

SISTEM REKOMENDASI PENJURUSAN SEKOLAH MENENGAH KEJURUAN DENGAN ALGORITMA C4.5

Indra Mukti Prabowo dan Subiyanto
Fakultas Teknik Universitas Negeri Semarang
email: blaqerz11@gmail.com

Abstrak

Penelitian ini bertujuan untuk menyajikan model *decision tree* dengan algoritma C4.5 dalam mengembangkan sistem rekomendasi pemilihan jurusan untuk calon siswa baru Sekolah Menengah Kejuruan (SMK). Pendekatan yang digunakan adalah *research and development (R&D)*. Pengumpulan data dilakukan dengan teknik angket dan studi dokumentasi. Variabel *input* yang digunakan dalam penelitian ini antara lain: minat, bakat akademik, nilai ujian nasional, dan jenis kelamin. Pilihan jurusan menjadi variabel target. *Decision tree* digunakan dalam menganalisis data siswa kelas 10 SMK se-Kecamatan Batang. Algoritma C4.5 digunakan untuk membangun *decision tree* yang menggambarkan hubungan antara variabel *input* dengan variabel target dalam bentuk pola. Pola tersebut digunakan sebagai aturan untuk proses klasifikasi variabel *input* ke dalam variabel target. Data penelitian dianalisis dengan cara membandingkan hasil *output system* dengan data siswa kelas 10 dengan tiga besar *ranking* paralel sebagai data uji. Hasil uji sistem menunjukkan bahwa sistem dapat memberikan rekomendasi yang tepat sebesar 83,33% dari 48 data uji.

Kata kunci: *rekomendasi, pemilihan minat, jurusan, algoritma C4.5*

RECOMMENDATION SYSTEM FOR VOCATIONAL MAJOR STREAMING BY C4.5 ALGORITHM

Abstract

This study was aimed at presenting decision tree model using C4.5 algorithm in developing a major selection system for vocational schools. The study was research and development using questionnaires and documentation as data collection instruments. The input variables were: interest, academic talent, National Exam score, and gender. The target variable was choice of majors. Decision trees were used to analyze the data from grade 10 of vocational schools Batang in District. The C4.5 Algorithm was used to build decision trees in describing the relationship between the input variables and the target variable in the form of patterns. The patterns were used as a guide for the classification of the input variables into the target variable. The data were analyzed by comparing results of the output system and students' highest parallel ranking. Results show that the system is able to provide appropriate recommendations up to 83.33% out of the 48 tested data.

Keywords: *recommendation system, major selection, vocational, algoritma C4.5*

PENDAHULUAN

Hambatan dalam menentukan jurusan sering dialami oleh siswa yang akan meneruskan studi ke jenjang Sekolah Menengah Kejuruan (SMK). Hambatan dapat berasal dari dalam maupun luar diri siswa. Siswa merasa tidak yakin terhadap kemampuannya menjadi salah satu dari hambatan yang berasal dari dalam diri siswa. Hambatan dari luar yang sering menimpa siswa salah satunya adalah paksaan dari orang tua. Paksaan yang terjadi biasanya dilatarbelakangi oleh pekerjaan di masa depan yang diharapkan orang tua untuk anaknya (Hartinah, 2010).

Salah satu solusi yang ditawarkan oleh pihak sekolah adalah dengan mengikuti bimbingan karier. Namun, bimbingan karier masih terkendala oleh terbatasnya jumlah dan minimnya keterampilan konseling guru bimbingan dan konseling (selanjutnya ditulis BK) serta kesalahan persepsi siswa tentang bimbingan karier. Keterbatasan jumlah guru BK masih terjadi di berbagai daerah dan menghambat berjalannya bimbingan karier (Kamaluddin, 2011; Simamora & Suwarjo, 2013). Kusmaryani (2010) memaparkan kegiatan bimbingan konseling yang dilaksanakan oleh guru BK masih kurang optimal karena beberapa guru BK belum menguasai sepenuhnya keterampilan konseling. Selain itu, kesalahpahaman siswa juga menjadi kendala berjalannya bimbingan karier. Siswa sering menganggap bahwa siswa yang bertemu dengan guru BK merupakan siswa nakal sehingga siswa lebih memilih berbagai cara dengan teman daripada dengan guru BK (Setyaningrum & Setiawati, 2013). Falentini (2013) menjelaskan bahwa pemberian masukan dari teman tentang pilihan karier menjadi hambatan paling dominan dalam siswa menentukan pilihan karier.

Penelitian yang bertujuan untuk membangun sistem yang dapat memberikan

rekomendasi jurusan SMK berdasarkan klasifikasi kemampuan, bakat, dan minat siswa. Klasifikasi-klasifikasi tersebut sangat dibutuhkan oleh siswa dalam menghadapi hambatan pemilihan jurusan.

