

# Penerapan Algoritma C4.5 Dalam Memprediksi Kesiapan Siswa SMP IT PAPB Semarang Menghadapi Ujian Nasional

## *Implementation of C4.5 Algorithm to Predict the Ability of Junior High School Student to Face National Exam*

Nadia Zulfa Rahma<sup>1</sup> Andik Setyono<sup>2</sup>

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro  
Jl. Nakula 1, No. 5-11 Semarang, Kode Pos 50131, Telp. (024) 3515261, Fax: 3569684  
[nadiazulfa@gmail.com](mailto:nadiazulfa@gmail.com)<sup>1</sup>, [andik.setyono@dsn.dinus.ac.id](mailto:andik.setyono@dsn.dinus.ac.id)<sup>2</sup>

### **Abstrak**

Kelulusan siswa dalam menghadapi ujian nasional dapat dijadikan sebagai tolak ukur sejauh mana siswa memahami materi-materi yang di dapat saat di sekolah. Untuk mendapatkan nilai tuntas pada ujian nasional, sekolah mengadakan latihan try out yang bertujuan agar siswa lebih siap menghadapi ujian. Dalam pelaksanaan try out, tidak semua siswa dapat menyelesaikan soal-soal dengan benar. Hal tersebut berdampak pada hasil nilai try out yang buruk. Maka dari itu, sekolah membutuhkan sebuah data mining dengan metode klasifikasi yang dapat membantu memprediksi kesiapan siswa menghadapi ujian nasional. Metode tersebut mengolah data yang memiliki perbedaan atribut kemudian disusun ke dalam kategori yang sesuai. Data tersebut di prediksi menggunakan algoritma Decision Tree, yang merupakan salah satu machine learning menggunakan perhitungan probabilitas. Penggunaan algoritma ini didukung dengan simulasi yang dilakukan menggunakan aplikasi RapidMiner dan mendapatkan nilai akurasi sebesar 99,48%. Dari hasil simulasi tersebut, diolah kembali menjadi sebuah aplikasi yang ditujukan untuk membantu pihak sekolah. Untuk menguji kegunaan dari aplikasi tersebut dilakukan penyebaran kuesioner berjumlah 10 soal ke 20 guru dan memperoleh hasil index 83,3% yang berarti memuaskan. Dengan begitu, aplikasi tersebut berguna bagi pihak sekolah.

**Kata Kunci**— Ujian Nasional, Tryout, Data Mining, Klasifikasi, Decision Tree.

### **Abstract**

Graduate students in national exam can be used as a benchmark the extent to which students understand the material in the can while in school. To get the complete value on the national exam, the school held a workout tryouts are intended to make students better prepared for the exam. In the implementation of the tryout, not all students can solve problems correctly. It has an impact on the results of the tryout a bad value. Therefore, schools need a data mining with classification methods that can help predict a student's readiness to face national exam. The method of processing data that have different attributes then organized into appropriate categories. The data in the prediction using Decision Tree algorithm, which is one of the machine learning using probability calculations. The use of this algorithm is supported by simulations were done using the application RapidMiner and scored an accuracy of 99.48%. From the simulation results, reprocessed into an application that is intended to help the school. To test the usefulness of the application is carried out questionnaires about the amount to 10 to 20 teachers and obtained 83.3% of the index, which means satisfactory. By doing so, the application is useful for those schools.

**Keywords**— National Exam, Tryout, Data Mining, Classification, Decision Tree.

---

## 1. PENDAHULUAN

Tryout merupakan tahap evaluasi yang dilakukan untuk menghadapi ujian nasional sesungguhnya. Soal-soal yang terdapat di dalam *tryout* mengacu pada materi dalam ujian nasional. Penyusunan soal diperoleh dari materi kelas 1, 2 dan 3, dan tidak menutup kemungkinan di dapat dari media massa yang masih berkaitan dengan kisi-kisi ujian nasional tersebut. Namun, siswa tidak semuanya dapat menyelesaikan soal-soal *tryout* dengan mudah. Maka dari itu *tryout* dilakukan berulang kali sampai mendekati masa ujian nasional agar siswa terbiasa menghadapi soal-soal yang sama dan mudah menjawabnya [1].

Siswa yang mengalami kesulitan dalam mengerjakan *tryout* perlu mendapat perhatian dari pihak sekolah. Perhatian yang dimaksud yaitu dengan mengadakan kelas tambahan agar mereka lebih siap menghadapi ujian nasional. Guna meminimalisir kejadian tersebut, penulis memberikan solusi berupa sistem pendukung keputusan dengan menggunakan pendekatan metode data mining. Data mining yang digunakan terfokus pada metode klasifikasi. Metode klasifikasi disarankan sebagai solusi dari masalah ini. Dengan menggunakan metode klasifikasi, pengambilan keputusan menjadi mudah dan sederhana. Dengan demikian, keputusan akan fokus terhadap solusi saja.

