropy* lebih besar dari nol, ulangi lagi proses dari awal dengan *Node* sebagai syarat sampai semua anggota dari *Node* bernilai nol.

Node adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *gain* suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut:

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

Keterangan:

S = Himpunan Kasus

A = Atribut

n = jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

|S| = jumlah kasus dalam S

Sementara itu, untuk menghitung nilai *Entropy* dapat dilihat pada persamaan berikut ini:

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

Keterangan :

S= Himpunan Kasus

n = jumlah partisi S

P_i = proporsi S_i terhadap S

HASIL DAN PEMBAHASAN

Decision system yang akan diproses pada sistem tertuang dalam tabel 1. Tabel 1 memperlihatkan *decision system* yang akan diproses pada penelitian ini. Tabel tersebut menjelaskan sejumlah *n* objek, E1, E2, E3, E4,.....,E39 dan dengan *attribute* kondisi yaitu sekolah, jurusan, perinngkat. Sedangkan IPK merupakan atribut keputusan(*decision attribute*).

Berdasarkan tabel 1 selanjutnya dihasilkan perhitungan dari setiap atribut kondisi dan atribut keputusan berupa nilai *entropy* dan nilai *gain* yang menjadi dasar pembentukan *node*.

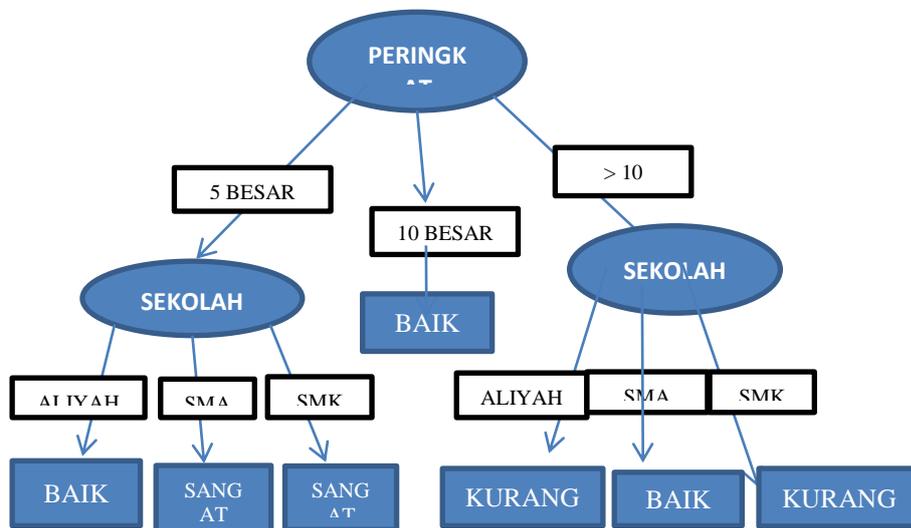
Dari tabel 2 tampak jelas hitungan jumlah kemunculan dari setiap atribut serta perhitungan nilai *entropy* dan nilai *gain* untuk setiap atribut. Untuk atribut yang memiliki nilai *gain* tertinggi maka akan menjadi *node*. Selanjutnya buatlah cabang untuk tiap-tiap anggota dari *Node*. Periksa apakah nilai *entropy* dari anggota *Node* ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai *entropy* anggota *Node* adalah nol, maka proses pun berhenti. Jika ada anggota *Node* yang memiliki nilai *entropy* lebih besar dari nol, ulangi lagi proses dari awal dengan *Node* sebagai syarat sampai

Tabel 1. *Decision System*

OBJEK	SEKOLAH	JURUSAN	PERINGKAT	IPK
E1	SMA	IPA	10 BESAR	BAIK
E2	SMA	IPA	10 BESAR	BAIK
E3	SMA	IPA	10 BESAR	BAIK
E4	SMA	IPA	5 BESAR	SANGAT
E5	SMA	IPA	10 BESAR	BAIK
E6	SMA	IPA	5 BESAR	SANGAT
E7	SMA	IPA	10 BESAR	BAIK
E8	SMA	IPA	10 BESAR	BAIK
...
E38	SMK	TEKNIK	5 BESAR	BAIK
E39	SMK	TEKNIK	> 10 BESAR	KURANG