Klasifikasi dan prediksi menjadi teknik dari pemerolehan (*data mining*) yang sering digunakan. *Decision tree* menjadi salah satu metode klasifikasi dari *data mining* yang memiliki bentuk seperti struktur pohon, yakni setiap cabang mempresentasikan nilai variabel dan daun mempresentasikan label kelas (Han, Kamber, & Pei, 2012, p. 330). Pembangunan *decision tree* menggunakan himpunan data (tupel) sebagai data training. Data tupel yang digunakan harus memiliki atribut-atribut dan label kelas. Atribut yang ada dapat bersifat diskrit maupun kontinu (Kabra & Bichkar, 2011). *Decision tree* digunakan untuk mengklasifikasikan data tupel yang belum memiliki label kelas dengan cara menelusuri cabang hingga mencapai daun dari *decision tree*.

Decision tree menjadi salah satu teknik klasifikasi dalam *data mining* yang memiliki performa yang baik (Muangnak, Pukdee, & Hengsanunkun, 2010; Pandey & Taruna, 2014). *Decision tree* sudah diterapkan dalam banyak karya tulis seperti Jantan, Hamdan, dan Othman (2010); Gu dan Zhou (2012); Amin, Indwiarti, dan Sibaroni (2015); dan Soliman, Abbas, dan Salem (2015).

Decision tree dibangun menggunakan data dengan variabel *input* antara lain minat, bakat, nilai UN, dan jenis kelamin siswa dengan label kelas jurusan. Hasil rekomendasi data uji didapatkan dengan menelusuri setiap cabang dari *decision tree* berdasarkan data variabel *input* sampai daun untuk mengetahui jurusan yang menjadi rekomendasi.

Algoritma C4.5 menjadi salah satu teknik pembentukan *decision tree* yang sangat populer. Semua penelitian yang

dipaparkan di atas menggunakan algoritma C4.5 dalam membentuk *decision tree*. Algoritma C4.5 diciptakan oleh Quinlan (1993) untuk membangun *decision tree*. Algoritma C4.5 dikembangkan dari algoritma ID3 (Larose, 2005, p. 116). Pemilihan atribut sebagai cabang pada algoritma C4.5 menggunakan hasil perhitungan dari *gain ratio*.

Performa algoritma C4.5 lebih baik dibandingkan dengan teknik pembentuk *decision tree* lain seperti SimpleCart, REPTree, NBTree, ADTree, dan ID3 (Sharma & Sahni, 2011; Yadav & Pal, 2012). Selain itu, beberapa penelitian sudah menerapkan algoritma C4.5 dalam membentuk *decision tree* di berbagai bidang. Pada bidang pendidikan, Andriani (2013) menyajikan sistem klasifikasi pemberian beasiswa menggunakan metode *decision tree* dengan algoritma C4.5. Hasil klasifikasi penerima beasiswa tersebut yang dievaluasi dengan *confusion matrix* menghasilkan tingkat akurasi hasil klasifikasi penerima beasiswa dengan *decision tree* sebesar 71,43%. Swastina (2013) menerapkan algoritma C4.5 dalam penentuan jurusan mahasiswa. *Decision tree* yang dibentuk menggunakan algoritma C4.5 menghasilkan nilai akurasi 93,31% lebih baik dibandingkan Naive Bayes menghasilkan nilai akurasi sebesar 82,64%.

Penelitian yang mempelajari tentang penerapan suatu metode dalam pemilihan jurusan sudah banyak dilakukan. Swastina (2013) melakukan penelitian tentang penerapan algoritma C4.5 dalam penentuan jurusan mahasiswa. Fartindiyyah dan Subiyanto (2014) mengembangkan sistem pendukung keputusan peminatan SMA sesuai dengan Kurikulum 2013 menggunakan metode *weighted product*.

Dengan pertimbangan di atas, algoritma C4.5 digunakan dalam membangun *decision tree* untuk memberikan rekomendasi

jurusan SMK. Dengan demikian diharapkan calon siswa yang akan melanjutkan ke SMK dapat melewati hambatan pemilihan jurusan dengan mudah.