Penelitian sebelumnya yang dilakukan oleh Kamagi dan Hansun, telah membahas tentang prediksi tingkat kelulusan mahasiswa menggunakan algoritma C4.5. Penelitian tersebut menunjukkan bahwa hasil prediksi kelulusan menggunakan algoritma C4.5 dengan atribut IP semester VI sebesar 87.5% dari 60 data *training* dan 40 data *testing* [2]. Penelitian selanjutnya dilakukan oleh Novianti dan Aziz dengan topik menampilkan laporan hasil nilai akhir mahasiswa menggunakan algoritma *Decision Tree* C4.5, menghasilkan evaluasi dan validasi dengan *Cross Validation* menunjukkan tingkat akurasi sebesar 100% [3]. Kemudian, penelitian selanjutnya yang dilakukan oleh Ainurrofiq mengangkat tema klasifikasi data kesiapan siswa dalam menghadapi ujian nasional menggunakan algoritma *Naive Bayes* menghasilkan akurasi sebesar 83,8798% dan error rate 16,1202% [1].

Penulis melakukan penelitian ini sebagai pengembangan dari penelitian sebelumnya yang telah disebutkan di atas terutama yang dilakukan oleh Ainurrafiq [1], dimana memiliki topik berkaitan dengan kesiapan siswa dalam menghadapi ujian nasional yaitu dengan melakukan penambahan variable dan algoritma yang berbeda. Penelitian yang penulis kembangkan memiliki beberapa perbedaan dengan penelitian sebelumnya. Atribut yang digunakan dalam penelitian ini yaitu data *tryout* selama 3 tahun kebelakang. Kemudian, label yang digunakan ada 2 macam yaitu siap dan belum siap. Perbedaan selanjutnya dari segi data yang digunakan berupa data *tryout* siswa SMP (Sekolah Menengah Pertama).

## 2. METODE PENELITIAN

### 2.1 Prosedur Pengumpulan Data

Dataset yang digunakan diperoleh berdasarkan data primer yakni diperoleh secara langsung dari narasumber [4]. Narasumber dalam penelitian ini yaitu Kepala sekolah SMP Islam Terpadu PAB Semarang dan guru yang ditunjuk untuk mengurus persiapan siswa menghadapi ujian nasional. Data bersifat kuantitatif karena informasi didalamnya berupa angka atau bilangan.

Tabel 1. Contoh Dataset Nilai Tryout

NO	KELAS	NAMA	NO INDUK	B.IND	MAT	B. ING	IPA	JUML AH	RATA- RATA	KRITE RIA
1	IX A	AHMAD BAIHAQI HAKIM	0974	4,00	5,00	2,80	4,70	16,50	4,13	TL

2	IX A	AISYAH IZZA TAMAMILA	0944	6,20	8,00	7,60	7,30	29,10	7,28	TL
3	IX A	AL JAILANIY JUSUF DIASTAMA	0945	6,80	3,30	5,80	4,30	20,20	5,05	TL
4	IX A	ALFAREL SAMUDRA HARIYANTO	1002	7,00	5,00	5,40	4,70	22,10	5,53	TL
5	IX A	ALFIN TOTI KHOIRI	1003	4,40	3,00	4,20	4,30	15,90	3,98	TL
6	IX A	CHALIMATUS SA'DIYYAH	0918	8,00	6,30	4,80	4,70	23,80	5,95	TL
7	IX A	DAFA BAYU PRATAMA	0981	5,80	6,70	3,00	3,30	18,80	4,70	TL
8	IX A	DANY SHAFFAN DAFFA	0982	6,20	6,70	6,80	5,00	24,70	6,18	TL
9	IX A	DEVI RIZKY ARDYAPUTRI PARAMITA	1009	7,80	8,70	6,60	5,30	28,40	7,10	TL
10	IX A	ELANG DIWA WIJAYA	1041	6,40	2,00	3,60	2,70	14,70	3,68	TL

## 2.2 Desain Penelitian

Penulis mengaplikasikan model standarisasi data mining yaitu *CRISP-DM* (*Cross Industry Standart Process for Data Mining*), dengan beberapa tahapan sebagai berikut [5]:

### 1. Pemahaman Bisnis

Penulis mengedepankan penentuan tujuan proyek dan kebutuhan lengkap dalam memahami proses bisnis atau tujuan bisnis mengenai penelitian tersebut. Setelah itu, penelitian tersebut mampu mengolah sebuah data yang didapat dari SMP Islam Terpadu PAPB Semarang menjadi sebuah formula atau rules.

### 2. Pemahaman Data

Penelitian ini menggunakan sumber data primer (sumber yang berasal langsung dari sekolah), yaitu data hasil *try out* siswa kelas IX SMP Islam Terpadu PAPB Semarang tahun ajaran 2013/2014 – 2015/2016.

