



Pemodelan Data Mining *Decision Tree* Dengan *Classification Error* Untuk Seleksi Calon Anggota Tim Paduan Suara

Patmi Kasih

¹Universitas Nusantara PGRI Kediri, Jl. KH.Ahmad Dahlan 76 Mojoroto, Kediri, Jawa Timur, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 07 September 2019

Revisi Akhir: 11 Oktober 2019

Diterbitkan Online: 14 Oktober 2019

KATA KUNCI

Decision Tree,
Classification Error,
Pemodelan,
Seleksi,
Rule's base

KORESPONDENSI

Telepon +6281210385211

E-mail: fatkasih@gmail.com

A B S T R A C T

The process of accepting new members in a team requires clarity of the evaluation criteria and the accuracy of the process in the assessment. Likewise, what is wanted by the Harmoni Nusantara Choir Team Builder at Nusantara University PGRI Kediri. During this time the selection process for the recruitment of team members is carried out conventionally with interviews and live voice tests. To decide whether a selection participant is accepted or not, the guiding team must conduct discussions that often occur when disputes arise between coaches when one participant has the same or balanced results with the other selection participants. Therefore we need a system (application) with the implementation of certain algorithms that can help in the selection process for new members. To realize the planned system, a Decision Tree algorithm modeling with the Classification Error concept is used to analyze an algorithm that is suitable to be used as an assistive algorithm in decision making. Decision Tree modeling with the concept of Classification Error is done by training data with a total of 60 records with attributes input gender, interpretation, technique, appearance, commitment, octave. The target class specified is accepted or not accepted. Based on the modeling done, a decision tree is obtained that produces 5 (five) rule's bases that cover all records in the training data. So it can be concluded that 100% training data records (60 records) are covered in the rule's base, and concluded that the Decision Tree with the Classification Error concept can be used as an assist algorithm that will be implemented in the system (application) to assist selection of new members of the Harmoni Nusantara Choir Team on Planned Nusantara Nusantara PGRI Kediri.

1. PENDAHULUAN

Unit Kegiatan Mahasiswa (UKM) Paduan Suara Harmoni Nusantara berdiri di bawah naungan BEM Universitas Nusantara PGRI Kediri dan merupakan kegiatan mahasiswa di bidang seni suara atau vocal. Setiap tahun tim paduan suara mengadakan seleksi penerimaan anggota baru dari mahasiswa baru yang ingin bergabung dalam UKM Paduan Suara Harmoni Nusantara. Tim seleksi adalah tim Pembina dan pengurus UKM Paduan suara.

Seperti umumnya pada tim paduan suara, dalam tim paduan suara Harmoni Nusantara terdapat 4 (empat) jenis suara yang dibutuhkan dari anggota tim yaitu *sopran* dan *alto* untuk suara perempuan, *tenor* dan *bass* untuk suara laki-laki. Kebutuhan jenis suara untuk tim ditentukan oleh tim Pembina paduan suara kampus. Proses penerimaan anggota baru dilakukan dengan cara konvensional, tes

suara dan wawancara langsung oleh Pembina tim. Penentuan diterima atau tidak juga masih bersifat subjektif dan memakan waktu yang lama karena tim pembina harus melakukan diskusi untuk memberikan hasil peserta audisi/ seleksi di terima atau tidak.

Terinspirasi dari penelitian Ratih Kumalasari Niswatin, 2017, dengan judul "Penerapan algoritma *Decision Tree* pada penentuan keberhasilan akademik mahasiswa", yang memanfaatkan *Decision Tree C4.5* untuk melakukan prediksi/ klasifikasi mahasiswa untuk pengelompokan mahasiswa berdasarkan kriteria latar belakang pendidikan mahasiswa yaitu rata-rata nilai raport, nilai uan dan jurusan disekolah asal. Hasil penelitian digunakan oleh panitia penerimaan mahasiswa baru dan pihak program studi untuk menentukan pengelompokkan calon mahasiswa baru. [1]

