

# PENENTUAN PENERIMAAN SISWA BARU MENGUNAKAN DECISION TREE

<sup>1</sup>Fitroh Rizky Muwardah, <sup>2</sup>Ricardus Anggi Pramunendar, M.Cs

Program Studi Teknik Informatika – S1

Fakultas Ilmu Komputer

Universitas Dian Nuswantoro, Jl. Nakula 1 No. 5-11. Semarang

<sup>1</sup>[111201106141@mhs.dinus.ac.id](mailto:111201106141@mhs.dinus.ac.id), <sup>2</sup>[ricardus.anggi@dsn.dinus.ac.id](mailto:ricardus.anggi@dsn.dinus.ac.id)

## ABSTRAK

Penerimaan siswa baru pada suatu institusi pendidikan merupakan sebuah kegiatan yang selalu dilaksanakan setiap tahun ajaran baru, dimana data calon siswa baru tersebut selalu meningkat dari tahun ke tahun dan hasil penyeleksian masih dilakukan secara manual, hal tersebut menimbulkan banyak kekeliruan, akurasi pemilihan yang tidak konsisten, kecurangan dan penilaian yang bersifat subjektif. Berdasarkan paper-paper sebelumnya yang membahas hal tersebut diperlukan penambahan parameter penduga selain parameter yang ada seperti parameter Asal Sekolah. Oleh karena itu diterapkan *data mining* dengan metode C.4.5 untuk menentukan siswa yang diterima di sekolah tersebut dengan menambahkan parameter Asal Sekolah. Penelitian menggunakan data calon siswa yang mendaftar melalui Jalur Reguler di Madrasah Aliyah Negeri Purbalingga tahun ajaran 2014/ 2015. Proses evaluasi dan validasi dilakukan dengan menghitung akurasi, *precision*, *recall* dan *f-measure*. Hasil menunjukkan bahwa *dataset* yang menggunakan Variabel Asal Sekolah nilai akurasinya meningkat dari 91,9% menjadi 94,6% dengan *ratio data training* 85%.

Kata Kunci : siswa baru, *decision tree* C4.5, *data mining*, *data mining* klasifikasi

## I. PENDAHULUAN

Proses penerimaan siswa baru yang ada di Madrasah Aliyah Negeri Purbalingga disediakan dalam beberapa jalur seperti Jalur Beasiswa, Mandiri dan Reguler. Pada penelitian ini akan membahas khususnya Jalur Reguler.

Dalam kegiatannya proses seleksi penerimaan siswa baru Jalur Reguler masih dilakukan secara manual

sehingga memungkinkan adanya hasil seleksi yang subjektif dan tentunya akan memakan waktu pengerjaan yang lama. Sehingga diperlukan sebuah metode yang dapat mengatasi masalah tersebut agar hasil yang diperoleh menjadi lebih objektif, akurat dan dapat mempersingkat waktu.

Berdasarkan paper-paper tersebut penambahan parameter penduga selain parameter yang ada seperti Asal Sekolah perlu dilakukan agar data yang diolah dapat memperoleh pemodelan data yang lebih baik.

Metode yang akan digunakan pada penelitian ini dengan penerapan data mining. Adapun teknik data mining yang dapat dimanfaatkan dalam proses tersebut adalah klasifikasi C4.5.

## II. DASAR TEORI

### A. Data Mining

Data mining merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer (machine learning)* untuk menganalisis dan mengekstraksi pengetahuan secara otomatis [3].

#### Decision Tree

Algoritma pohon keputusan yang terkenal adalah C4.5. Pada akhir tahun 1970 sampai awal tahun 1980 J. Ross Quinlan, membuat sebuah algoritma *decision tree* yang dikenal dengan ID3 (*Iterative Dichotomiser*). Quinlan kemudian membuat algoritma C4.5 yang merupakan pengembangan dari

algoritma ID3 [11]. Algoritma ini memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan) [7].

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan Algoritma C4.5 [9], yaitu :

1. Menyiapkan *data training*.
2. Menentukan akar dari pohon dengan *Entropy* dan *Gain*.

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

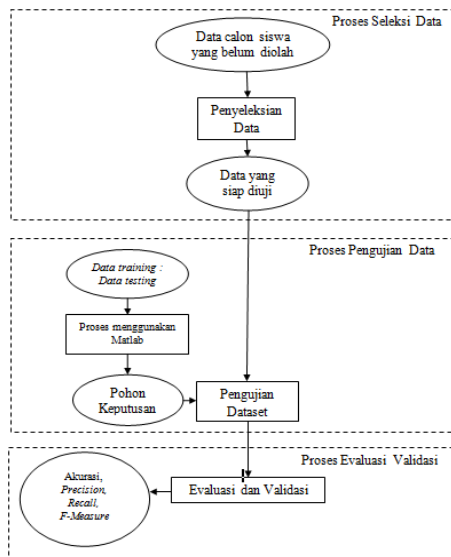
$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i).$$

3. Ulangi langkah ke-2 hingga semua *record* terpartisi.

## III. METODE YANG DIUSULKAN

Berikut blok diagram untuk menggambarkan metode yang diusulkan oleh peneliti :

Gambar 1. Blok Diagram Usulan



### 1. Proses Penyeleksian Data

Data calon siswa yang telah didapat diseleksi untuk diambil mana yang akan digunakan atau tidak digunakan. Ada dua data yang dibuang karena data tersebut tidak lengkap. Sehingga data sekarang berjumlah 248. Dan Atribut yang dipakai sekarang hanya enam, yaitu Sekolah Asal, B Indo, B Ingg, Mat, Ipa.