### 3. Pengolahan Data

Pengolahan data berlangsung dari data mentah menjadi data bentuk hasil normalisasi. Berikut ini penjelasan dalam tahapan pengolahan data [6] :

- Data cleaning and data selection*, merupakan tahapan awal dalam memproses data mining. Kegiatan ini bertujuan untuk membuang atribut yang tidak dibutuhkan agar lebih efisien, seperti jumlah dan keterangan.
- Data integration and data transformation*, merupakan tahap selanjutnya dalam proses data mining yaitu memindahkan database dan memproses transformasi data. Pada data *try out* kelas IX, data dari file yang berbeda seperti data pada tahun ajaran 2013/2014, 2014/2015 dan 2015/2016 disusun menjadi satu kesatuan dalam *file* yang sama.
- Data Reduction*, yaitu penggunaan data sesuai dengan atribut dan jumlah *record* yang dibutuhkan. Jadi, terdapat beberapa bagian *field* yang perlu dihilangkan karena tidak lengkapnya data atau kurang relevan dalam penelitian.

### 4. Pemodelan

Tahap awal dilakukan pemrosesan dataset seperti konversi data dan *replacing data*, kemudian dilakukan pemodelan menggunakan algoritma C4.5.

### 5. Evaluasi

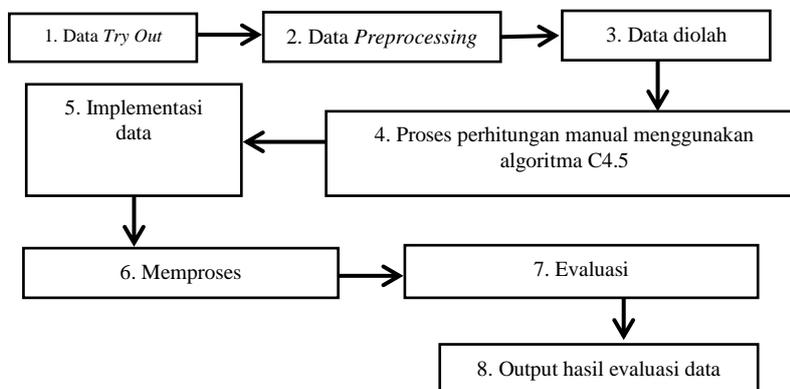
Dilakukan validasi, yaitu pengukuran akurasi dari hasil sebuah pemodelan algoritma yang digunakan dengan memanfaatkan *tools RapidMiner*.

### 6. Penyebaran

Penulis menghasilkan penelitian yang membentuk *Decision Support System* (DSS). Hasil dari analisa diterapkan sebagai dasar perancangan pada sistem pengambil keputusan untuk mengidentifikasi dan mengklasifikasi kesiapan siswa dalam menghadapi ujian nasional.

### 2.3 Metode yang Diusulkan

Desain metode pada penelitian ini menggunakan algoritma C4.5 untuk melakukan perhitungan klasifikasi dan prediksi



Gambar 1. Metode yang diusulkan

Berdasarkan gambar 1 di atas, maka langkah-langkah metode dapat dijelaskan seperti dibawah:

1. Data yang akan diimplementasikan sebagai data penelitian adalah hasil nilai *try out* kelas IX SMP Islam Terpadu PAPB Semarang tahun ajaran 2013/2014, 2014/2015 dan 2015/2016.
2. Data diolah dengan cara pre-processing, yaitu dengan seleksi data, transformasi data, dan reduksi data. Pengolahan tersebut dilakukan supaya data terhindar dari *missing value* (data yang kurang lengkap).
3. Tahap selanjutnya setelah dilakukan data pre-processing, maka akan menghasilkan data *training* yang siap diolah di tahap berikutnya.
4. Pada tahap ini dilakukan perhitungan manual dengan mulai menghitung *entropy* dan *gain* berdasarkan data pre-processing.
5. Hasil dari perhitungan tersebut di selaraskan dengan hasil yang dilakukan dalam *tools RapidMiner*.
6. Dalam *tools* tersebut dilakukan proses mengolah data sesuai dengan algoritma yang digunakan. Algoritma tersebut adalah C4.5.
7. Dilakukan simulasi untuk melihat prosentase akurasi data yang di dapat dari algoritma tersebut.
8. Output akan memperlihatkan hasil dari pengolahan data tersebut dalam aplikasi *RapidMiner*. Hasil dari *RapidMiner* itulah yang akan digunakan sebagai rumus dalam pembuatan aplikasi Java Netbeans

## 3. HASIL DAN PEMBAHASAN

### 3.1 Perhitungan Manual

Berikut ini perhitungan manual penerapan algoritma C4.5 untuk memprediksi kesiapan siswa SMP Islam Terpadu PAPB Semarang dalam menghadapi ujian nasional.

#### 1. Mengambil sampel data acak

Sampel data acak digunakan sebagai data awal untuk melakukan perhitungan manual algoritma C4.5. Berikut ini data training yang digunakan:

Tabel 2. Data Training Tahap 1

NO INDUK	B. IND	MAT	B. ING	IPA	KRITERIA
0944	3	3	4	2	BELUM SIAP
0945	3	3	1	3	BELUM SIAP
1002	2	1	3	2	BELUM SIAP
1003	1	3	2	1	BELUM SIAP
0918	4	3	3	3	SIAP
0981	4	1	3	3	BELUM SIAP
1041	3	4	4	3	SIAP
1009	4	3	3	4	SIAP
1042	4	4	4	2	BELUM SIAP
1010	3	3	4	3	SIAP

