

# Evaluasi klasifikasi penerima bidikmisi menggunakan algoritma *iterative dichotomiser 3 (ID3)*

Jainuddin<sup>1</sup>, Islamiyah<sup>2</sup>, Gubtha Mahendra Putra<sup>3</sup>, Havaluddin<sup>4</sup>, Vina Zahrotun Kamila<sup>5</sup>

*Jurusan Teknologi Informasi dan Komunikasi, Universitas Mulawarman, Samarinda, Indonesia*

<sup>1</sup> jainuddin911@gmail.com; <sup>2</sup> .islamiyah@fkti.unmul.ac.id; <sup>3</sup> gubthamp@fkti.unmul.ac.id; <sup>4</sup> havaluddin@unmul.ac.id; <sup>5</sup> vinakamila@fkti.unmul.ac.id

---

## INFORMASI ARTIKEL

### Histori Artikel

Diterima : 15 Januari 2020

Direvisi : 2 Februari 2020

Diterbitkan : 4 April 2020

---

### Kata Kunci:

Kinerja Algoritma

Data Mining Klasifikasi

*Iterative Dichotomiser 3 (ID3)*

## ABSTRAK

Teknik klasifikasi data mining yang cukup populer adalah *Decision Tree* diantaranya menggunakan algoritma *Iterative Dichotomiser 3 (ID3)*. Klasifikasi didapatkan dari pohon keputusan yang terbentuk melalui algoritma *Iterative Dichotomiser 3 (ID3)* yang akan diukur tingkat akurasi dan error rate algoritma dalam menentukan klasifikasi. Hal ini dapat dilakukan dengan cara membentuk model pohon keputusan pada mesin learning *RapidMiner* menggunakan data training dan evaluasi membandingkan data nyata dengan data testing klasifikasi untuk mengukur akurasi algoritma. Tujuan dalam penelitian ini adalah untuk menghasilkan informasi klasifikasi kelayakan penerima bidikmisi menggunakan algoritma *Iterative Dichotomiser 3 (ID3)* di Institut Agama Islam Negeri (IAIN) Samarinda dan untuk mengetahui akurasi algoritma yang digunakan. Variabel penelitian terdiri pekerjaan orang tua, jumlah penghasilan orang tua, jumlah anggota keluarga, status kepemilikan rumah, jumlah pengeluaran keluarga, dan status kepemilikan SKTM/KIP. Berdasarkan hasil analisis dengan mengukur kinerja algoritma menggunakan metode *confusion matrix*, dengan menghasilkan akurasi 98.3% dan error rate 1.7% dalam menentukan klasifikasi kelayakan penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda.

2019 SAKTI – Sains, Aplikasi, Komputasi dan Teknologi Informasi.

Hak Cipta.

---

## I. Pendahuluan

Bidikmisi merupakan bantuan dana biaya pendidikan yang berfokus kepada calon-calon mahasiswa yang terkendala ekonomi untuk melanjutkan pendidikan ke perguruan tinggi. Proses seleksi penerima Bidikmisi diserahkan kepada masing-masing perguruan tinggi berdasarkan persyaratan-persyaratan yang telah ditentukan oleh Jenderal Pendidikan Tinggi Kementerian Pendidikan Nasional Republik Indonesia. Institut Agama Islam Negeri (IAIN) Samarinda merupakan salah satu perguruan tinggi keagamaan negeri yang mendapatkan program penerimaan mahasiswa Bidikmisi.

Proses seleksi penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda terdiri dari tiga tahapan yaitu seleksi *online* berbasis sistem, survei lapangan dan wawancara. Berdasarkan data penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda yang telah dinyatakan lolos seleksi, penulis berkeinginan melakukan penelitian dengan tujuan untuk mendapatkan informasi berupa pohon keputusan klasifikasi kelayakan penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda. Sehingga, informasi tersebut dapat digunakan sebagai evaluasi klasifikasi kelayakan penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda.

Penelitian yang akan dilakukan penulis yaitu merancang *data mining* klasifikasi menggunakan algoritma *Iterative Dichotomiser 3 (ID3)* dengan melakukan analisis perhitungan data, sehingga dapat menghasilkan informasi kelayakan penerima Bidikmisi yang telah dinyatakan lolos seleksi berdasarkan variabel data yang dimiliki. Selain itu, penelitian ini juga diharapkan untuk mengetahui akurasi algoritma yang digunakan.

## II. Dasar Teori

*Data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1]–[3].

Klasifikasi merupakan proses penemuan model yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari obyek yang label kelasnya tidak diketahui [4], [5].