METODE

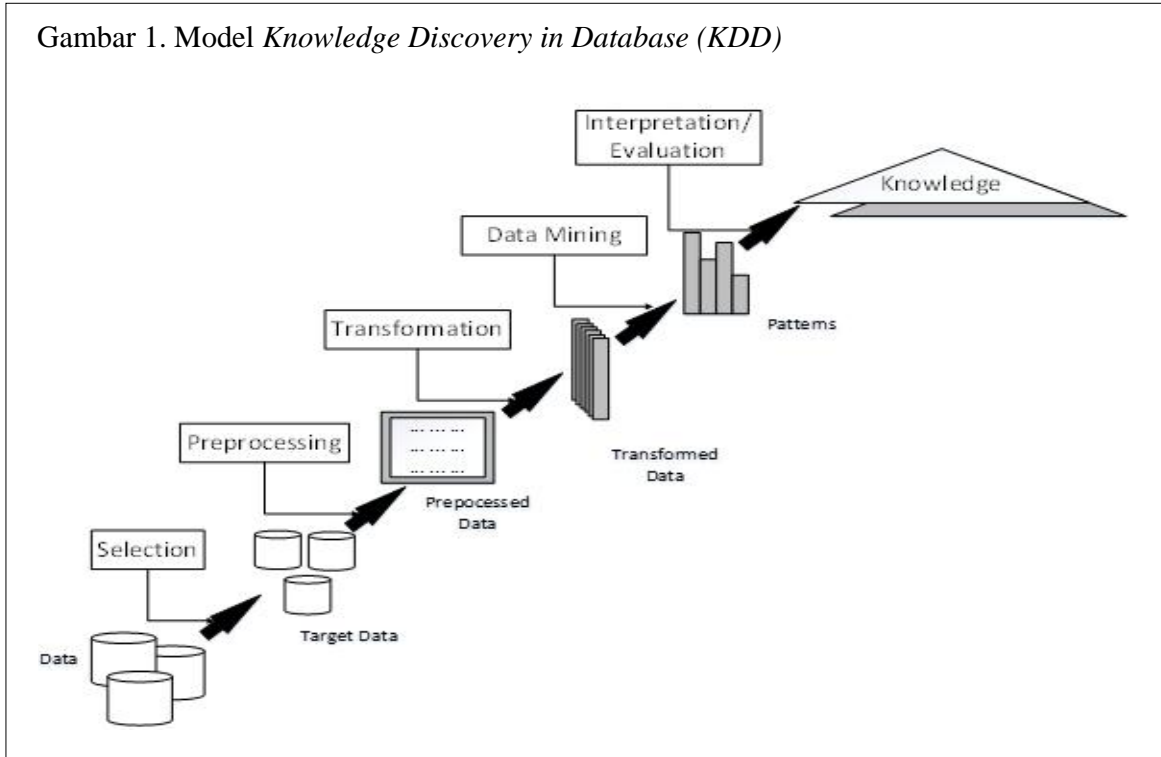
Pendekatan *research and development* digunakan dalam penelitian ini untuk mengembangkan sistem rekomendasi pemilihan minat jurusan bagi calon siswa baru SMK. Siswa kelas 10 Tahun Pelajaran 2015/2016 SMK di Kecamatan Batang menjadi subjek dari penelitian ini. Jumlah siswa kelas 10 yang digunakan sebagai responden dalam penelitian sejumlah 291 siswa. *Sampling* menggunakan tabel Krejcie sebagai dasar menentukan jumlah responden dalam penelitian ini. Waktu penelitian pada bulan Juni 2015 sampai bulan Maret 2016.

Pengumpulan data yang dilakukan menggunakan teknik angket dan studi dokumentasi. Angket digunakan untuk mengambil data minat siswa dengan berdasarkan Tes RIASEC. Data minat siswa yang diambil dibagi menjadi enam jenis kepribadian. Data bakat akademis siswa dengan menggunakan Tes Potensi Akademik (TPA). Studi dokumentasi digunakan dalam mengambil data identitas siswa antara lain NIS, nama, jenis kelamin, nilai Matematika UN SMP, nilai Bahasa Indonesia UN SMP, nilai Bahasa Inggris UN SMP, nilai IPA UN SMP, dan jurusan.

Penelitian ini menggunakan *Knowledge Discovery in Database (KDD)* sebagai prosedurnya. KDD yang dilakukan terbagi menjadi lima tahapan yang terdiri atas *data selection*, *data preprocessing*, *data transformation*, *data mining/modeling*, dan *evaluation* seperti terlihat pada Gambar 1 (Hermawati, 2009).

Data siswa kelas 10 digunakan sebagai *data training* atau data pembangun *decision tree*. Data siswa kelas 10 dengan peringkat

Gambar 1. Model *Knowledge Discovery in Database (KDD)*



akademis tiga besar digunakan sebagai *data testing* atau data yang digunakan untuk evaluasi dari *decision tree*. Data yang digunakan untuk membangun *decision tree* sebanyak 291 data siswa SMK di Kecamatan Batang. Data yang digunakan untuk menguji tingkat akurasi system sebanyak 48 data siswa kelas 10 SMK peringkat akademis tiga besar.

Dalam menganalisis data untuk memberikan rekomendasi pemilihan minat jurusan calon siswa baru SMK digunakan metode *decision tree* yang dibentuk dengan algoritma C4.5.