2. Menentukan entropy tahap 1

Setelah ditentukan data awal, langkah selanjutnya melakukan perhitungan dengan mencari nilai entropy dari masing-masing data. Berikut ini merupakan rumus entropy [7]:

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

Berdasarkan rumus (1) di atas, dapat dilakukan perhitungan sebagai berikut :

$$\text{Entropy(S)} = \left( - \left( \frac{5}{10} \right) * \log_2 \left( \frac{5}{10} \right) \right) + \left( - \left( \frac{5}{10} \right) * \log_2 \left( \frac{5}{10} \right) \right) = 0,97095059$$

- Entropy(b.ind-1) = 0
- Entropy(b.ind-2) = 0
- Entropy(b.ind-3) = 1
- Entropy(b.ind-4) = 1
- Entropy(mat-1) = 0
- Entropy(mat-2) = 0
- Entropy(mat-3) = 1
- Entropy(mat-4) = 1
- Entropy(b.ing-1) = 0
- Entropy(b.ing-2) = 0
- Entropy(b.ing-3) = 1
- Entropy(b.ing-4) = 1
- Entropy(ipa-1) = 0
- Entropy(ipa-2) = 0
- Entropy(ipa-3) = 0,97095059
- Entropy(ipa-4) = 0

3. Menentukan gain tahap 1

Setelah melakukan perhitungan entropy, langkah selanjutnya melakukan perhitungan gain. Berikut ini merupakan rumus gain [7]:

$$\text{Gain (A)} = \text{Entropi (S)} - \sum_{i=1}^k \frac{|S_i|}{|S|} * \text{Entropi}(S_i) \dots\dots\dots (2)$$

Berdasarkan rumus gain (2) di atas, dapat diperoleh perhitungan sebagai berikut :

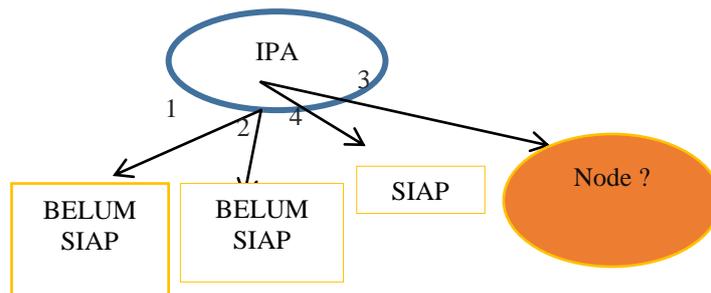
- Gain(b.ind) = 0,97095059-(0+0+((4/10)\*1)+((4/10)\*1)) = 0,17095059
- Gain(mat) = 0,97095059-(0+0+((4/10)\*1)+((4/10)\*1)) = 0,17095059
- Gain(b.ing)= 0,97095059-(0+0+((4/10)\*1)+((4/10)\*1)) = 0,17095059
- Gain(ipa) = 0,97095059-(0+0+((5/10)\* 0,97095059)+0) = 0,4854753

Setelah nilai entropy dan gain diperoleh hasilnya, kemudian dibuat sebuah tabel yang fungsinya untuk mempermudah perhitungan menjadi lebih rinci dan lebih jelas. Tabel 3 merupakan tabel yang dibuat berdasarkan hasil nilai-nilai entropy dan gain diatas.

Tabel 3. Node Tree Tahap1

Node 1	Jumlah	Siap	Belum	Entropy	Gain
Total	10	4	6	0,97095059	
B.IND					0,17095059
1	1	0	1	0	
2	1	0	1	0	
3	4	2	2	1	
4	4	2	2	1	
MAT					0,17095059
1	2	0	2	0	
2	0	0	0	0	
3	6	3	3	1	
4	2	1	1	1	
B.ING					0,17095059
1	1	0	1	0	
2	1	0	1	0	
3	4	2	2	1	
4	4	2	2	1	
IPA					0,4854753
1	1	0	1	0	
2	3	0	3	0	
3	5	3	2	0,97095059	
4	1	1	0	0	

Berdasarkan tabel 3, terlihat hasil nilai-nilai entropy dan gain yang telah disusun rapi dalam satu tabel untuk di eksekusi. Setelah dilakukan perhitungan, didapat hasil dalam tabel berwarna kuning yang berarti tabel tersebut digunakan sebagai acuan untuk diolah ke perhitungan selanjutnya. Hasil pohon keputusan :



Gambar 2. Pohon Keputusan Node 1

4. Mengambil data training tahap 2

Setelah diperoleh hasil pohon keputusan berdasarkan tabel diatas. Kemudian langkah selanjutnya dilakukan perhitungan ulang dengan menggunakan data baru dari hasil tabel 3. Tabel tersebut diberi nama tabel training tahap 2.