Tabel 2. Perhitungan Nilai Atribut

	JUMLAH	SANGAT BAIK	KURANG BAIK	ENTROPHY	GAIN
TOTAL	39	8	22	9	1,4229
SEKOLAH					0,909
SMA	21	6	15	0	0
ALIYAH	4	0	2	2	0
SMK	14	2	5	7	1,4316
JURUSAN					0,5778
IPA	16	4	12	0	0
IPS	9	2	5	2	1,4355
TEKNIK	14	2	5	7	1,4316
PERINGKAT					1,4229
>10	12	0	3	9	0
10	16	0	16	0	0
5	11	8	3	0	0



Gambar 2 Pohon Keputusan Manual

anggota dari *Node* bernilai nol. Sehingga menghasilkan pohon keputusan seperti pada gambar 2.

Berdasarkan pohon keputusan pada gambar 3.1 maka dapat disimpulkan yang menjadi rule keputusan adalah:

1. Jika Peringkat = >10 besar dan Sekolah = SMK maka IPK = Kurang
2. Jika Peringkat = >10 besar dan Sekolah = SMA maka IPK = Baik
3. Jika Peringkat = >10 besar dan Sekolah = Aliyah maka IPK = Kurang
4. Jika Peringkat = 5 besar dan sekolah = SMK maka IPK = Sangat Baik
5. Jika peringkat = 5 besar dan sekolah = SMA maka IPK = Sangat Baik
6. Jika Peringkat =5 besar dan Sekolah = Aliyah maka IPK = Baik
7. Jika Peringkat = 10 besar maka IPK = Baik

Selanjutnya pengujian dengan *tools* yang telah ditentukan yaitu *Rapidminer 5*. Proses decision tree pada *Rapidminer 5* dimulai dari menginput

tabel *decision system* sehingga tampak seperti gambar 3.

SEKOLAH	JURUSAN	PERINGKAT	IPK
polyno...	polyno...	polyno...	polyno...
[attribute]	[attribute]	[attribute]	[label]
SMA	IPA	10 BESAR	BAIK
SMA	IPA	10 BESAR	BAIK
SMA	IPA	10 BESAR	BAIK
SMA	IPA	5 BESAR	SANGAT BAI
SMA	IPA	10 BESAR	BAIK
SMA	IPA	5 BESAR	SANGAT BAI
SMA	IPA	10 BESAR	BAIK
ALYAH	IPS	5 BESAR	BAIK
ALYAH	IPS	>10 BESAR	KURANG

Gambar 3. *Decision System* Rapidminer 5

Gambar 3. menunjukkan atribut kondisi yaitu sekolah, jurusan dan peringkat. Sedangkan IPK merupakan atribut keputusan yang ditandai dengan mengganti inisial attribute menjadi inisial label. Tahapan selanjutnya di dalam *tools Rapidminer 5* melakukan koneksi antara *decision system* yang telah diproses dengan metode *decision tree* yang telah tersedia sehingga menghasilkan pohon keputusan pada gambar 4.