Nilai *gain ratio* digunakan untuk menentukan variabel *input* yang menjadi cabang dalam pembentukan *decision tree* dengan menggunakan algoritma C4.5. Sebelum melakukan perhitungan *gain ratio*, perhitungan nilai entropy dari keseluruhan data siswa dilakukan dengan Persamaan (1).

$$I_j(s) = \sum_{i=1}^n -\frac{J_i}{S} * \log_2 \frac{J_i}{S} \quad (1)$$

Keterangan:

- I_j : entropy keseluruhan data siswa
- J_i : jumlah data siswa dengan jurusan ke- i
- n : jumlah jurusan
- S : jumlah keseluruhan data siswa

Perhitungan nilai entropy dilakukan untuk setiap nilai variabel *input* dengan menggunakan Persamaan (2), sesudah diketahui nilai entropy dari keseluruhan data siswa

$$I_v = (S_v) = \sum_{i=1}^n -\frac{J_{v_i}}{|S_v|} * \log_2 \frac{J_{v_i}}{|S_v|} \quad (2)$$

Keterangan:

- I_v : Nilai entropy nilai variabel *input*
- J_{v_i} : Jumlah data siswa untuk jurusan ke- i dengan kondisi variabel *input* sama dengan nilai variabel *input*

n : Jumlah Jurusan
 S_v : Jumlah keseluruhan data siswa dengan kondisi variabel *input* sama dengan nilai variabel *input*

Perhitungan nilai *gain* setiap variabel *input* dapat dilakukan dengan menggunakan Persamaan (3), sesudah nilai entropy dari semua nilai variabel diketahui.

$$Gk(j, v) = I_j(j) - \sum_{i=1}^n \frac{|S_{v_i}|}{|S|} * I_{v_i} \quad (3)$$

Keterangan:

Gk : nilai *gain* variabel *input*
 j : keseluruhan data siswa
 v : variabel *input*
 I_j : nilai entropy keseluruhan data siswa
 I_{v_i} : nilai entropy dari nilai variabel *input* ke- i
 S_{v_i} : jumlah keseluruhan data siswa dengan kondisi variabel *input* sama dengan nilai variabel *input* ke- i
 S : jumlah keseluruhan data siswa
 n : jumlah nilai variabel *input*

Perhitungan nilai *split* info untuk tiap variabel *input* dengan Persamaan (4) dilakukan sebelum perhitungan *gain ratio*.

$$SI(v) = \sum_{i=1}^n \frac{|S_{v_i}|}{|S|} \log_2 \left(\frac{|S_{v_i}|}{|S|} \right) \quad (4)$$

Keterangan

SI : nilai *split* info setiap variabel *input*
 v : variabel *input*
 S_{v_i} : jumlah keseluruhan data siswa dengan kondisi variabel *input* sama dengan nilai variabel *input* ke- i
 S : jumlah keseluruhan data siswa
 n : jumlah nilai variabel *input* untuk setiap variabel *input*

Kita dapat mencari nilai *gain ratio* untuk setiap variabel *input* menggunakan persamaan (5), sesudah didapat nilai *split* info untuk tiap variabel *input*.

$$gain\ ratio(v) = \frac{Gk(v)}{SI(v)} \quad (5)$$

Keterangan

v : variabel *input*
 Gk : nilai *gain* variabel *input*
 SI : nilai *split* info variabel *input*

Penentuan variabel *input* sebagai akar dan cabang dengan cara membandingkan nilai *gain ratio* tiap variabel *input*. Variabel *input* dengan nilai *gain ratio* terbesar digunakan sebagai akar atau cabang dari *decision tree* sistem rekomendasi SMK.

Variabel *input* dengan nilai *gain ratio* tertinggi juga digunakan sebagai syarat kondisi dalam penggunaan data siswa pada iterasi berikutnya. Perhitungan iterasi pembentukan *decision tree* berhenti saat semua variabel *input* tidak tersisa atau tidak dapat dipartisi lagi dan jurusan yang memiliki jumlah data siswa tersisa atau terbanyak dijadikan sebagai daun dari *decision tree*.

HASIL PENELITIAN DAN PEMBAHASAN

Tahap *pertama* dalam penelitian ini adalah tahap *data selection*. Pada tahap *data selection* dilakukan proses penentuan variabel data yang digunakan dalam pembentukan *decision tree*. Data jurusan dipilih menjadi variabel keputusan dari pembentukan *decision tree*. Variabel yang digunakan sebagai penentu dalam pembentukan *decision tree* terdiri dari minat, bakat, nilai UN, nama, dan jenis kelamin.

Tahap *kedua* adalah *data preprocessing* yang terdiri dari dua tahap, yaitu tahap *data intergration* dan *data cleaning*. Tahap *data intergration* dilakukan penggabungan 4 *database*, yaitu identitas, nilai UN SMP, minat, dan bakat akademis siswa berdasarkan NIS untuk setiap sekolah. Selanjutnya, data delapan sekolah

digabungkan menjadi satu *database* untuk digunakan menjadi *data training*.

Tahap selanjutnya adalah *data cleaning*. Pada tahap *data cleaning* dilakukan penghilangan data yang memiliki *missing value*. Data yang memiliki *missing value* ditemukan sebanyak 72 data sehingga data yang digunakan dalam pembentukan *decision tree* sebanyak 219 data siswa.

Tahap *ketiga* adalah tahap *data transformation*. Tahap *data transformation* diterapkan agar data yang digunakan sesuai dengan format *data mining*. Data yang mengalami proses *data transformation* yaitu data nilai UN untuk mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA dikelompokkan seperti Tabel 1.