### 2. Proses Pengujian

Pengujian akan dilakukan menggunakan Tool Matlab. Dataset dibagi menjadi dua bagian, yaitu *data testing* dan *data training*. Pengujian dilakukan sesuai skenario berikut :

- a. Pengujian tanpa Variabel Asal Sekolah , *Data training* 85% : *data testing* 15%
- b. Pengujian dengan Variabel Asal Sekolah
  - a. *Data training* 85% : *data testing* 15%
  - b. *Data training* 75% : *data testing* 25%
  - c. *Data training* 65% : *data testing* 35%

### 3. Proses Evaluasi dan Validasi

Setelah data diuji, dilakukan evaluasi dan validasi melalui penghitungan akurasi , *precision*, *recall* dan *f-measure*.

- a. Akurasi

Akurasi =

$$\frac{\text{Jumlah yang diklasifikasi benar}}{\text{Total sample testing yang diuji}}$$

- b. *precision* dan *recall*.

$$p = \frac{TP}{TP+FP} \quad r = \frac{TP}{TP + FN}$$

- c. *F-Measure*.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

## IV. HASIL PENGUJIAN

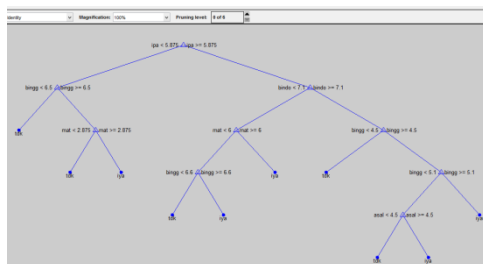
- A. Pengujian Rules Terhadap Data Calon Siswa dengan Variabel Asal Sekolah

Pengujian terhadap validitas sistem bertujuan untuk mengetahui apakah solusi yang dihasilkan oleh pohon keputusan tersebut *valid* atau tidak.

### 1. Pengujian I

*Data set* dibagi menjadi dua 85% *data training* dan 15% *data testing*.

Gambar 2. Pohon Keputusan 1



Kemudian dilakukan evaluasi dan validasi dengan cara menghitung nilai akurasi, *precision*, *recall*, dan *f-measure*. Hasilnya tertera pada Tabel 2.

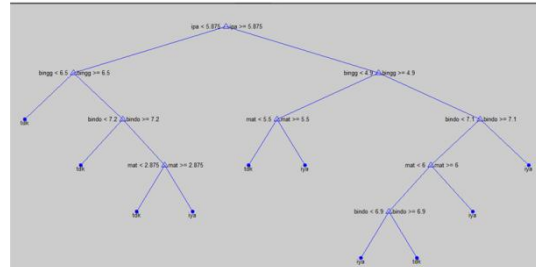
Tabel 1. Evaluasi Validasi 1

Persentase Data	Data training	Data testing	Akurasi	Precision	Recall	F-Measure
85% : 15%	211	37	0.9459459459	1	0.875	0.9333333333

### 2. Pengujian II

*Data set* dibagi menjadi dua 75% *data training* dan 25% *data testing*. Berikut pohon keputusan yang dihasilkan :

Gambar 3. Pohon Keputusan 2



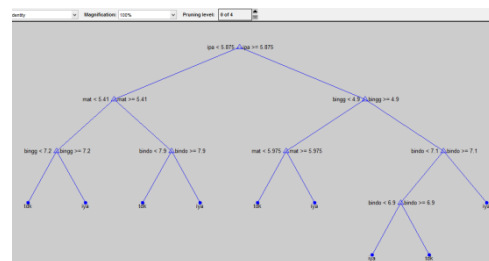
Tabel 2. Evaluasi Validasi 2

Persentase Data	Data training	Data testing	Akurasi	Precision	Recall	F-Measure
75% : 25%	186	62	0.935483871	0.973684211	0.925	0.948717949

### 3. Pengujian III

*Data set* dibagi menjadi dua 65% *data training* dan 35% *data testing*. Berikut pohon keputusan yang dihasilkan :

