

**PREDIKSI KELAYAKAN MASUK PENJURUSAN IPA SISWA
SEKOLAH MENENGAH ATAS MENGGUNAKAN C4.5
(Studi Kasus: SMA Tarakanita Gading Serpong)**

Rasi Rahwali¹, Seng Hansun², Yustinus Widya Wiratama³

^{1,2,3}Program Studi Teknik Informatika
Fakultas Teknik dan Informatika
Universitas Multimedia Nusantara

Email : rasi.rahwali@gmail.com¹, hansun@umn.ac.id², yustinus.widya@umn.ac.id³

ABSTRAK

Penjurusan Sekolah Menengah Atas adalah masa yang penting dalam menentukan masa depan pendidikan di jenjang berikutnya, yang dimulai dari kelas 10 naik ke kelas 11, dan untuk menentukan penjurusan tersebut digunakan variabel-variabel penentuan penjurusan. Untuk mendukung hasil prediksi penentuan penjurusan SMA, maka dibuatlah penelitian ini. Pada penelitian ini, algoritma yang digunakan untuk memprediksi kelayakan masuk penjurusan IPA pada Sekolah Menengah Atas, khususnya pada SMA Tarakanita Gading Serpong adalah algoritma C4.5, yang merupakan salah satu algoritma data mining yang populer digunakan. Variabel yang digunakan dalam proses penentuan keputusan adalah nilai materi pelajaran Matematika dan IPA siswa kelas XI, XII, serta hasil psikotest berupa bakat dan minat siswa. Hasil dari algoritma C4.5 ini dapat membentuk suatu pohon keputusan yang dapat menentukan hasil prediksi kelayakan masuk penjurusan IPA. Berdasarkan uji coba yang telah dilakukan, aplikasi prediksi kelayakan masuk penjurusan IPA telah berhasil diterapkan dengan hasil yang sesuai dengan kriteria, dan memiliki tingkat error 16,65%.

Kata kunci: Algoritma C4.5, Data Mining, Penjurusan SMA, Prediksi

ABSTRACT

Senior High School is an important time in determining the future of education in the next level, which starts from grade 10 up to 11th grade, and to determine the majors are used determination variables. To support the prediction results of the determination of SMA majors, then this study was made. In this research, the algorithm used to predict the feasibility of entering IPA in High School, especially at Tarakanita Gading Serpong SMA is algorithm C4.5, which is one of the popular data mining algorithm used. The variables used in the decision making process are the value of Mathematics and Science subject of the students of grade XI, XII, and the result of psychotest in the form of students' talents and interests. The result of C4.5 algorithm can form a decision tree that can determine the prediction result of IPA majors. Based on the experiments that have been performed, the feasibility app for IPA majors has been successfully applied with results that match the criteria, and has an error rate of 16.65%.

Keywords: Algorithm C4.5, Data Mining, Senior High School, Prediction

PENDAHULUAN

Sekolah Menengah Atas Tarakanita Gading Serpong Tangerang merupakan salah satu sekolah menengah atas yang sukses menarik banyak siswa di setiap periodenya dan telah memiliki akreditasi A (<http://sma-gs.tarakanita.or.id/about.html>). Selain itu juga Sekolah Menengah Atas Tarakanita Gading Serpong Tangerang memiliki siklus dalam pengambilan jurusan dari kelas X ke kelas XI khususnya dalam penjurusan Ilmu Pengetahuan Alam yaitu pada saat kelas X siswa mempunyai kesempatan satu tahun (semester 1 dan 2) yang mana harus memiliki nilai mata pelajaran matematika dan Ilmu Pengetahuan Alam lebih dari kriteria ketuntasan minimal yang ditentukan yaitu 75.