*Decision Tree* merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari *node* keputusan yang dihubungkan dengan cabang-cabang dari simpul akar sampai ke *node* daun (akhir) [6], [7]. Pada *node* keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke *node* lain atau ke *node* akhir untuk menghasilkan suatu keputusan [3].

Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Pohon keputusan dimulai dengan sebuah *root node* (titik awal) yang dipakai oleh *user* untuk mengambil tindakan. Dari *node root* ini, *user* memecahkan sesuai dengan algoritma *Decision Tree*. Hasil akhirnya adalah sebuah pohon keputusan dengan setiap cabangnya menunjukkan kemungkinan skenario dari keputusan yang diambil hasilnya. Sementara itu, algoritma ID3 merupakan algoritma *decision tree* yang paling dasar [8], [9]. Algoritma ini melakukan pencarian secara rakus/menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan [10].

### III. Metode Penelitian

#### A. Data dan Metode Pengumpulan Data

Data yang digunakan dan metode pengumpulan data dalam penelitian ini akan dijabarkan pada bagian ini. Penjelasan adalah sebagai berikut:

##### 1) Data

Data dalam penelitian ini didapatkan dari data mahasiswa penerima Bidikmisi tahun angkatan 2015-2018 Institut Agama Islam Negeri Samarinda. Data tersebut berjumlah 120 dengan atribut yang dimiliki berupa nama, nim, pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan keluarga, kepemilikan rumah, pengeluaran bulanan, dan keterangan keluarga tidak mampu.

##### 2) Metode Pengumpulan Data

###### a) Studi Literatur

Tahapan ini dilakukan dengan membaca studi pustaka berupa buku pembelajaran, jurnal dan karya ilmiah lainnya yang serupa dan berkaitan dengan *data mining*. Sebagai referensi dalam menunjang penelitian ini agar tidak menyimpang.

###### b) Observasi

Tahapan ini ialah survei langsung ke lapangan. Penulis mengunjungi Institut Agama Islam Negeri Samarinda, untuk mengambil data-data yang diperlukan dalam penelitian.

###### c) Studi Literatur

Tahapan ini yaitu dengan melakukan wawancara secara langsung dengan satu staf pegawai bidang akademik dan kemahasiswaan Institut Agama Islam Negeri (IAIN) Samarinda untuk memperoleh sumber variabel sesuai dengan ketepatan dalam klasifikasi kelayakan penerima Bidikmisi.

#### B. Variabel

Terdapat dua jenis variabel pada penelitian ini, yaitu variabel bebas dan variabel terikat. Variabel bebas dalam penelitian ini sebanyak enam variabel terdiri dari penghasilan orang tua, pekerjaan orang tua, jumlah anggota keluarga, pengeluaran keluarga, tempat tinggal, dan keadaan tidak mampu. Variabel terikat dalam penelitian ini adalah label kelayakan calon penerima bidikmisi yaitu layak dan tidak layak.

#### C. Analisis Data

Pada tahap analisis data penelitian ini menggunakan model standarisasi *data mining* yaitu *Cross Industry Standart Process for Data Mining* (CRISP-DM). Langkah-langkahnya adalah sebagai berikut:

##### 1) *Bussiness Understanding*

Pada tahap ini berfokus pada tujuan penelitian yaitu untuk mengetahui akurasi dalam penentuan klasifikasi kelayakan penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda, dengan menerjemahkan data penerima Bidikmisi, sehingga didapatkan model yang terbaik untuk memenuhi dari tujuan penelitian.

##### 2) *Bussiness Understanding*

Data diperoleh dari data penerima bidikmisi di Institut Agama Islama Negeri (IAIN) Samarinda di bagian akademik, atribut yang disajikan pada Tabel 1.

Tabel 1 Keterangan Kriteria Data

No.	Atribut	Keterangan
1	Pekerjaan Orang Tua	Semua pekerjaan orang tua
2	Penghasilan Orang Tua	Nominal penghasilan orang tua
3	Jumlah Anggota Keluarga	Semua jumlah anggota keluarga yang tercatat
4	Rumah	Status kepemilikan rumah
5	Pengeluaran	Nominal pengeluaran keluarga
6	SKTM	Status keadaan keluarga tidak mampu
7	Asal Daerah	Semua daerah kabupaten/kota
8	Jurusan	Semua jenis jurusan mahasiswa

### 3) Data Preparation

Dari data yang didapatkan maka tahap selanjutnya yaitu pemilihan atribut yang akan digunakan untuk menentukan kemiskinan kemudian nantinya akan menggunakan aplikasi *data mining* yaitu menggunakan *RapidMiner*. Untuk memudahkan dalam mengolah data dan mendapatkan data yang bagus maka ada beberapa tahap yang dilakukan, diantaranya yaitu: (1) *data cleaning*, dan (2) *data integration* dan data transformasi. Pertama *data cleaning*, dari 120 data tidak ada data yang dihapuskan. Karena data yang didapatkan tidak ada yang bernilai kosong maka tahapan *data cleaning* tidak dilakukan karena data sudah siap untuk dipakai. Kedua *data integration* dan data transformasi, Untuk meningkatkan dan memudahkan dalam proses analisis maka dari 8 Kriteria yang diperoleh akan dipilih beberapa kriteria inti yang digunakan untuk penentuan tingkat kemiskinan yaitu seperti pada Tabel 2.