Gambar 4. Pohon Keputusan 3



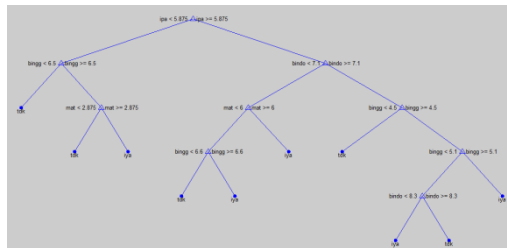
Tabel 3. Evaluasi Validasi 3

Persentase Data	Data training	Data testing	Akurasi	Precision	Recall	F-Measure
65% : 35%	161	87	0.908045977	0.981818181	0.885245901	0.931034482

### B. Pengujian Rules Terhadap Data Calon Siswa Tanpa Variabel Asal Sekolah

Data set dibagi menjadi dua 85% data training dan 15% data testing. Keseluruhan data set berjumlah 248, data training sebanyak 211 data dan data testing sebanyak 37 data. Variabel yang digunakan Bindo, Bingg, Mat, Ipa, Hasil. Berikut pohon keputusan yang dihasilkan :

Gambar 5. Pohon Keputusan 4



Tabel 4. Evaluasi Validasi 4

Persentase Data	Data training	Data testing	Akurasi	Precision	Recall	F-Measure
85% : 15%	211	37	0.91891899	1	0.875	0.9333333

### C. Analisa Hasil Pengujian

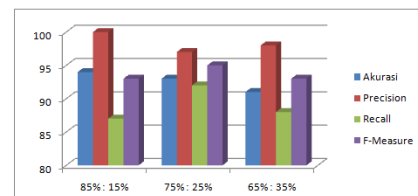
Pengujian I, II, dan III menghasilkan akurasi, precision, recall dan f-measure sebagai berikut:

Tabel 5. Analisa Pengujian

	Persentase Data	Data training	Data testing	Akurasi	Precision	Recall	F-Measure
De ng an Va r	85% : 15%	211	37	0.945945 9459	1	0.875	0.9333 33333
	75% : 25%	186	62	0.935483 871	0.97368421 1	0.925	0.9487 17949
r As al	65% : 35%	161	87	0.908045 977	0.98181818 1	0.885 24590 1	0.9310 34482

Berikut grafik yang dihasilkan dari pengujian 1, 2 dan 3 :

Gambar 6. Grafik Hasil Pengujian



Grafik tersebut menunjukkan bahwa semakin banyak data training yang digunakan maka semakin akurat hasil yang didapat.

Hasil pengujian menggunakan variabel Asal tersebut mengalami kenaikan akurasi dari 0.91891899 menjadi 0.9459459459. Berikut perbandingan hasilnya :

Tabel 6. Perbandingan Akurasi

No	Persentase Data	Data training	Data testing	Akurasi	Precision	Recall	F-Measure
Ta np a As al	85% : 15%	211	37	0.918918 99	1	0.875	0.9333 33333
De ng an As al	85% : 15%	211	37	0.945945 9459	1	0.875	0.9333 33333

## V. KESIMPULAN

### A. Kesimpulan

Berdasarkan hasil pengujian metode *decision tree* terhadap data calon siswa dapat disimpulkan :

Pengimplementasian metode *decision tree* menggunakan variabel asal sekolah terhadap data calon siswa dapat digunakan untuk menentukan penerimaan siswa baru dengan baik, akurasi yang dihasilkan meningkat dari 91,9% menjadi 94,6%.

### B. Saran

Penelitian dapat dilanjutkan dengan menambahkan parameter tambahan lainnya , misalnya prestasi siswa pada lomba akademik atau non akademik, sehingga hasilnya akan lebih selektif.

Pengujian metode ini belum sampai pada implementasi, maka perlu dibuat sistem aplikasi yang diperuntukkan untuk pihak sekolah.

## VI. REFERENSI

[1] Pudjo Prabowo,  
“PENERAPAN DATA

MINING DENGAN  
MATLAB,” Rekayasa Sains ,  
Bandung, 2013.

- [2] Astuti Fajar, “Data Mining,” Andi, Surabaya, 2009.
- [3] Santosa Budi, “Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis”, Yogyakarta : Graha Ilmu, 2007.
- [4] Anwar Arifin, Shaufiah, Dwi Dawam, “Analisis dan Implementasi Metode Single Layer Perceptron Pada Data Mining Penerimaan Siswa Baru Jalur JPPAN,” pp. 1–7, 2012.
- [5] Ardhi Fauzan, “Analisis dan Implementasi Algoritma Cruise dengan Metode ID pada Penentuan Seleksi Penerimaan Mahasiswa Baru di Perguruan Tinggi melalui JPPAN,” pp. 1–7, 2009.
- [6] Dwi Aditya and Shaufiah, “Pemanfaatan Data Mining pada Penerimaan Mahasiswa Baru Menggunakan Metode AHP dan Algoritma C4.5 Decision Tree,” pp. 1–12, 2013.
- [7] Hamid Abdul, Al-Ghamdi Abdullah, Naoman Amin, Madbouly Ayman, “A Comparative Analysis of Classification Algorithm for student college enrollment approval using data mining” pp. 1–8, 2012.