Nilai 75 tersebut diambil dari jumlah nilai semester 1 dan 2 yang telah dibagi dua menjadi nilai rata-rata, lalu siswa diberikan angket untuk mengetahui minatnya yang mana angket tersebut diketahui oleh orang tua siswa, kemudian siswa juga dapat layanan konsultasi dengan psikolog untuk mendapatkan hasil psikotest dan hasil psikotest menghasilkan bakat, yang mana bakat tersebut diberikan kepada guru bimbingan konseling. Terjadi masalah bila ada siswa yang memiliki nilai kurang dari kriteria ketuntasan minimal, tetapi bakat atau minat menunjukkan di Ilmu Pengetahuan Alam. Kemudian pihak sekolah mengadakan rapat dan pada saat rapat tidak jarang mengalami kendala untuk menentukan penjurusan Ilmu Pengetahuan Alam untuk siswa yang memiliki masalah seperti ini.

Cara menyelesaikan masalah penjurusan tersebut perlu dibuat adanya suatu sistem pendukung keputusan untuk membantu memprediksi kelayakan siswa dalam pengambilan penjurusan Ilmu Pengetahuan Alam. *Data mining* adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Di sisi lain, arti dari *data mining* adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa memperbaiki pengambilan keputusan di masa depan (Swastina, 2013).

Swastina (2013) dalam penelitiannya yang berjudul Penerapan Algoritma C4.5 untuk Penentuan Jurusan Mahasiswa mengatakan bahwa algoritma C4.5 memiliki tingkat akurasi lebih tinggi daripada algoritma Naive Bayes dengan perbandingan tingkat akurasinya adalah 82,64% dan 66,36%. Dalam penelitian tersebut, Swastina (2013) menggunakan variabel data mahasiswa dengan atribut NIM, Nama, Tanggal Lahir, Asal Sekolah, Nilai UN, Jurusan yang dipilih, serta data Indeks Prestasi Kumulatif (IPK) mahasiswa semester 1 dan semester 2.

Algoritma C4.5 adalah contoh algoritma dari *data mining* yang digunakan untuk membuat keputusan, dan keputusan itu berbentuk pohon (*tree*). Proses *data mining* adalah untuk mengekstraksi informasi dari sekumpulan *dataset* yang besar dan diubah menjadi struktur yang dapat dimengerti dan digunakan untuk masa depan, dan *data mining* itu digunakan terutama untuk membuat model klasifikasi dan prediksi (Agrawal dkk., 2008). Selanjutnya untuk mengolah sekumpulan data *history* yang besar digunakan algoritma C4.5 untuk mendapatkan *rule* (aturan) agar dapat menentukan atau memprediksi keputusan untuk masuk dalam penjurusan Ilmu Pengetahuan Alam. Apabila dihubungkan dengan masalah yang ada, nantinya algoritma ini akan memproses atribut seperti nilai akademik, hasil psikotest, dan minat siswa yang digunakan untuk memprediksi kelayakan penjurusan Ilmu Pengetahuan Alam siswa Sekolah Menengah Atas Tarakanita Gading Serpong Tangerang.

TELAAH LITERATUR

1. Pohon Keputusan

Pohon keputusan atau *decision tree* adalah model prediksi menggunakan struktur pohon atau hirarki. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun juga mudah untuk dipahami sehingga *decision tree* ini merupakan metode klasifikasi yang paling populer digunakan (Ginting dkk., 2014).

Tipikal dari *decision tree* adalah gabungan dari *internal nodes*, *edges*, dan *leaf nodes*. Setiap *internal node* dalam *decision tree* merepresentasikan sebuah *test* dari atribut atau subset atribut, dan setiap *edges* menandakan nilai spesifik

atau *range* nilai dari *input* atributnya. Setiap *leaf node* adalah sebuah sambungan *node* dari *tree* berupa *class label* (Dai dkk., 2014).

2. Data Mining

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menemukan keteraturan, pola, dan hubungan dalam dataset berukuran besar. Maksud dari pengertian ini adalah proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data yang besar. Karakteristik *data mining* adalah sebagai berikut.