Tabel 4 Tabel Data Training Tahap 2

NO INDUK	IPA	B.IND	MAT	B.ING	KRITERIA
----------	-----	-------	-----	-------	----------

0945	3	3	3	1	BELUM SIAP
0918	3	4	3	3	SIAP
0981	3	4	1	3	BELUM SIAP
1041	3	3	4	4	SIAP
1010	3	3	3	4	SIAP

5. Menentukan entropy tahap 2

Dilakukan perhitungan entropy sesuai dengan data training diatas.

$$\text{Entropy}(\text{ipa-3}) = \left(-\left(\frac{3}{5}\right) * \log_2\left(\frac{3}{5}\right)\right) + \left(-\left(\frac{2}{5}\right) * \log_2\left(\frac{2}{5}\right)\right) = 0,97095059$$

$$\text{Entropy}(\text{b.ind-3}) = 0,91829583$$

$$\text{Entropy}(\text{b.ind-4}) = 1$$

$$\text{Entropy}(\text{mat-1}) = 0$$

$$\text{Entropy}(\text{mat-3}) = 0,91829583$$

$$\text{Entropy}(\text{mat-4}) = 0$$

$$\text{Entropy}(\text{b.ing-3}) = 0$$

$$\text{Entropy}(\text{b.ing-3}) = 1$$

$$\text{Entropy}(\text{b.ing-4}) = 0$$

6. Menentukan gain tahap 2

Setelah menghitung entropy, dilakukan perhitungan gain. Sebagai berikut :

$$\text{Gain}(\text{b.ind}) = 0,97095059 - (((3/5) * 0,91829583) + ((2/5) * 1)) = 0,01997309$$

$$\text{Gain}(\text{mat}) = 0,97095059 - (0 + ((3/5) * 0,91829583) + 0) = 0,41997309$$

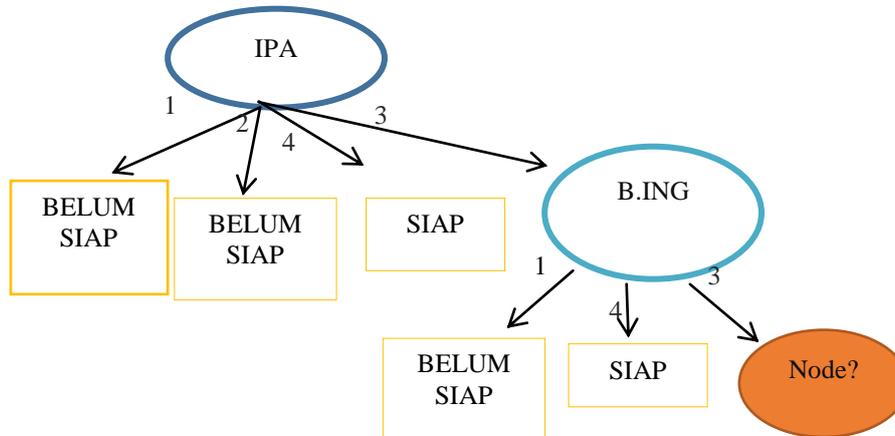
$$\text{Gain}(\text{b.ing}) = 0,97095059 - (0 + ((2/5) * 1) + 0) = 0,57095059$$

Setelah ditemukan nilai entropy dan gain pada tahap 2, kemudian didapatkan hasil seperti pada tabel 5 di bawah ini.

Tabel 5. Node Tree Tahap 2

Node	Jumlah	siap	blm	Entropy	Gain
1.1 IPA-3	5	3	2	0,97095059	
B.IND					0,01997309
	3	3	1	0,91829583	
	4	2	1	1	
MAT					0,41997309
	1	1	0	0	
	3	3	2	0,91829583	
	4	1	1	0	
B.ING					0,57095059
	1	1	0	0	
	3	2	1	1	
	4	2	2	0	

Berdasarkan tabel 5 di atas, terlihat bahwa tabel yang terdapat warna kuning terletak pada gain milik mata pelajaran Bahasa Inggris. Maka, yang digunakan sebagai acuan pada perhitungan selanjutnya adalah nilai skala 3 milik mata pelajaran Bahasa Inggris. Tabel berwarna kuning tersebut juga digunakan sebagai acuan untuk menentukan pohon keputusan pada node tree tahap 2. Dibawah ini adalah gambar pohon keputusannya :



Gambar 3. Pohon Keputusan Node Tahap 2

Pada gambar 3 di atas, menjelaskan bahwa mata pelajaran IPA dalam skala 3 memiliki nilai yang masuk ke dalam kelas Bahasa Inggris. Dimana skala 4 Bahasa Inggris berarti siap dan skala 1 berarti belum siap, sedangkan skala 3 Bahasa Inggris masih belum diketahui nilainya.