Tabel 2 Tahap *Integration* dan Transformasi Data

No.	Atribut	Proses	Penjelasan
1	Pekerjaan	Digunakan Sebagai Atribut	Status pekerjaan orang tua
2	Penghasilan	Digunakan Sebagai Atribut	Nominal penghasilan orang tua
3	Jumlah Anggota Keluarga	Digunakan Sebagai Atribut	Jumlah anggota di dalam rumah tangga
4	Rumah	Digunakan Sebagai Atribut	Status kepemilikan rumah
5	Pengeluaran	Digunakan Sebagai Atribut	Nominal pengeluaran keluarga
6	SKTM	Digunakan Sebagai Atribut	Status keadaan keluarga tidak mampu
7	Asal Daerah	<i>Data Cleaning</i>	Asal daerah mahasiswa
8	Jurusan	<i>Data Cleaning</i>	Jenis jurusan mahasiswa

Adapun tahapan selanjutnya yang harus dilakukan yaitu transformasi data ke dalam atribut. Transformasi tersebut dibagi menjadi enam tahapan sesuai dengan proses yang digunakan sebagai atribut. Data atribut tersebut disajikan pada Tabel 3 sampai dengan Tabel 8. Tabel 3 menunjukkan tentang kriteria pekerjaan orang tua yang dibagi menjadi tiga interval. Tabel 4 menunjukkan tentang kriteria penghasilan orang tua yang dibagi menjadi tiga interval. Tabel 5 menunjukkan tentang kriteria kepemilikan rumah yang dibagi menjadi tiga interval. Tabel 6 menunjukkan tentang status kepemilikan rumah yang dibagi menjadi tiga interval. Tabel 7 menunjukkan tentang jumlah pengeluaran keluarga yang dibagi menjadi tiga interval. Dan Tabel 8 menunjukkan tentang surat keterangan tidak mampu yang dibagi menjadi tiga interval. Setelah tahap *integration* dan transformation data maka menghasilkan data transformasi atribut.

Tabel 3 Pekerjaan Orang Tua

No.	Kriteria Pekerjaan	Interval
1.	Petani, Nelayan, Kuli dan pekerjaan serabutan lainnya	3
2.	Pedagang, Pegawai Swasta, Honorer, Pebengkel, Satpam, dan pekerjaan swasta lainnya.	2
3.	PNS, Kontraktor, Pengusaha, Manager, dan Direktur.	1

Tabel 4 Penghasilan Orang Tua

No.	Kriteria Penghasilan	Interval
1.	< atau = Rp. 2.000.000,00	3
2.	> Rp. 2.000.000,00 – Rp. 4.000.000,00	2
3.	> Rp. 4.000.000,00	1

Tabel 5 Jumlah Anggota Keluarga

No.	Kriteria Tanggungan	Interval
1.	Total anggota keluarga > atau = 8	3
2.	Total anggota keluarga 4 sampai 7	2
3.	Total anggota keluarga < 4	1

Tabel 6 Status Kepemilikan Rumah

No.	Kriteria Rumah	Interval
1.	Sewaan / belum memiliki rumah sendiri	3
2.	Rumah milik sendiri	2
3.	Numpang di rumah keluarga atau mertua dan sebagainya	1

Tabel 7 Pengeluaran Keluarga

No.	Kriteria Pengeluaran	Interval
1.	Jika besar pengeluaran > Rp. 2.000.000,00	3
2.	Jika besar pengeluaran Rp. 1.000.000,00 – Rp. 2.000.000,00	2
3.	Jika besar Pengeluaran < Rp. 1.000.000,00	1

Tabel 8 Tidak Mampu

No.	Kriteria Keterangan Tidak Mampu	Interval
1.	Memiliki surat keterangan tidak mampu atau surat jenis lainnya yang menunjukkan kondisi ekonomi keluarga lemah	3
2.	Memiliki Kartu Indonesia Pintar (KIP)	2
3.	Memiliki surat keterangan tidak mampu atau surat jenis lainnya yang menunjukkan kondisi ekonomi keluarga lemah	1

#### D. Implementasi Algoritma Iterative Dichotomiser 3 (ID3)

*Iterative Dichotomiser 3 (ID3)* adalah algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang paling dasar. Algoritma ini melakukan pencarian secara rakus/menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan [10], dengan menghitung nilai *entropy* dan *gain*. *Gain* tertinggi dinyatakan sebagai *node* awal (akar).