Tabel 1

Transformasi Nilai Variabel Nilai UN

No	Nilai UN	Klasifikasi
1	85 < Nilai UN < 100	Sangat Baik (SB)
2	70 < Nilai UN < 85	Baik (B)
3	55 < Nilai UN < 70	Cukup (C)
4	UN < 55	Kurang (K)

Sumber: Buku Saku UN 2015

Penentuan variabel dan nilai variabel, serta label kelas yang digunakan dalam tahap *data mining/modeling* seperti Tabel 2 dilakukan sesudah tahap *data selection*, *data preprocessing* dan *data transformation*.

Tahap *keempat* yang dilakukan yaitu tahap *data mining* atau *modeling*. Pada tahap ini analisis data yang dilakukan berupa pembentukan *decision tree* dilakukan dengan menggunakan Algoritma C4.5 dengan 219 data siswa kelas X. Pembentukan tersebut dilakukan untuk mengetahui hubungan antara variabel *input* dengan variabel target yang digunakan sebagai aturan dalam rekomendasi jurusan SMK.

Pembentukan *decision tree* yang sudah dilakukan, menghasilkan *decision tree* dengan jumlah daun sejumlah 71 daun seperti terlihat pada Gambar 2. *Decision tree* yang terbentuk digunakan sebagai aturan dalam memberikan rekomendasi jurusan oleh sistem berdasarkan variabel *input* yang dimasukkan ke dalam sistem.

Tahap *evaluation* dilakukan dengan membandingkan hasil rekomendasi menggunakan *decision tree* dengan data jurusan yang ada. Data yang digunakan dalam tahap ini adalah *data testing* yang terdiri dari 48 data siswa kelas 10 dengan 3 besar prestasi akademis SMK se-Kecamatan Batang. Hasil dari perbandingan tersebut digunakan sebagai validasi hasil rekomendasi dari sistem yang berupa nilai akurasi.

Tahap *kelima*, *evaluation* yang dilakukan bertujuan untuk melihat tingkat kepercayaan dari sistem rekomendasi pemilihan minat jurusan SMK dengan algoritma C4.5. Perbandingan antara hasil *output* sistem rekomendasi dengan hasil pemilihan jurusan oleh siswa dengan peringkat akademis tiga besar.

Penelitian ini mengujikan 48 data sampel siswa dengan tiga besar peringkat akademis dari SMK se-Kecamatan Batang untuk dibandingkan dengan rekomendasi dari Algoritma C4.5 pada tahap *evaluation*. Data pengujian dari 48 siswa peringkat akademis tiga besar, diproses oleh sistem rekomendasi jurusan SMK dengan algoritma C4.5. Nilai variabel setiap data dimasukkan ke *decision tree* untuk mencari label kelas sebagai rekomendasi jurusan. Kemudian hasil rekomendasi jurusan dibandingkan dengan jurusan sebenarnya dari data siswa tersebut. Hasil uji tersebut disajikan ke dalam sebuah *confusion matrix* terlihat pada Tabel 3.