- a. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. *Data mining* menggunakan data yang besar. Data yang besar digunakan untuk membuat hasil lebih dipercaya.
- c. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Secara umum ada dua jenis metode pada *data mining*, metode *predictive* dan metode *descriptive*. Metode *predictive* adalah proses untuk menentukan pola dari data yang menggunakan beberapa variabel untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya. Teknik yang termasuk dalam *predictive mining* antara lain klasifikasi, regresi, dan deviasi. Metode *descriptive* adalah proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data. Teknik *data mining* yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining* (Ginting dkk., 2014).

3. Algoritma C4.5

Algoritma C4.5 digunakan untuk membangun pohon keputusan. Bening (Bening, 2014) berpendapat bahwa algoritma C4.5 merupakan salah satu algoritma *machine learning*. Dengan algoritma ini, mesin (komputer) akan memberikan sekelompok data untuk dipelajari yang disebut *learning dataset*. Kemudian hasil dari pembelajaran tersebut akan digunakan untuk mengolah data-data yang baru disebut *test dataset*. Karena algoritma C4.5 digunakan untuk

melakukan klasifikasi, jadi hasil dari pengolahan *test dataset* berupa pengelompokan data ke dalam kelas-kelasnya.

Andriani (Andriani, 2012) berpendapat bahwa ada 4 tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5.

1. Mempersiapkan *data training*. *Data training* biasanya diambil dari data *history* yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus (1).

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

Keterangan :

Entropy: Pengukuran yang berdasarkan kemungkinan yang digunakan untuk menghitung jumlah ketidaktentuan.

S : Himpunan kasus

n : Jumlah partisi S

p_i : Proporsi S_i terhadap S

Kemudian hitung nilai *gain* menggunakan rumus (2):

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

Keterangan :

Gain: Salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*.

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus dalam S

3. Ulangi langkah ke-2 dan langkah ke-3 hingga semua *record* terpartisi.
4. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong.

Marwana (2013) berpendapat bahwa kunci pencarian *entropy* ada dua, yaitu sebagai berikut.

1. Jika di antara kolom “benar” atau “tidak” ada yang bernilai 0 (nol) maka *entropy*-nya dipastikan juga bernilai 0 (nol).
2. Jika kolom “benar” dan “tidak” mempunyai nilai yang sama maka *entropy*-nya dipastikan juga bernilai 1 (satu).

METODOLOGI PENELITIAN

Metodologi penelitian yang diterapkan dalam penelitian yang dilakukan meliputi tahapan-tahapan berikut.

1. Studi Literatur

Bertujuan untuk membantu dalam menemukan teori-teori yang berkaitan dengan topik yang dibahas dalam penelitian ini, baik mengenai pohon keputusan, *data mining*, dan algoritma C4.5. Referensi yang digunakan pada studi literatur bersumber dari buku, artikel, jurnal ilmiah, serta sumber lainnya.

2. Wawancara

Merupakan pencarian informasi dengan pemberian pertanyaan langsung, sehingga diperoleh informasi yang dibutuhkan. Dalam penelitian ini, narasumbernya adalah Kepala Sekolah SMA Tarakanita Gading Serpong, bernama adalah Ibu HY. Shinta Natalini B., S.Si. Wawancara dilakukan untuk mendapatkan alur penerimaan siswa penjurusan dan data-data yang terkait di dalamnya.

3. Analisis Sistem

Analisis sistem merupakan tahap dimana berbagai kebutuhan yang diperlukan dalam pembangunan sistem dianalisis. Analisis tersebut berupa *platform* pengembangan sistem, basis data yang digunakan, dan perancangan sistemnya.

4. Perancangan Sistem

Perancangan yang dilakukan pada tahap ini adalah perancangan tampilan atau antarmuka aplikasi serta fitur-fitur yang akan disediakan pada aplikasi ini yang akan dijabarkan lebih lanjut pada bab berikutnya.

5. Pemrograman Sistem

Setelah proses perancangan dilalui, maka pemrograman sistem dilakukan. Pemrograman sistem adalah realisasi atau pembuatan sistem yang telah dirancang sebelumnya, yakni aplikasi prediksi kelayakan masuk penjurusan IPA bagi siswa SMA.