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

Keterangan:

S = himpunan kasus.

n = jumlah partisi S.

pi = proporsi dari Si terhadap S.

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

Keterangan:

S = himpunan kasus.

A = atribut

|Si| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

Entropy (Si) = *entropy* untuk sampel-sampel yang memiliki nilai i.

#### E. Pembentukan Model

Pembentukan model menggunakan *data training* dan algoritma *Iterative Dichotomiser 3 (ID3)*. Sedangkan untuk teknik evaluasi algoritma menggunakan *K-fold cross-validation* dilakukan pada mesin *learning RapidMiner* [11].

#### F. Pengujian

Pengujian dilakukan dengan membandingkan data nyata dengan data testing klasifikasi. Langkah selanjutnya adalah menghitung nilai akurasi menggunakan *confusion matrix* dalam memprediksi klasifikasi.

#### G. Hasil Decision Tree

Pohon keputusan yang terbentuk dari proses *data mining* klasifikasi dan telah dilakukan pengujian, selanjutnya disajikan berupa informasi klasifikasi kelayakan penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda, serta tingkat kelayakannya yaitu dinyatakan layak, dan tidak layak. Selain itu, juga memberikan persentase akurasi klasifikasi yang dihasilkan penelitian ini.

### IV. Hasil dan Pembahasan

#### A. Pengumpulan Data

Data yang digunakan pada penelitian ini diperoleh dari Institut Agama Islam Negeri (IAIN) Samarinda. Data yang digunakan adalah data mahasiswa penerima Bidikmisi tahun angkatan 2015-2018 di Institut Agama Islam Negeri (IAIN) Samarinda yang berjumlah 120 data, dengan atribut yang dimiliki seperti pada Tabel 9.

Tabel 9 Atribut Data

No.	Jenis Atribut	Alias
1	Status pekerjaan orang tua	C1
2	Jumlah penghasilan orang tua	C2
3	Jumlah anggota keluarga	C3
4	Status kepemilikan rumah	C4
5	Jumlah pengeluaran keluarga	C5
6	Status kepemilikan SKTM/KIP	C6

### B. Analisis Data

Pada tahap analisis data, penelitian ini menggunakan teknik *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dalam melakukan *data cleaning* dan transformasi data [12]. Tahap pertama yaitu *cleaning data*. Pada tahap ini dilakukan dengan memeriksa atribut data yang telah didapatkan dari Institut Agama Islam Negeri (IAIN) Samarinda. Setelah melakukan pemeriksaan atribut data tersebut, tidak ditemukan atribut yang kosong sehingga dilanjutkan ke tahap selanjutnya. Langkah kedua adalah transformasi data. Pada tahap ini melakukan transformasi nilai setiap masing-masing atribut berdasarkan interval yang telah ditentukan seperti contoh pada Tabel 10.

Tabel 10 Transformasi Atribut

C1	C2	C3	C4	C5	C6	Hasil
3	3	2	2	1	2	Layak
3	3	2	2	2	2	Layak
2	3	2	3	2	2	Layak
2	2	1	2	2	2	Layak

### C. Implementasi Algoritma ID3

Tabel 11 Hitung Manual Node (Akar)

Node Akar	Jumlah	Layak	Tidak Layak	Entropy	Gain
	120	118	2	0.122	
C1					0.053
3	77	76	1	0.018	
2	42	41	1	0.016	
1	0	0	0	0	
C2					0.110
3	110	110	0		
2	9	8	1	0.151	
1	1	0	1	0	
C3					0.043
3	7	7	0	0	
2	98	97	1	0.082	
1	15	14	1	0.092	
C4					0.043
3	29	29	0	0	
2	71	69	2	0.185	
1	20	20	0	0	
C5					0.077
3	6	5	1	0.650	
2	55	54	1	0.025	
1	59	59	0		
C6					0.002
2	107	105	2	0.134	
1	3	3	0	0	

Pada tahap ini melakukan perhitungan manual menggunakan data yang sudah di transformasikan. Pada perhitungan manual ini melakukan pencaharian *node* dengan menghitung nilai *entropy* dan *gain*. *Gain* tertinggi menjadi rujukan untuk menentukan *node*.