7. Mengambil data training tahap 3

Setelah diperoleh hasil pohon keputusan berdasarkan tabel training tahap 2, kemudian dilakukan perhitungan ulang dengan menentukan data baru berdasarkan hasil diatas. Berikut tabel data training tahap 3 :

Tabel 6. Data Training Tahap 3

NO INDUK	B.ING	B.IND	MAT	KRITERIA
0918	3	4	3	SIAP
0981	3	4	1	BELUM SIAP

8. Menentukan entropy tahap 3

Perhitungan entropy dilakukan berdasarkan data training diatas. Berikut perhitungannya :

$$\text{Entropy}(b.ing-3) = \left(-\left(\frac{1}{2}\right) * \log_2\left(\frac{1}{2}\right)\right) + \left(-\left(\frac{1}{2}\right) * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy}(b.ind-4) = 1$$

$$\text{Entropy}(mat-1) = 0$$

$$\text{Entropy}(mat-3) = 0$$

9. Menentukan gain tahap 3

Setelah nilai entropy didapat, langkah selanjutnya mencari nilai gain seperti dibawah ini :

$$\text{Gain}(b.ind) = 1 - \left(\left(\frac{2}{2}\right) * 1\right) = 0$$

$$\text{Gain}(mat) = 1 - (0+0) = 1$$

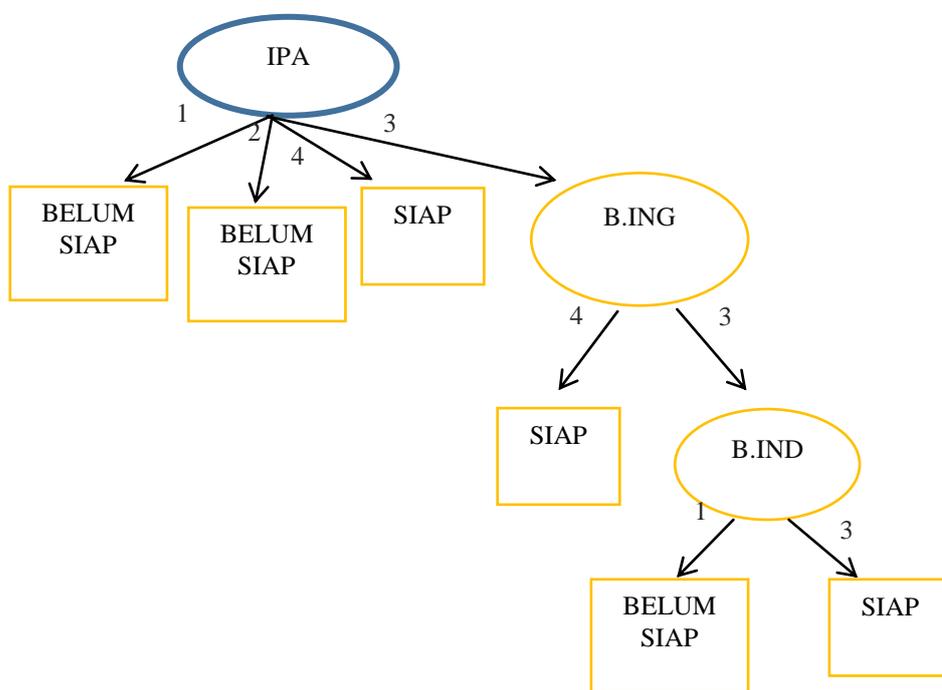
Setelah ditemukan nilai entropy dan gain pada tahap 2, kemudian didapatkan hasil seperti pada tabel dibawah.

Tabel 7. Node Tree Tahap 3

Node	Jumlah	siap	blm	Entropy	Gain
------	--------	------	-----	---------	------

1.2	Total	2	1	1	1
	B.ING-3				
	B.IND				0
		4	2	1	1
	MAT				1
		1	1	0	1
		3	1	1	0

Berdasarkan tabel 7 di atas, terlihat bahwa gain dengan nilai tertinggi dimiliki oleh mata pelajaran bahasa indonesia. Dengan begitu dapat disimpulkan pada pohon keputusan cabang terakhir adalah milik mata pelajaran bahasa indonesia.



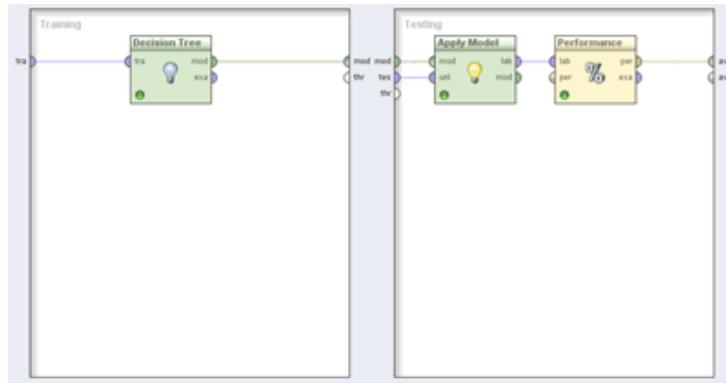
Gambar 4. Pohon Keputusan Node Tahap 3

Gambar 4 di atas merupakan hasil pohon keputusan terakhir setelah dilakukan perhitungan panjang algoritma C4.5. Dari gambar diatas dapat dilihat bahwa nilai ipa 1 yaitu belum siap, nilai ipa 2 yaitu belum siap, nilai ipa 4 yaitu siap, sedangkan nilai ipa 3 masuk kedalam nilai bahasa inggris. Nilai bahasa inggris 4 berarti siap, sedangkan nilai bahasa inggris 3 berarti masuk ke nilai bahasa indonesia. Nilai bahasa indonesia 1 berarti belum siap, sedangkan nilai bahasa indonesia 3 berarti siap.