6. Uji Coba dan Evaluasi

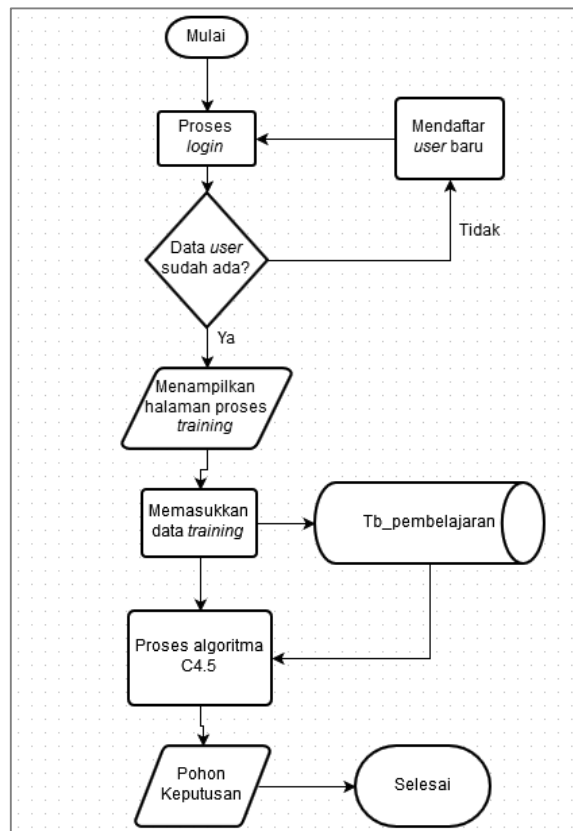
Setelah proses pemrograman sistem selesai dilakukan, maka proses selanjutnya yang dilakukan adalah uji coba terhadap aplikasi yang dibuat. Selain itu dibuat juga evaluasi atas hasil dari uji coba tersebut yang dalam penelitian ini menggunakan metode *k-fold cross validation* untuk mengetahui tingkat akurasi prediksi yang dihasilkan oleh aplikasi.

7. Penulisan Laporan dan Publikasi

Penulisan laporan bertujuan sebagai dokumentasi dari sebuah penelitian yang telah dilakukan sehingga dapat memberikan informasi bagi penelitian sejenis. Hasil penelitian juga dipublikasikan dalam bentuk makalah ilmiah sebagai acuan penelitian selanjutnya.

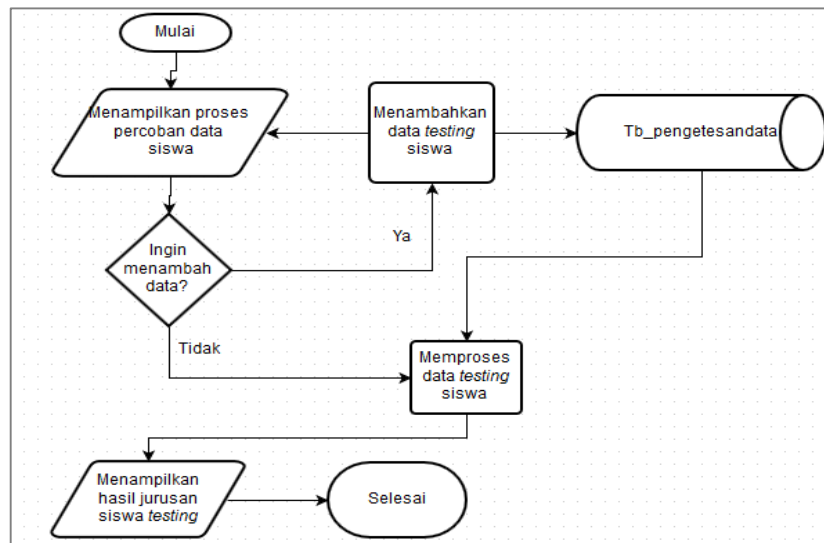
DESKRIPSI SISTEM

Flowchart adalah cara penyajian visual aliran data melalui sistem informasi, operasi dilakukan dalam sistem dan urutan dimana mereka dilakukan. *Flowchart* dari aplikasi ini ditunjukkan pada Gambar 1, 2, dan 3.