#### 1) Node (akar)

- Menghitung *Entropy* total.  $Entropy(\text{Total}) = -(118/120) * (\log_2(118/120)) + -(2/120) * (\log_2(2/120)) = 0.122$ ;
- Setelah itu menghitung nilai *entropy* dari setiap atribut, sebagai contoh yaitu menghitung nilai *entropy* dari atribut skala penghasilan seperti berikut:  $Entropy(\text{skala penghasilan} = \text{Rating } 2) = -(8/9) * (\log_2(8/9)) + -(1/9) * (\log_2(1/9)) = 0.503$ ;

- Setelah menghitung semua nilai *entropy* semua atribut, kemudian menghitung nilai *gain*. Sebagai contoh yaitu pada atribut penghasilan seperti berikut:  $\text{Gain (Penghasilan)} = 0.122 - (((110/120) * 0) + ((9/120) * 0.151) + ((1/120) * 0)) = 0.110$ .

Berdasarkan langkah langkah tersebut, hasil perhitungan dapat diketahui. Nilai *entropy* dan *gain* setiap atribut seperti pada Tabel 11. Dari Tabel 11 dapat diketahui bahwa *gain* tertinggi dimiliki oleh variabel Penghasilan (C2). Dari data atribut penghasilan di atas dapat diketahui kelas tujuan (label) status yang dinyatakan LAYAK berjumlah 110 dari 118 dan status yang dinyatakan TIDAK LAYAK berjumlah 1 dari 2. Sehingga masih terdapat 9 data lagi yang harus dilakukan perhitungan untuk menentukan *node* (cabang) kedua.

### 2) Node (Cabang)

- Menghitung *entropy* total (Penghasilan).  $\text{Entropy(Total)} = (- (8/9) * (\log_2(8/9))) + (- (1/9) * (\log_2(1/9))) = 0.503$ ;
- Setelah itu menghitung nilai *entropy* dari setiap atribut, sebagai contoh yaitu menghitung nilai *entropy* dari atribut skala penghasilan seperti berikut:  $\text{Entropy (Skala Anggota Keluarga = Rating 1)} = (- (1/2) * (\log_2(1/2))) + (- (1/2) * (\log_2(1/2))) = 1$ ;
- Setelah menghitung semua nilai *entropy* semua atribut, kemudian menghitung nilai *gain*. Sebagai contoh yaitu pada atribut pekerjaan seperti berikut:  $\text{Gain (Penghasilan)} = 0.503 - (((2/9) * 0) + ((7/9) * 1)) = 0.043$ ;

Dari langkah tersebut, hasil perhitungan dapat diketahui. Hasil perhitungan *entropy* dan *gain* skala penghasilan seperti pada Tabel 12. Dari Tabel 12 dapat diketahui bahwa *gain* tertinggi dimiliki oleh variabel Anggota Keluarga (C3). Dari data atribut penghasilan di atas dapat diketahui kelas tujuan (label) status yang dinyatakan LAYAK berjumlah 7, dan masih tersisa 2 data lagi yang belum diketahui status kelayakannya sehingga harus dilakukan perhitungan untuk menentukan *node* (daun) ketiga.

Tabel 12 Hitungan Manual *Node* (Cabang)

Node Cabang (C2 – 2)	Jumlah	Layak	Tidak Layak	Entropy	Gain
	9	8	1	0.503	
3	2	2	0	0	
2	7	6	1	0.591	
1	0	0	0	0	
<b>C3</b>					0.281
3	0	0	0	0	
3	7	7	0	0	
1	2	1	1	1	
<b>C4</b>					0.020
3	0	0	0	0	
2	8	7	1	0.543	
1	1	0	1	0	
<b>C5</b>					0.142
3	4	4	0	0	
2	4	3	1	0.811	
1	1	1	0	0	
<b>C6</b>					0
2	9	8	1	0.503	
1	0	0	0	0	

### 3) Node (Daun)

- Menghitung *Entropy* total (Anggota Keluarga).  $\text{Entropy(Total)} = (- (1/2) * (\log_2(1/2))) + (- (1/2) * (\log_2(1/2))) = 1$ ;
- Setelah itu menghitung nilai *entropy* dari setiap atribut, sebagai contoh yaitu menghitung nilai *entropy* dari atribut skala rumah seperti berikut:  $\text{Entropy (Skala Rumah = Rating 1)} = (- (1/1) * (\log_2(1/1))) + (- (1/1) * (\log_2(1/1))) = 0$ ;
- Setelah menghitung semua nilai *entropy* semua atribut, kemudian menghitung nilai *gain*. Sebagai contoh yaitu pada atribut pekerjaan seperti berikut:  $\text{Gain (Rumah)} = 1 - (((1/2) * 0) + ((1/2) * 1)) = 1$ ;