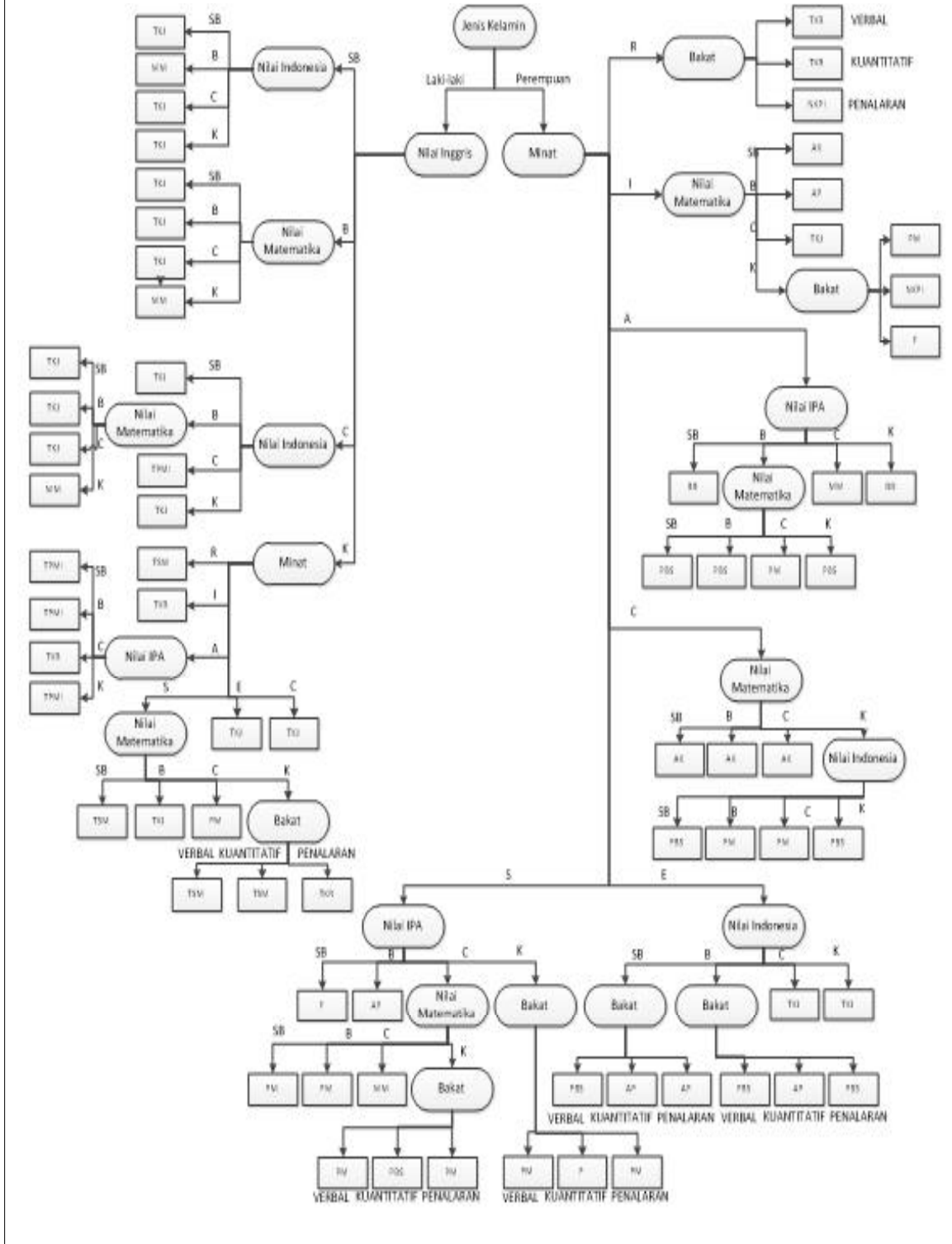
Tabel 2
Variabel Input dan Target

No	Nama	Nilai
1	Jenis Kelamin	Laki-laki, Perempuan
2	Minat (Librarians, 2008)	Realistik (R), Investigatif (I), Artistik (A), Sosial (S), Enterprise (E), Conventional (C),
3	Bakat (Gunarto, 2014)	Verbal, Kuantitatif, Penalaran
4	Nilai UN Matematika	SB, B, C, K
5	Nilai UN IPA	SB, B, C, K
6	Nilai UN B Indonesia	SB, B, C, K
7	Nilai UN B Inggris	SB, B, C, K
8	Jurusan (Variabel target)	Akutansi (AK), Administrasi Perkantoran(AP), Pemasaran (PM), Perbankan Syariah(PBS), Teknik Komputer dan Jaringan (TKJ), Multimedia (MM), Farmasi (F), Teknik Kapal Penangkap Ikan (TKPI), Neutika Kapal Penangkap Ikan (NKPI), Teknik Pemeliharaan Mekanik Industri (TPMI), Teknik Sepeda Motor (TSM), Teknik Kendaraan Ringan (TKR), Busana Butik (BB)

Tabel 3
Confusion Matrix Hasil Pengujian

	AK	TKJ	PM	MM	PBS	TSM	TPMI	F	TKR	NKPI	TKPI	BB	AP	Class Precision (%)
AK	6	0	0	0	0	0	0	0	0	0	0	0	0	100,0
TKJ	0	3	0	0	0	0	0	0	0	0	0	0	0	100,0
PM	0	0	3	0	0	0	0	0	0	0	0	0	0	100,0
MM	0	0	0	3	0	0	0	0	0	0	0	0	0	100,0
PBS	0	0	0	0	3	0	0	0	0	0	0	0	0	100,0
TSM	0	1	0	0	0	8	0	0	0	0	0	0	0	88,9
TPMI	0	0	0	0	0	1	2	0	0	0	0	0	0	67,3
F	0	0	0	0	0	0	0	3	0	0	0	0	0	100,0
TKR	0	0	0	0	0	0	0	0	3	0	0	0	0	100,0
NKPI	0	1	0	0	0	0	1	0	1	0	0	0	0	0
TKPI	0	1	0	1	0	0	1	0	0	0	0	0	0	0
BB	0	0	0	0	0	0	0	0	0	0	0	3	0	100,0
AP	0	0	0	0	0	0	0	0	0	0	0	0	3	100,0
Class Recall (%)	100	50	100	75	100	88,9	50	100	75	0	0	100	100	