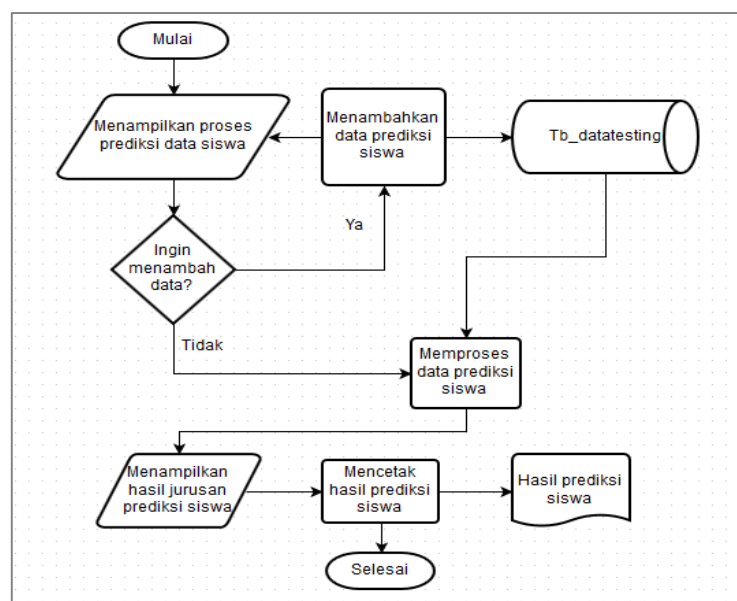
Gambar 1. *Flowchart* Pembelajaran Aplikasi

Gambar 1 merupakan *flowchart* pembelajaran yang menunjukkan cara aplikasi menghasilkan pohon keputusan, dimulai dari “proses *login*”, kemudian diperiksa dan masuk ke proses pemilihan “data *user* sudah ada?” jika “tidak”, masuk proses “mendaftar” lalu mengulangi masuk, tetapi jika “ya” masuk dan “menampilkan proses training”, selanjutnya memasukkan data ke proses “memasukkan data *training*”, kemudian data *training* tersebut disimpan ke dalam *database*, dan data tersebut akan diolah dengan cara masuk ke “proses algoritma C4.5”, lalu setelah data tersebut diolah terbentuk “pohon keputusan”, kemudian “selesai”.



Gambar 2. Flowchart Percobaan Data Aplikasi

Flowchart percobaan data yang diperlihatkan pada Gambar 2 menunjukkan cara aplikasi dalam mendapatkan hasil percobaan data berupa jurusan siswa, dimulai dari “menampilkan proses percobaan data siswa”, kemudian masuk ke pemilihan “ingin menambah data?” jika “ya” masuk ke proses “menambahkan data *testing* siswa” dan data tersebut disimpan ke dalam *database* lalu mengulangi ke “menampilkan proses percobaan data siswa”, namun jika “tidak” masuk ke “memproses data *testing* siswa”, setelah itu muncul *output* berupa “menampilkan hasil jurusan siswa *testing*”, kemudian “selesai”.



Gambar 3. Flowchart Prediksi Data Aplikasi

Gambar 3 merupakan *flowchart* prediksi data yang memperlihatkan cara mendapatkan hasil prediksi data berupa jurusan siswa, dimulai dari masuk “menampilkan proses prediksi data siswa”, kemudian masuk proses pemilihan “ingin menambah data?” jika “ya” masuk ke proses “menambahkan data prediksi siswa”, dan data tersebut disimpan ke dalam *database*, kemudian mengulangi ke “menampilkan proses prediksi siswa”, tetapi jika “tidak” masuk ke “memproses data prediksi siswa”, kemudian setelah data tersebut diproses akan menampilkan *output* “menampilkan hasil jurusan prediksi siswa”, setelah itu akan dilanjutkan ke proses “mencetak hasil prediksi siswa”, lalu proses mencetak hasil akan menghasilkan *output* satu *document* yaitu “hasil prediksi siswa”, dan kemudian dari proses “mencetak hasil” semua proses telah terpenuhi dan selesai.