Berdasarkan alasan yang telah dipaparkan dan inspirasi penelitian sebelumnya, peneliti melakukan

pemodelan dengan algoritma *Decision Tree* untuk permasalahan pemilihan anggota baru tim paduan suara agar proses seleksi dan pengambilan keputusan penerimaan anggota baru dapat dilakukan dengan lebih cepat dan fair. Solusi yang direncanakan adalah pembuatan suatu sistem (aplikasi) yang dapat digunakan sebagai sistem bantu dalam proses seleksi dan pengambilan keputusan penerimaan anggota baru dalam tim paduan suara Harmoni Nusantara. Sebagai langkah awal mewujudkan solusi sistem bantu bagi permasalahan tim Pembina paduan suara, dibuatlah suatu penelitian pemodelan data mining *Decision Tree* dengan *Classification Error* sebagai model analisa untuk melakukan klasifikasi peserta seleksi/ audisi diterima atau tidak berdasarkan kriteria penilaian yang telah ditetapkan oleh tim seleksi.

2. LANDASAN TEORI

2.1 Perencanaan Pemodelan Kasus dengan *Decision Tree* dengan *Classification Error*.

Perencanaan pemodelan ini digunakan sebagai pedoman peneliti dalam pelaksanaan penelitian agar hasil yang dicapai tetap mengarah ke tujuan awal. Pemodelan *Decision Tree* dengan implementasi *Classification Error* (selanjutnya disingkat *C. Error*) untuk seleksi anggota baru dalam tim paduan suara ini, dilakukan dengan tahap awal sebagai berikut:

- Mencari informasi mengenai cara seleksi anggota baru dalam tim paduan suara Harmoni Nusantara, apa saja yang menjadi kriteria penilaian, dan bagaimana pengambilan keputusan yang dilakukan.
- Melakukan kajian literatur dan kajian artikel dari penelitian terdahulu dengan kasus serupa, serta mempelajari dan menganalisa kemungkinan algoritma yang tepat untuk digunakan.
- Melakukan analisa algoritma klasifikasi data mining yang cocok untuk dijadikan algoritma bantu dalam melakukan pengambilan keputusan akhir seleksi anggota baru tim paduan suara.
- Menentukan kriteria penilaian dari tim pembina Paduan Suara Harmoni Nusantara (gender, teknik, interpretasi, penampilan, komitmen, oktaf) sebagai atribut dalam dataset, dengan target atribut "diterima" dan "tidak diterima". Jumlah data yang akan digunakan data pembelajaran (training) adalah 60 data peserta.
- Melakukan proses pemodelan dengan *Decision Tree* menggunakan konsep nilai *C. Error* dengan atribut yang telah ditentukan.
- Melakukan analisa hasil pemodelan dan membuat kesimpulan akhir untuk langkah selanjutnya dalam mewujudkan sistem bantu bagi permasalahan dalam proses seleksi penerimaan anggota baru tim paduan suara.

2.2 *Decision Tree* dengan *Classification Error*

2.2.1 Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. [2]

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. [3]

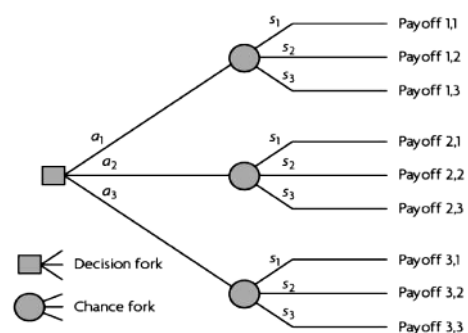
Data mining mengolah data mentah yang tersimpan di dalam basis data sehingga menghasilkan informasi yang berguna. Data mentah merupakan data yang disimpan hanya sebagai dokumentasi. Pengolahan data mentah dapat dimanfaatkan untuk pembuatan laporan, mengontrol kebutuhan, menemukan informasi dan pengetahuan baru. Data mining menelusuri data pada database untuk membangun model dan menggunakannya untuk mengenali pola data lain yang tidak tersimpan dalam basis data.