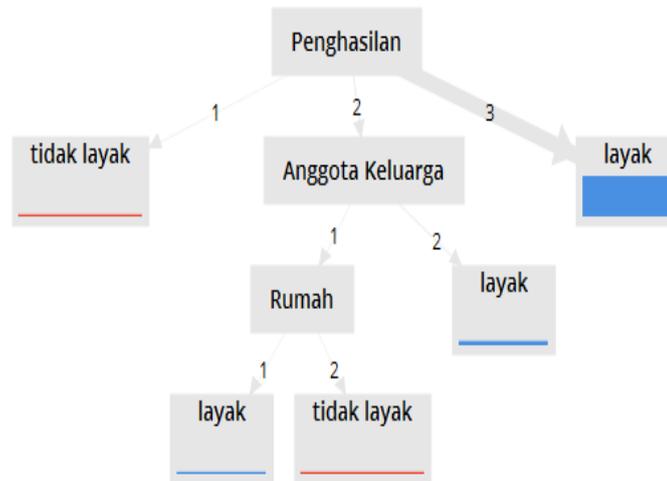
Dari perhitungan tersebut, hasil perhitungan dapat diketahui. Hasil perhitungan *entropy* dan *gain* skala penghasilan seperti pada Tabel 13. Dari Tabel 13 dapat diketahui bahwa *gain* tertinggi dimiliki oleh variabel Rumah (C4) karena lebih dekat dari *node* sebelumnya. Dari data atribut rumah di atas dapat diketahui kelas tujuan (label) status yang dinyatakan LAYAK berjumlah 1, dan status yang dinyatakan TIDAK LAYAK berjumlah 1.

Tabel 13 Hitungan Manual *Node* (Cabang)

Node Cabang (C3 – 1)	Jumlah	Layak	Tidak Layak	Entropy	Gain
	2	1	1	1	
C1					0
3	0	0	0	0	
2	2	1	1	1	
1	0	0	0	0	
C4					1
3	0	0	0	0	
2	1	0	1	0	
1	1	1	0	0	
C5					1
3	1	1	0	0	
2	1	0	1	0	
1	0	0	0	0	
C6					0
2	2	1	1	1	
1	0	0	0	0	

D. Pembentukan Model

Pada tahap ini pembentukan model menggunakan *RapidMiner* yang akan menghasilkan *output* berupa *rule* pohon keputusan. Pembentukan model tersebut dapat dilihat pada Gambar 1.



Gambar. 1. *Rule* Pohon Keputusan

E. Pengujian

Pada tahap pengujian dilakukan perbandingan hasil prediksi klasifikasi ID3 dengan klasifikasi kenyataan menggunakan data uji atau *data testing*. Hal ini dapat dilihat pada Tabel 14. Berdasarkan Tabel 14 menggunakan data uji atau *data testing* dengan membandingkan data nyata, maka dapat dihitung nilai akurasinya melalui Tabel 15.

Tabel 14 Evaluasi Prediksi ID3

C1	C2	C3	C4	C5	C6	Data Nyata	Hasil ID3
3	3	1	2	1	2	layak	layak
3	3	2	2	2	2	layak	layak
3	3	2	2	2	2	layak	layak
3	3	2	2	2	2	layak	layak
2	3	1	2	1	2	layak	layak
2	2	1	2	2	2	layak	Tidak layak
3	3	2	3	2	2	layak	layak
3	3	3	2	1	2	layak	layak
3	3	2	2	1	2	layak	layak
3	2	2	2	2	2	layak	layak
3	3	2	3	1	2	layak	layak
2	3	2	1	2	2	layak	layak
2	1	2	2	3	2	layak	Tidak layak
2	3	2	1	2	2	layak	layak

C1	C2	C3	C4	C5	C6	Data Nyata	Hasil ID3
2	3	3	2	1	2	layak	layak
2	3	1	2	1	2	layak	layak
3	3	3	2	2	2	layak	layak
2	2	1	1	3	2	layak	layak
2	3	2	1	2	2	layak	layak
2	3	2	2	3	1	layak	layak
3	3	2	2	1	2	layak	layak
3	3	2	1	2	2	layak	layak
2	3	3	2	2	2	layak	layak
3	3	3	2	3	2	layak	layak
2	3	1	1	1	2	layak	layak
2	3	2	2	1	2	layak	layak
2	3	2	3	2	2	layak	layak
2	3	3	3	1	2	layak	layak
3	3	2	2	1	2	layak	layak

Tabel 15 Confusion Matrix Dataset

	true layak	true tidak layak
pred. layak	118	0
pred. tidak layak	2	0

Dari tabel 15 *confusion matrix* dapat dihitung nilai akurasi menggunakan persamaan Akurasi (AC) =  $\frac{118+0}{118+0+2+0} \times 100\% = 98.3\%$  dan tingkat kesalahan (*error rate*) menggunakan persamaan *Error rate* =  $\frac{2+0}{118+0+2+0} \times 100\% = 1.7\%$ .