HASIL DAN PEMBAHASAN

Setelah melalui tahap perancangan dan pembangunan sistem, maka aplikasi dapat diterapkan dan diujicoba oleh pihak Sekolah. Hasil implementasi diperlihatkan pada Gambar 4 yang merupakan halaman pembelajaran (*learning*) dimana user dapat menambah data baru, menambah data dengan data yang sudah ada, melihat hasil *tree* yang sudah melewati proses *learning*, melihat bantuan, dan menggunakan *link* ke halaman percobaan data atau ke prediksi data.

The screenshot shows a web application window titled "Aplikasi Prediksi Penjurusan IPA SMA Tarakanita Gading Serpong - Halaman Data Pembelajaran". The interface includes the school logo and name, a table of student data, and a decision tree visualization on the right side. The table has columns for NIS, Nama, Matematika, Fisika, Biologi, Kimia, Bakat, Minat, and Jurusan. The decision tree shows a root node "Bakat" with branches for "IPA : IPA" and "IPS & Kimia". Further sub-nodes include "nilai < 75 : IPS" and "nilai >= 75 & Matematika", with further sub-nodes for "nilai < 75 : IPA" and "nilai >= 75 : IPS".

NIS	Nama	Matematika	Fisika	Biologi	Kimia	Bakat	Minat	Jurusan
11120110...	Riki	84.185	81.25	85.25	83.25	IPA	IPA	IPA
11120110...	Maiden	70.0	72.25	76.0	73.75	IPA	IPA	IPA
11120110...	Medusa	78.835	73.75	74.25	78.5	IPA	IPA	IPA
11120110...	Enigma	72.665	74.5	80.0	72.5	IPA	IPA	IPA
11120110...	Centa	74.335	75.75	70.0	77.0	IPA	IPA	IPA
11120110...	Ember	75.0	76.0	77.5	74.0	IPA	IPA	IPA
11120110...	Storm	83.835	81.25	83.75	77.75	IPA	IPA	IPA
11120110...	Antimage	72.835	72.75	71.0	75.0	IPA	IPA	IPA
11120110...	Tiny	85.5	82.5	80.0	81.5	IPA	IPA	IPA
11120110...	Legion	71.83	71.25	70.0	70.75	IPA	IPS	IPA
11120110...	Skywarth	70.0	70.75	72.25	71.25	IPA	IPA	IPA
11120110...	Shamman	83.835	83.75	82.75	82.25	IPA	IPA	IPA
11120110...	Dazzle	84.185	81.25	85.25	83.25	IPA	IPA	IPA
11120110...	Bounty	83.835	81.25	83.75	77.75	IPA	IPA	IPA

Gambar 4. Halaman Pembelajaran (*Learning*)

Setelah tahap pembelajaran selesai dilakukan, *tree* yang terbentuk dapat digunakan untuk mengevaluasi tingkat akurasi aplikasi prediksi kelayakan masuk penjurusan IPA dengan C4.5. Gambar 5 memperlihatkan halaman percobaan (*testing*) dari data yang telah tersimpan di *database* dengan hasil prediksinya masing-masing, sementara Gambar 6 memperlihatkan halaman prediksi untuk data baru yang belum ada di *database*.