### 3.2 Pengujian Algoritma C4.5

Percobaan yang dilakukan pada implementasi algoritma C4.5 dalam penelitian ini menggunakan aplikasi RapidMiner. Data yang diolah menggunakan data try out kelas IX yang berjumlah 383 data. Atribut pada data tersebut diantaranya nomor induk, nilai b.ind, nilai mat, nilai b.ing, nilai ipa, dan kriteria. Proses validation tersebut dihubungkan pada bagian model, training, dan average ke bagian result yang memiliki tujuan untuk menghasilkan nilai akurasi dan nilai eror (error rate). Dalam proses validation terdapat proses lain yang masih dilakukan

untuk menyempurnakan pengujian data tersebut. Berikut ini merupakan proses yang terjadi di dalam validation.



Gambar 5. Decision Tree Model

Pada gambar 5 di atas terlihat bagian training terdapat blok algoritma decision tree yang dihubungkan dengan garis dan terhubung ke blok apply model. Dalam blok tersebut digunakan pada data training yang kemudian diterapkan ke dalam data testing dengan menggunakan algoritma C4.5. Dari blok apply model dihubungkan kembali dengan blok performance yang bertujuan untuk menevaluasi nilai akurasi dari model.

Berkaitan dengan data training yang berjumlah 383 data yang berisi variabel antara lain nomor induk, nilai b.ind, nilai mat, nilai b.ing, dan nilai ipa, kemudian dilakukan pemodelan menggunakan algoritma decision tree C4.5. Berdasarkan pemodelan tersebut diperoleh hasil dalam bentuk Confusion Matrix sebagai berikut :

	true BELUM SIAP	true SIAP	class precision
pred. BELUM SIAP	190	2	98.96%
pred. SIAP	0	191	100.00%
class recall	100.00%	98.96%	

Gambar 6. Nilai Akurasi Model C4.5

Berdasarkan gambar 4.3 di atas, diketahui jumlah true positif (tp) sebanyak 190 record, sedangkan jumlah false positif (fp) sebanyak 0 record. Lalu pada jumlah true negatif (tn) sebanyak 191 record, sedangkan false negatif (fn) jumlahnya sebanyak 2 record. Terlihat pada gambar hasil nilai akurasi prediksi kesiapan siswa dalam menghadapi ujian nasional menggunakan perantara aplikasi RapidMiner sebesar 99,48%. Adapun perhitungan manual menentukan akurasinya adalah sebagai berikut :

$$\begin{aligned}
 \text{Akurasi} &= ((tp + tn) / (tp + tn + fp + fn)) * 100\% \\
 &= ((190 + 191) / (190 + 191 + 0 + 2)) * 100\% \\
 &= (381 / 383) * 100\% \\
 &= 0,99477 * 100\% \\
 &= 99,48\%
 \end{aligned}$$

### 3.3 Implementasi Program

Aplikasi yang akan dibuat penulis untuk membantu sekolah diaplikasikan menggunakan java netbeans. Berikut ini desain interface aplikasi yang telah penulis buat.

**PREDIKSI KESIAPAN SISWA MENGHADAPI UN**

NO. INDUK: 1005  
 NAMA SISWA: mei  
 Try Out ke: Try Out 1

MATEMATIKA: 8.3  
 B. INDONESIA: 9.0  
 B.INGGRIS: 7.9  
 IPA: 7.6

DONE      RESULT: SIAP  
 Pertahankan nilai di atas KKM

Data Kesiapan UN Siswa

No Induk	Nama	Try Out 1	Try Out 2	Try Out 3	Try Out 4	Try Out 5	Try Out 6	Try Out 7	Try Out 8
1001	nadia	BELUM ...	BELUM ...	BELUM ...	SIAP	SIAP	SIAP	SIAP	SIAP
1002	deby	BELUM ...	SIAP	BELUM ...	SIAP	BELUM ...			
1003	linda	BELUM ...							
1004	fika	BELUM ...							

EDIT      HAPUS      OUT

Gambar 7. Desain Interface Aplikasi

Terlihat pada gambar 7, pada form inputan pertama pengguna harus memasukkan nomor induk, nama siswa, dan data try out. Hal tersebut bertujuan agar data yang dihasilkan dapat diketahui melalui inputan yang digunakan sebagai identitas data. Kemudian pada form inputan kedua pengguna harus memasukkan nilai-nilai try out yang telah diperoleh siswa berdasarkan try out yang di ikuti siswa. Nilai-nilai tersebut meliputi nilai matematika, nilai Bahasa Indonesia, nilai Bahasa Inggris, dan nilai IPA.

### 3.4 Evaluasi Kinerja

Tahapan terakhir setelah memaparkan implementasi program, selanjutnya perlu dituangkan dalam bentuk evaluasi kinerja sistem. Hal tersebut dikarenakan, salah satu tujuan dalam pembuatan tugas akhir ini adalah untuk memudahkan pihak sekolah dalam menentukan prediksi kesiapan siswa SMP Islam Terpadu PAPB Semarang menghadapi ujian nasional. Maka dari itu, penulis ingin mengetahui seberapa jauh peran aplikasi tersebut dalam memudahkan pihak sekolah. Evaluasi yang dilakukan penulis bertujuan untuk mengetahui respon pihak sekolah terhadap aplikasi yang telah dibuat. Dengan adanya evaluasi tersebut dapat membantu sebagai koreksi dan perbaikan terhadap aplikasi penulis apabila terdapat kesalahan yang tidak diinginkan oleh pihak sekolah.