2.2.2 Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang populer adalah dengan *Decision Tree* yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. *Decision Tree* adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

2.2.3 *Decision Tree*

Decision Tree adalah struktur flowchart yang menyerupai *Tree* (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada *Decision Tree* di telusuri dari simpul akar ke simpul daun yang memegang prediksi. [5]



Gambar 1. Bentuk *Decision Tree* Secara Umum

Decision Tree adalah sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal

node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian tersebut dan leaf node menyatakan kelas-kelas/ distribusi kelas. Node yang paling atas disebut sebagai *root node* atau node akar, memiliki beberapa edge keluar tetapi tidak memiliki edge masuk. Internal node akan memiliki satu edge masuk dan beberapa edge keluar, sedangkan leaf node hanya akan memiliki satu edge masuk tanpa memiliki edge keluar. Leaf node adalah hasil akhir yang memiliki label kelas dari kombinasi atribut yang terbentuk menjadi *rule*.

Dalam membangun sebuah *Decision Tree* secara top-down (dari atas ke bawah), tahap awal yang dilakukan adalah mengevaluasi semua atribut yang ada menggunakan suatu ukuran statistik (yang biasa digunakan adalah *information gain*) untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan suatu kumpulan sampel data. Atribut yang diletakkan pada *root_node* adalah atribut yang memiliki *information gain* terbesar. Semua atribut adalah bersifat kategori yang bernilai diskrit. Atribut dengan nilai continuous harus didiskritkan.[4]

Nilai Gain adalah *Information Gain* yang digunakan untuk mencari satu variable/ atribut dari dataset (S) untuk dijadikan *root/ node* dan *branch node*, yaitu satu atribut yang mempunyai nilai gain tertinggi. Untuk pencarian nilai gain (*information Gain*) dapat dilakukan dengan konsep **Entropy**, **Gini Index** dan **Classification Error**. Nilai gain terbesar yang di dapat dari atribut pada dataset (data pembelajaran) pertama kali digunakan untuk mencari atribut yang layak menjadi root (akar) pohon keputusan (*decision Tree*). Selanjut proses akan diulang untuk mencari atribut yang menjadi cabang hingga menemukan leaf yang merupakan label kelas.

- Information Gain dengan Nilai Entropy
Entropy adalah formula untuk menghitung homogenitas atribut (A) dari sebuah sample data (S). Dengan formula:

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

S = Himpunan Kasus dalam dataset

A = Fitur (atribut)

n = jumlah partisi atribut S

Pi = proposi dari Si terhadap S

maka:

$$Entropy(S) = \sum_{i=1}^n p(i|s) \log_2 p(i|s)$$

Maka Gain(S,A) adalah Information Gain dari attribute A pada koleksi contoh S.

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

- Information Gain dengan Nilai Gini Index

$$Gini(t) = 1 - \sum_{i=1}^n [p(i|s)]^2 \quad (3)$$

$$Gain(S,A) = Gini(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Gini(S_i) \quad (4)$$

- Information Gain dengan Classification Error
Nilai C. Error diperoleh dari nilai value atribut yang terkecil dari terhadap class label.

$$C. Error(S) = 1 - \max_i [p(i|s)] \quad (5)$$

$$Gain(S,A) = Error(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Error(S_i) \quad (6)$$

Layaknya sebuah pohon, *Decision Tree* mempunyai akar (*root/ node*), batang/cabang (*branch node*), dan daun (*leaf*). Strategi yang dapat digunakan untuk pembentukan pohon keputusan dengan *Decision Tree* adalah:

- Pohon dimulai sebagai node tunggal (akar/ root) yang merepresentasikan semua data.
- Sesudah node root dibentuk, maka data pada node akar akan diukur dengan *information gain* untuk dipilih atribut mana yang akan dijadikan atribut pembagiannya.
- Sebuah cabang dibentuk dari atribut yang dipilih menjadi pembagi dan data akan didistribusikan ke cabang masing-masing.
- Algoritma ini akan terus menggunakan proses yang sama (rekursif) untuk dapat membentuk sebuah *Decision Tree* . Ketika sebuah atribut telah dipilih menjadi node pembagi/ cabang, maka atribut tersebut tidak diikuti lagi dalam penghitungan nilai *information gain*.
- Proses pembagian rekursif akan berhenti jika salah satu dari kondisi dibawah ini terpenuhi:
 1. Semua data dari anak cabang telah termasuk dalam kelas yang sama.
 2. Semua atribut telah dipakai, tetapi masih tersisa data dalam kelas yang berbeda. Dalam kasus ini, diambil data yang mewakili kelas yang terbanyak untuk menjadi label kelas pada node daun.
 3. Tidak terdapat data pada anak cabang yang baru. Dalam kasus ini, node daun akan dipilih pada cabang sebelumnya dan diambil data yang mewakili kelas terbanyak untuk dijadikan label kelas.