NIS	Nama	Matematika	Fisika	Biologi	Kimia	Bakat	Minat	Jurusan
11110110334	Russ	75.0	74.75	73.25	74.0	IPS	IPS	IPS
11110110335	Morton	80.165	79.0	78.0	78.0	IPA	IPA	IPS
11110110336	Federico	85.67	77.25	81.0	78.25	IPS	IPS	IPS
11110110337	Adam	76.17	77.75	80.75	73.25	IPS	IPS	IPS
11110110338	Miriam	91.5	85.0	85.25	87.75	IPA	IPA	IPS
11110110339	Enrico	87.33	81.0	75.5	81.0	IPA	IPA	IPS
11110110340	Will	85.0	80.75	80.0	80.0	IPA	IPA	IPS
11110110341	Marcella	74.5	76.5	78.75	74.0	IPS	IPS	IPS
11110110342	Joseph	65.33	76.0	76.5	80.75	IPA	IPA	IPS
11110110343	Tina	69.665	68.75	69.25	70.25	IPS	IPS	IPS
11110110344	Richard	95.5	83.25	85.0	75.75	IPA	IPA	IPS
11110110345	Henry	77.17	74.75	76.0	73.25	IPS	IPS	IPS
11110110346	Finney	79.17	78.5	80.75	78.0	IPA	IPA	IPS
11110110347	Fields	76.0	77.0	80.0	77.0	IPS	IPA	IPS
11110110348	Frederick	73.335	73.0	75.5	75.75	IPS	IPS	IPS

Gambar 5. Halaman Percobaan (*Testing*)

NIS	Nama	Matematika	Fisika	Biologi	Kimia	Bakat	Minat	IQ	Jurusan
11110110700	Joko	86.7	78.7	87.7	78.367	IPA	IPA	120	IPA
11110110701	Subur	78.0	77.0	76.0	76.0	IPS	IPS	115	IPS
11110110702	Enik	88.0	88.0	88.0	88.0	IPS	IPS	112	IPS
11110110703	Lewi	80.0	70.0	80.0	73.0	IPA	IPS	114	IPS

Gambar 6. Halaman Prediksi untuk Data Baru

Cross-validasi (*cross validation*) atau yang sering disebut dengan estimasi rotasi merupakan teknik validasi model untuk mengukur seberapa tingkat optimal

hasil analisis, selain itu cross-validasi juga merupakan teknik komposisi dalam penentuan banyaknya *data training* dan *data testing* yang akan digunakan. Salah satu metode *cross validation* adalah metode *k-fold*. Dalam metode *k-fold*, data disegmentasi secara *random* ke dalam k partisi yang berukuran sama. Selama proses, salah satu dari partisi dipilih untuk menjadi data *testing*, sedangkan sisanya digunakan untuk *training*. Kemudian rata-rata *error* dari setiap percobaan k dihitung. Variasi dari total *error* bergantung pada *data training* yang digunakan dan *data testing* (Rodriguez dkk., 2010). Penggunaan nilai k terbaik untuk menguji validitas *data training* adalah sebesar sepuluh (Breiman dkk., 1992).

Tabel 1. *Cross validation*

Data		Error	% Error
Training	Testing		
433	48	4	8,16
433	48	6	12,50
433	48	0	0,00
433	48	9	18,75
433	48	11	22,92
433	48	10	20,83
433	48	12	25,00
433	48	7	14,58
433	48	13	27,08
433	48	8	16,67
Total Data			481
Total % Error			166,50
K cross validation			10
% Error cross validation			16,65

Tabel 1 menggunakan k-partisi bernilai sepuluh, sehingga total keseluruhan data dibagi menjadi sepuluh partisi; sembilan partisi digunakan untuk *data training*, dan salah satunya digunakan sebagai *data testing*. Total data yang digunakan adalah 481 yang terdiri dari tahun ajaran 2011-2013. Data *training* pertama diambil dari data ke-49, dan data *testing* pertama diambil dari data ke-1 sampai ke-48, data *training* kedua diambil dari data ke-1 sampai data ke-48 dan dari data ke-98 sampai data ke-481, data *testing* kedua diambil dari data ke-49 sampai data ke-97, dan selanjutnya seperti itu terus sampai data *training* dan data *testing* kesepuluh.