#### F. Pembahasan

Dari hasil implementasi dan pengujian algoritma ID3 maka dapat diketahui bahwa variabel penghasilan merupakan atribut utama dalam memprediksi klasifikasi kelayakan penerima bidikmisi, kemudian dilanjutkan ketahap analisis variabel *node* berikutnya. Aturan yang terbentuk hasil analisis sebagai berikut: (1) Jika penghasilan orang tuanya dibawah/maksimal Rp. 2.000.000,00 (*Rating 3*) maka mahasiswa tersebut dinyatakan layak menerima bidikmisi; (2) Jika penghasilan orang tuanya di atas Rp. 4.000.000,00 (*Rating 1*) maka mahasiswa tersebut dinyatakan tidak layak menerima bidikmisi; (3) Jika penghasilan orang tuanya di atas Rp. 2.000.000,00 – Rp. 4.000.000 maka mahasiswa tersebut akan ditinjau kembali dengan pertimbangan yang telah ditetapkan: jika anggota keluarganya berjumlah 4 sampai 7 orang maka mahasiswa tersebut dinyatakan layak menerima bidikmisi, sedangkan jika anggota keluarganya berjumlah kurang dari 4 orang dan status rumahnya belum punya rumah/numpang dengan keluarga maka mahasiswa tersebut dinyatakan layak menerima bidikmisi; dan (4) Jika penghasilan orang tuanya diatas Rp. 2.000.000,00 – Rp. 4.000.000, anggota keluarganya berjumlah kurang dari 4 orang, dan status rumahnya memiliki rumah sendiri maka mahasiswa tersebut dinyatakan tidak layak menerima bidikmisi.

Berdasarkan aturan yang dibentuk melalui pohon keputusan klasifikasi kelayakan penerima bidikmisi, maka 120 data mahasiswa penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda selanjutnya dikelompokkan. Pengelompokannya adalah sebagai berikut:

##### 1) Penghasilan

110 dinyatakan layak menerima Bidikmisi dengan penghasilan orang tuanya di bawah Rp. 2.000.000 rupiah, dan 1 dinyatakan tidak layak menerima Bidikmisi dengan penghasilan orang tuanya melebihi Rp. 4.000.000 rupiah.

##### 2) Anggota Keluarga

7 dinyatakan layak menerima Bidikmisi dengan penghasilan keluarga Rp. 2.000.000 – 4.000.000 rupiah dan jumlah anggota keluarganya 4-7 orang.

##### 3) Kepemilikan Rumah

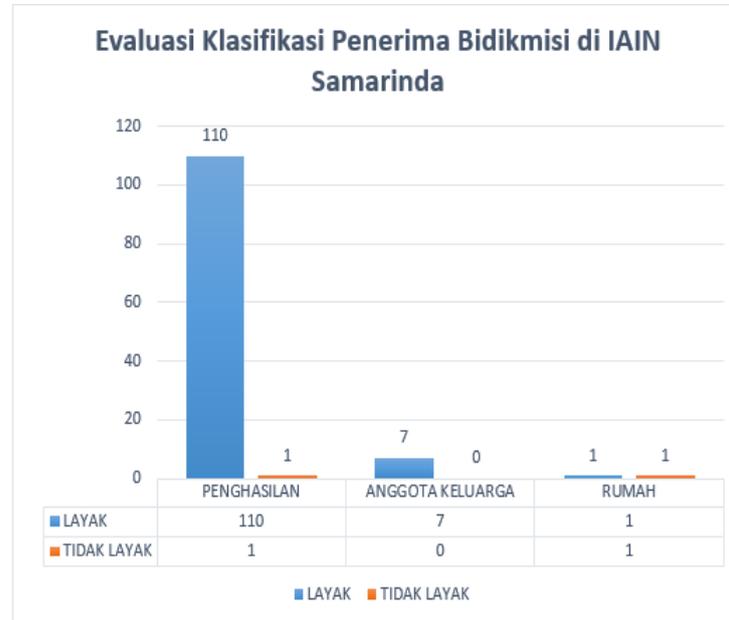
1 dinyatakan layak menerima Bidikmisi dengan penghasilan keluarga Rp.2.000.000 – 4.000.000 rupiah, jumlah anggota keluarganya 4-7 orang dan status kepemilikan rumah (tempat tinggal) belum punya rumah. Kemudian 1 dinyatakan tidak layak menerima Bidikmisi dengan penghasilan keluarga Rp.2.000.000 – 4.000.000 rupiah, jumlah anggota keluarganya 4-7 orang dan status kepemilikan rumah (tempat tinggal) rumah sendiri.