Gambar 2 Hasil Pembentukan *Decision Tree*

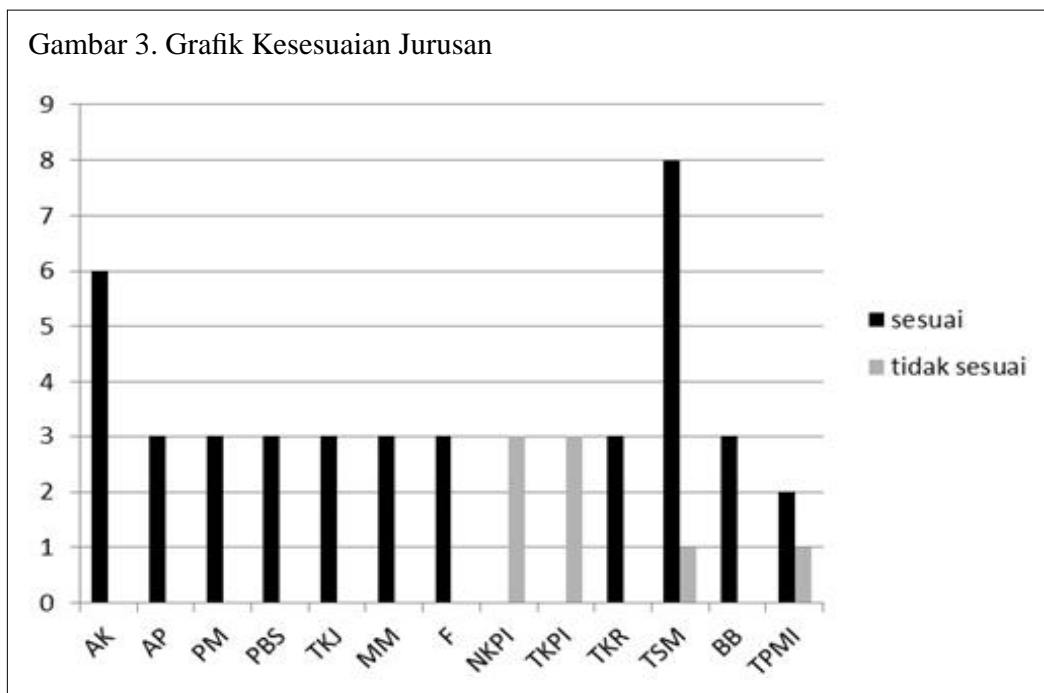


Berdasarkan *confusion matrix*, sistem berhasil memberikan rekomendasi jurusan yang sesuai dengan data jurusan siswa akademis tiga besar berprestasi pada data eksperimen sebanyak 40 data siswa. Jumlah data yang hasil rekomendasi jurusan dari sistem tidak sesuai dengan data jurusan siswa berprestasi adalah 8 data. Dengan demikian tingkat akurasi dari sistem rekomendasi jurusan SMK sebesar 83,33% dan *error rate* 16,67%.

Hasil rekomendasi jurusan yang tepat 100% terjadi pada data pengujian dengan jurusan AK, TKJ, PM, MM, PBS, F, TKR, BB, dan AP. Data pengujian dengan jurusan TPMI memiliki nilai akurasi rekomendasi 67,33% yaitu dua dari tiga memiliki hasil rekomendasi yang sesuai dengan data kenyataan siswa, yaitu TPMI. Satu data lain direkomendasikan jurusan TSM. Data dengan jurusan TSM sebanyak sembilan data dapat direkomendasikan sesuai dengan jurusan siswa sebanyak delapan data, sedangkan satu data direkomendasikan oleh

sistem sebagai jurusan TKJ. Nilai akurasi sistem dalam uji data dengan jurusan TSM adalah 88,9%. Keseluruhan data pengujian dengan jurusan TKPI dan NKPI tidak dapat direkomendasikan sesuai dengan data kenyataan jurusan siswa. Tiga data dengan jurusan TKPI direkomendasikan sistem sebagai jurusan TKR, TPMI, dan TKJ. Tiga data dengan jurusan NKPI direkomendasikan oleh sistem sebagai jurusan TPMI, MM, dan TKJ. Secara ringkas data dapat dilihat seperti pada Gambar 3.

Artikel ini belum dapat menyajikan rekomendasi yang sesuai dengan beberapa jurusan karena artikel ini mempunyai tujuan untuk pemanfaatan algoritma C4.5 dalam pengembangan sistem rekomendasi jurusan SMK yang berada di wilayah Kecamatan Batang. Artikel penelitian selanjutnya diharapkan dapat mencakup objek penelitian yang lebih luas dan besar untuk mengatasi masalah sedikitnya data dari jurusan yang kurang populer.



SIMPULAN

Artikel ini menyajikan pentingnya pengembangan sistem rekomendasi pemilihan jurusan SMK yang dapat membantu siswa yang mengalami kesulitan atau hambatan dalam memilih jurusan pada sekolah lanjutan (SMK). Sudah disajikan hasil analisis dari penerapan algoritma C4.5 untuk pengembangan sistem rekomendasi jurusan SMK memiliki tingkat akurasi sebesar 83,33%. Oleh karena itu, *decision tree* yang dibangun dengan algoritma C4.5 memiliki potensi yang besar untuk menyelesaikan masalah siswa dalam memilih jurusan SMK.