Dari hasil evaluasi kinerja, dapat dilakukan perhitungan menggunakan skala likert dengan hasil sebagai berikut:

1. Responden yang menjawab sangat setuju (5) =  $11 \times 5 = 55$
2. Responden yang menjawab setuju (4) =  $8 \times 4 = 32$
3. Responden yang menjawab netral (3) =  $1 \times 3 = 3$
4. Responden yang menjawab tidak setuju (2) =  $0 \times 2 = 0$
5. Responden yang menjawab sangat tidak setuju (1) =  $0 \times 1 = 0$

$$\text{Total skor} = 55 + 32 + 3 + 0 + 0 = 90$$

Setelah mendapatkan total skor untuk tiap-tiap soalnya, kemudian langkah selanjutnya menentukan prosentase yang didapat dari keseluruhan kuesioner tersebut. Untuk mendapatkan hasil interpretasi, perlu diketahui dahulu skor tertinggi (X) dan skor terendah (Y) untuk tiap item penilaian.

$$Y = 5 \times 20 = 100$$

$$X = 1 \times 20 = 20$$

$$\text{Rumus index} = \frac{\text{total skor}}{Y} \times 100 \dots \dots \dots (9)$$
$$\text{Rumus index} = \frac{(90+81+82+76+84+78+77+77+100+88)}{100} \times 10 = 83,3\%$$

Berdasarkan penyebaran kuesioner yang dilakukan oleh penulis diperoleh hasil indexnya sebesar 83,3%. Dari hasil tersebut, terbukti bahwa aplikasi yang diberikan kepada sekolah dapat memudahkan pihak sekolah dalam mempersiapkan siswanya menghadapi ujian nasional dengan baik melalui hasil nilai try out siswa. Dengan begitu, tujuan dari tugas akhir dalam memudahkan pihak sekolah telah tercapai.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa :

1. Metode klasifikasi dapat diterapkan untuk memprediksi kesiapan siswa SMP Islam Terpadu PAPB Semarang menghadapi ujian nasional. Hal tersebut berdasarkan pengujian yang telah dijelaskan pada bab sebelumnya, pemodelan algoritma C4.5 menghasilkan akurasi sebesar 99,48%.
2. Penelitian ini dapat memudahkan pihak sekolah dalam memprediksi kesiapan siswa menghadapi ujian nasional dengan cara menyebar kuesioner kepada pihak sekolah mengenai aplikasi yang telah dibuat penulis menggunakan Java. Setelah melakukan penyebaran kuesioner tersebut memperoleh hasil presentase sebesar 83,3% yang berarti pihak sekolah puas terhadap aplikasi yang telah dibuat.

#### DAFTAR PUSTAKA

- [1] M. Ainurrofiq, "Implementasi Algoritma Naive Bayes Untuk Menentukan Kesiapan Siswa Dalam Menghadapi Ujian Nasional," Skripsi Teknik Informatika Universitas Dian Nuswantoro, Semarang, 2015.
- [2] D. H. Kamagi dan S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Ultimatics*, vol. VI, no. 6, pp. 15-20, 2014.
- [3] T. Novianti dan A. Azis, "Aplikasi Data Mining Menggunakan Metode Decision Tree untuk Menampilkan Laporan Hasil Nilai Akhir Mahasiswa (Studi Kasus di Fakultas Teknik UM Surabaya)," *Jurnal Ilmiah NERO*, vol. 1, no. 6, pp. 198-203, 2015.
- [4] Wandasari, N. D. (2013). Perlakuan Akutansi Atas PPH Pasal 21 Pada PT. ARTHA PRIMA FINANCE KOTAMOBAGU. *Jurnal EMBA*, 558-566.
- [5] I. Budiman, "Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," Program Pasca Sarjana Universitas Diponegoro, Semarang, 2012.
- [6] Juju, F. S. (2010). *Data Mining : Meramalkan Bisnis Perusahaan*. Jakarta: PT Elex Media Komputindo.
- [7] Erricha Paramitha Dewi, A. E. (2015). Klasifikasi Status Gizi Balita Menggunakan Metode Algoritma C4.5 Berbasis Web. *Universitas Muhammadiyah Sidoarjo*, 1-14.
- [8] N. S. Budanis Dwi Meilani, "Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes," *Jurnal Ilmiah NERO*, vol. 1, no. 8, pp. 182-189, 2015.
- [9] C. R. Sari, "Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA Negeri 1 Polewali," *Indonesian Journal on Networking and Security*, vol. 5, no. 7, pp. 48-54, 2016.
- [10] E. D. Diana Laily Fithri, "Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes," *Prosiding SNATIF*, vol. 1, no. 6, pp. 319-324, 2014.