Setelah melakukan analisis 120 data dengan mengacu pada aturan yang terbentuk dari pohon keputusan klasifikasi kelayakan penerima Bidikmisi maka 2 data mahasiswa penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda dinyatakan tidak layak dengan pertimbangan berikut:

- 1 mahasiswa dengan penghasilan orang tuanya melebihi Rp. 4.000.000 (*Rating 1*), kemudian

- mahasiswa dengan penghasilan orang tuanya di atas Rp. 2.000.000 – 4.000.000 (*Rating2*), kemudian jumlah anggota keluarganya kurang dari 4 orang (*Rating 1*), dan status tempat tinggal memiliki rumah sendiri (*Rating 2*).

Evaluasi klasifikasi penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda dikelompokkan berdasarkan penghasilan orang tua, jumlah anggota keluarga dan status kepemilikan rumah. Evaluasi klasifikasi kelayakan tersebut disajikan pada Gambar 2.



Gambar. 2. Evaluasi Hasil Klasifikasi

## V. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat ditarik kesimpulan yaitu data penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda yang diperoleh peneliti, maka pohon keputusan merumuskan variabel penghasilan orang tua sebagai faktor utama untuk menentukan klasifikasi kelayakan penerima Bidikmisi. Kemudian variabel jumlah anggota keluarga, dan yang terakhir variabel status kepemilikan rumah. Selain itu, algoritma *Iterative Dichotomiser 3* (ID3) memiliki akurasi senilai 98.3% dan tingkat kesalahan (*error rate*) senilai 1.7% dalam memprediksi klasifikasi kelayakan penerima Bidikmisi.

Lebih lanjut, terdapat saran yang diberikan untuk penelitian selanjutnya. Pengembangan penelitian selanjutnya dapat dilakukan dengan melakukan komparasi algoritma, sehingga dapat diketahui algoritma yang lebih baik dibandingkan algoritma *Iterative Dichotomiser 3* (ID3) dalam menentukan klasifikasi kelayakan penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda. Kemudian, implementasi ke dalam sistem pendukung keputusan penentuan kelayakan calon penerima Bidikmisi di Institut Agama Islam Negeri (IAIN) Samarinda. Terakhir, dari hasil akurasi algoritma yang digunakan pada penelitian ini yang cukup tinggi, memberikan dorongan untuk penelitian selanjutnya untuk menggunakan algoritma *Iterative Dichotomiser 3* (ID3) dalam menentukan klasifikasi pada objek lain.

## Daftar Pustaka

- [1] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2005.
- [2] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publisher, 2006.
- [3] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [4] S. Defiyanti and D. L. C. Pardede, "Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi Spam-mail," Universitas Gunadarma, 2014.
- [5] L. Dwi, B. Pratiwi, W. Wibowo, and J. Statistika, "Klasifikasi Nilai Peminat SBMPTN ( Seleksi Bersama Masuk Perguruan Tinggi Negeri ) ITS dengan Pendekatan Classification and Regression Trees ( CART )," vol. 4, no. 2, pp. 2–7, 2015.

- [6] A. Gupta, S. Gupta, and D. Singh, "A Systematic Review of Classification Techniques and Implementation of ID3 Decision Tree Algorithm," pp. 144–152, 2015.
- [7] M. Learning *et al.*, "Induction of Decision Trees," *Expert Syst.*, pp. 81–106, 2007.
- [8] J. Adiana, E. Sedyono, and O. Dwi, "Pendekatan Metode Pohon Keputusan Menggunakan Algoritma ID3 Untuk Sistem Informasi Pengukuran Kinerja PNS," vol. 02, pp. 75–86, 2014.
- [9] S. Singh and P. Gupta, "Comparative study ID3, cart and C4 . 5 Decision tree algorithm: a survey," *Int. J. Adv. Inf. Sci. Technol.*, vol. 27, no. 27, pp. 97–103, 2014.
- [10] Wahyudin, "Metode Iterative Dichotomizer 3 (ID3) untuk Penyeleksian Penerimaan Mahasiswa Baru," *J. Pendidik. Teknol. Inf. dan Komun.*, vol. 2, no. 2, pp. 5–15, 2009.
- [11] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of database systems*, L. Liu and M. T. Özsu, Eds. Boston, MA: Springer, 2009.
- [12] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 2000, pp. 29–39.