DAFTAR PUSTAKA

- Amin, R. K., Indwiarti, & Sibaroni, Y. (2015). Implementation of decision tree using c4.5 algorithm in decision making of loan application by Debtor (Case study: Bank Pasar of Yogyakarta Special Region). Dalam *Proceedings of 3rd International Conference of Information and Communication Technology (ICoICT 2015)* (pp.75-80). Denpasar, Bali.
- Andriani, A. (2013). Sistem pendukung keputusan berbasis decision tree dalam pemberian beasiswa. Studi kasus: Amik BSI Yogyakarta. Dalam *Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013)* (pp.163-168). UAJY, Yogyakarta.
- Badan Standar Nasional Pendidikan. (2015). *Buku saku UN 2015*. Jakarta: Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan.
- Departemen Pendidikan Nasional. (2008). *penataan pendidikan profesional konselor dan layanan bimbingan dan konseling dalam jalur pendidikan formal* (Nomor 27). Bandung: UPI Press.
- Falentini, F. Y., Taufik, & Mudjiran. (2013). Usaha yang dilakukan siswa dalam menentukan arah pilihan karier dan hambatan-hambatan yang ditemui (Studi Deskriptif terhadap Siswa SMA N 3 Payakumbuh). *Jurnal Ilmiah Konseling*, 2(1), 310-316.
- Fartindiyah, N., & Subiyanto. (2014). Sistem pendukung keputusan peminatan SMA menggunakan metode weighted product. *Jurnal Kependidikan*, 44(2), 139-145.
- Gu, P., & Zhou, Q. (2012). Student performances prediction based on improved C4.5 decision tree algorithm. Dalam M. Elwin, Xu Linli, & T. Wenya (Eds.). *Emerging Computation and Information technologies for Education* (pp.1-8). Hangzhou, China.
- Gunarto, D. (2014). *Panduan resmi tes TPA OTO BAPPENAS*. Jakarta: Bintang Wahyu.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques* (3rd ed.). Waltham, MA: Elsevier Inc.
- Hartinah, S. (2010). *Pengembangan peserta didik*. Bandung: PT. Rafika Aditama.
- Hermawati. (2009). *Data mining*. Yogyakarta: Andi Offset.
- Jantan, H., Hamdan, A.R, & Othman, Z. A. (2010). Human talent prediction in HRM using C4.5 classification algorithm. (*IJCSE*) *International Journal on Computer Science and Engineering*, 2(8), 2526-2534.
- Kabra, R. R., & Bichkar, R. S. (2011). Performance prediction of engineering students using decision trees. *International Journal of Computer Applications*, 36(11), 8-12.
- Kamaluddin, H. (2011). Bimbingan dan konseling sekolah. *Jurnal Pendidikan dan Kebudayaan*, 17(4), 447-454.

- Kusmaryani, R. E. (2010). Penguasaan keterampilan konseling guru pembimbing di Yogyakarta. *Jurnal Kependidikan*, 40(2), 175-188.
- Larose, D. T. (2005). *Discovering knowledge in data: An introduction to data mining*. New Jersey: John Willey & Sons, Inc.
- Librarians. (2008). *Careers and vocational guidance*. New York: Ferguson Publishing.
- Muangnak, N., Pukdee, W., & Hengsanunkun, T. (2010). Classification students with learning disabilities using naïve bayes classifier and decision tree. Dalam *Proceedings of 6th International Conference on Networked Computing and Advanced Information Management (NCM)* (pp. 189-192). Seoul, Korea Utara.
- Pandey, M., & Taruna, S. (2014). A multi-level classification model pertaining to the student's academic performance prediction. *International Journal of Advances in Engineering & Technology (IJAET)*, 7(4), 1329-1341.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann.
- Setyaningrum, D., & Setiawati, D. (2013). Pengaruh persepsi siswa tentang layanan konseling individu dan persepsi tentang kompetisi kepribadian konselor terhadap minat memanfaatkan layanan bimbingan dan konseling. *Jurnal BK UNESA*, 1(1), 245-252.
- Sharma, A. K., & Sahni, A. (2011). Comparative study of classification algorithms for spam email data analysis. *International Journal on Computer Science and Engineering (IJCSE)*, 3(5), 1890-1895.
- Simamora, A., & Suwarjo, L. (2013). Manajemen bimbingan dan konseling di SMAN 4 Yogyakarta. *Jurnal Akutabilitas Manajemen Pendidikan*, 1(2), 190-204.
- Soliman, S. A., Abbas, S., & Salem, A. B. M. (2015). Classification of thrombosis collagen diseases based on C4.5 algorithm. Dalam *Proceedings of 2015 IEEE 7th International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp.131-136). Kairo, Mesir.
- Swastina, L. (2013). Penerapan algoritma C4.5 untuk penentuan jurusan mahasiswa. *Jurnal GEMA AKTUALITA*, 2(1), 93-98.
- Yadav S. K., & Pal, S. (2012). *Data mining: A prediction for performance improvement of engineering students using classification*. *World of Computer Science and Information Technology Journal (WCSIT)*, 2(2), 51-56.