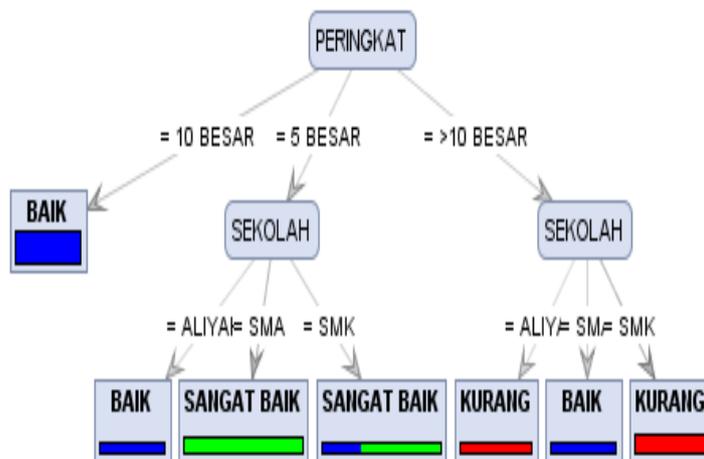
Hasil Pohon keputusan pada gambar 4 memiliki hasil yang sama

dengan proses manual yang telah dilakukan sesuai dengan gambar 1. Untuk mendapatkan rule keputusan secara teks cukup dengan memilih *text view* pada *rapidminer 5* sehingga tampak gambar 5.



Gambar 5. *Text View* Pohon Keputusan

Berdasarkan gambar 5 dapat dipastikan bahwa hasil pengujian dengan *tools rapidminer* sama dengan hasil manual sebagaimana gambar 3.1. Beberapa manfaat yang dirasakan sejak adanya sistem pengolahan *data mining* menggunakan *decision tree* dengan *algoritma C.45* sebagai metode pemecahan masalah dengan *tools Rapidminer* sebagai *tools* untuk menguji data dan menghasilkan *rule* sebagai landasan melakukan prediksi hasil belajar



Gambar 4. Pohon Keputusan Rapidminer 5

mahasiswa yang dihasilkannya adalah dengan memanfaatkan *Rapidminer 5*, *rule* yang ditemukan dapat terlihat dengan jelas, baik pohon keputusan maupun text view dari pohon keputusan tersebut dengan waktu pengerjaan yang cukup efektif.

SIMPULAN

Berdasarkan proses dan hasil penelitian yang telah dilakukan adalah sebagai berikut:

1. Implementasi data mining metode *decission tree algoritma C.45* mampu menghasilkan *rule* guna memprediksi pencapaian hasil belajar mahasiswa berdasarkan riwayat akademik terdahulu.
2. Pengujian *decission system* dengan menggunakan *Tools Rapidminer 5* sangat dirasakan dapat mempermudah proses *decission system* dalam menghasilkan *rule* keputusan sebagai dasar melakukan prediksi.

DAFTAR PUSTAKA

- Mabrur, A.G. 2012. Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit. *Jurnal Komputa*. I: 53 – 57.
- Fajar A.H. 2009. *Data Mining*. Ed.I.Yogyakarta: Andi. Hlm.2-3.
- Kartika, 2013. Implementasi *Data Mining* Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus: Kreditplus). *Pelita Informatika Budi Dharma*. IV(3): 26 – 27
- Prasetyo. 2012. *Data Mining* Konsep dan Aplikasi Menggunakan Matlab. Ed.I. Yogyakarta: Andi. Hlm.3 – 7.
- Sapta, A., 2017. Pengaruh Penggunaan Quiz Creator Terhadap Hasil Belajar Matematika Siswa. *Jurnal Mathematics Paedagogic*, 1(1), pp.91- 96.
- Sembiring, M.A., 2016. Penerapan Metode Decission Tree Algoritma C.45 Untuk Memprediksi Hasil Belajar Mahasiswa Berdasarkan Riwayat Akademik. *JURTEKSI*, 3(1), pp.60-65.
- Seruni. 2014. Pemberian Umpan Balik dalam Meningkatkan Hasil Belajar dan Minat Belajar Mahasiswa. ISSN: 2088-351X.
- Sibuea, M.F.L., 2017. Implementasi Model Pembelajaran Kooperatif Tipe Think Talk Write (Ttw) Sebagai Upaya Meningkatkan Hasil Belajar Siswa. *MES (journal of mathematics education and science)*, 2(2).