Prosentase *error* setiap *data testing*
 $= (\text{error} / \text{data testing}) * 100$

Prosentase *k cross validation*
 $= \sum \text{Prosentase error setiap data testing} / k\text{-fold}$
 $= 166,50 \% / 10$
 $= 16,65 \%$

Dari hasil yang didapatkan menunjukkan bahwa dengan menggunakan *k* partisinya sepuluh, maka prosentase *error 10-cross validation* yang didapatkan adalah 16,65%. Hasil ini menunjukkan tingkat akurasi aplikasi yang dibangun adalah 83,35% yang tergolong tinggi. Tingkat akurasi yang diperoleh ini sudah sesuai dengan hipotesis awal peneliti bahwa algoritma C4.5 dapat digunakan untuk memprediksi kelayakan masuk penjurusan IPA bagi siswa SMA. Hasil ini juga lebih tinggi dibandingkan hasil penelitian Swastina (2013) sebesar 82,64% untuk penentuan jurusan mahasiswa menggunakan algoritma C4.5.

SIMPULAN

Simpulan yang diperoleh dari hasil penelitian ini adalah implementasi *data mining* algoritma C4.5 dalam memprediksi kelayakan penjurusan IPA pada Sekolah Menengah Atas, khususnya pada Sekolah Menengah Atas Tarakanita Gading Serpong berhasil diimplementasikan dengan baik. Lalu dari hasil penelitian ini, aplikasi sistem prediksi penjurusan Ilmu Pengetahuan Alam (IPA) telah berhasil dibangun dengan menggunakan algoritma C4.5, maka dari itu algoritma C4.5 dinilai berhasil dalam memprediksi penjurusan Ilmu Pengetahuan Alam (IPA) pada jenjang Sekolah Menengah Atas, dan berdasarkan hasil uji coba validasi dengan menggunakan *10-cross validation* memiliki nilai tingkat *error* sebesar 16,65%.

Selain itu, dari hasil penelitian yang dilakukan, maka saran yang dapat disampaikan adalah menganalisis atribut lain seperti nilai Bahasa Indonesia atau Bahasa Inggris yang dapat mempengaruhi hasil penjurusan IPA karena nilai tersebut pada masa yang akan datang dapat membantu siswa dalam proses belajar di dalam kelas dan dapat menambah basis pengetahuan (*knowledge*) untuk algoritma C4.5 agar mendapatkan hasil yang lebih akurat.

DAFTAR PUSTAKA

- Agrawal, Gaurav L, Prof. Hitesh Gupta. 2008. "Optimization of C4.5 Decision Tree Algorithm for Data Mining Application". *International Journal of Emerging Technology and Advanced Engineering*, Volume 3, Issue 3, March 2013.
- Andriani, Anik. 2012. "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Drop Out". *Seminar Nasional Matematika 2012*.
- Bening, Rudi. 2014. "Belajar Mudah Algoritma Data Mining : C4.5". Tersedia dalam <https://www.academia.edu>. Diakses 16 Februari 2015.
- Breiman, Leo, Philip Spector. 1992. "Submodel Selection and Evaluation in Regression. The X-Random Case". *International Statistical Review*. Vol 60 No 3, 1992.
- Dai, Wei, Wei Ji. 2014. "A Mapreduce Implementasion of C4.5 Decision Tree Algorithm". *International Journal of Database Theory and Application*. Vol 7 No 1, 2014.
- Ginting, Selvia Lorena Br, Wendi Zarman, Ida Hamidah. 2014. "Analisis dan Penerapan Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik". *Prosiding Seminar Nasional Aplikasi Sains & Teknologi*, 15 November 2014.
- Marwana. 2013. "Algoritma C4.5 untuk Simulasi Prediksi Kemenangan Dalam Pertandingan Sepakbola". Tersedia dalam <http://jim.stimednp.ac.id/>. Diakses 16 Februari 2015.
- Rodriguez, Juan Diego, Aritz Perez, Jose Antonio Lozano. 2010. "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation". *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 32 No 3. 2010.
- Swastina, Liliana. 2013. "Penerapan Algoritma C4.5 untuk Penentuan Jurusan Mahasiswa". *Jurnal GEMA Aktualita*, Vol 2 No 1. 2013.