3. METODOLOGI

Tahapan penelitian yang dilakukan mulai dari pengumpulan data, studi literatur, pengembangan perangkat lunak, pengujian dan penarikan kesimpulan.

4. HASIL DAN PEMBAHASAN

Pemodelan *Decision Tree* untuk mencari algoritma bantu dalam seleksi anggota baru tim paduan Suara Harmoni Nusantara Universitas Nusantara PGRI Kediri menggunakan data training (data pembelajaran) yang telah disiapkan oleh Tim Pembina Paduan Suara harmoni Nusantara. Data training yang digunakan sejumlah 60 record, yang merupakan data pembelajaran dari hasil seleksi (audisi) yang dilakukan oleh tim pembina paduan suara pada tahun 2018. Data pembelajaran mempunyai 6 (enam) atribut input yang merupakan kriteria penilaian yaitu Gender (G), Teknik (T), Interpretasi (Ip), Penampilan (P), Komitmen (K), Oktaf (Ot) dan 1 (satu) atribut target yaitu Hasil dengan class target adalah "L" untuk lulus (diterima) dan "TL" untuk tidak lulus (tidak diterima).

Value_Atribut untuk masing-masing atribut adalah:

- Gender (G) = Perempuan (P), Laki-laki (Lk)
- Teknik (T) = Baik (B), Cukup (C), Kurang (K)
- Interpretasi (Ip) = Baik (B), Cukup (C)
- Penampilan (P) = Baik (B), Cukup (C), Kurang (K)
- Komitmen (K) = Ya (Y), Tidak (T)
- Oktaf (Ot) = Rendah (R), Sedang (S), Tinggi(T)

Tabel 1. Data Pembelajaran

No	Nama	G	T	Ip	P	K	Ot	Hasil
1	Wahyu	Lk	B	B	B	Y	S	L
2	Agustia	P	C	B	C	T	T	TL
3	Nadila	P	B	B	B	Y	T	L
4	Daniel	Lk	C	C	C	T	T	TL
5	Arum	P	B	B	B	T	R	TL
6	Anita	P	B	B	B	Y	S	L
7	Christin	P	B	B	B	Y	T	L
8	Sukma	P	B	B	B	Y	T	L
9	Heni	P	B	B	B	Y	T	L
10	Galih	Lk	B	B	B	Y	S	L
11	Rudi	Lk	C	C	B	Y	S	L
12	Calvin	Lk	B	B	B	Y	T	L
13	Hardo	Lk	B	B	B	Y	S	L
14	Nantha	Lk	C	B	C	T	S	TL
15	Riandika	Lk	B	C	B	T	R	TL
16	Youanda	Lk	B	B	B	Y	T	L
17	Monic	P	B	B	B	Y	S	L
18	Alliffah	P	B	B	B	Y	T	L
19	Adi	Lk	C	B	B	Y	R	L
20	Eko	Lk	C	B	C	Y	R	L
21	Renova	P	C	C	C	T	T	TL
22	Diana	P	C	C	B	T	R	TL
23	Rinda	P	C	C	C	T	R	TL
24	Chyntia	P	B	C	C	T	T	TL
25	Cici	P	B	B	B	Y	S	L
26	Bintang	P	B	B	C	Y	R	L
27	Irma	P	C	B	B	Y	S	L
28	Ira	P	B	C	B	Y	R	L
29	Bella	P	C	C	C	T	R	TL
30	Arif	Lk	B	B	C	Y	S	L
31	Djoni	Lk	B	B	B	Y	S	L
32	Galang	Lk	C	B	B	Y	S	L
33	Danny	Lk	C	B	C	Y	S	L
34	Aming	Lk	B	B	C	T	T	TL
35	Ari	Lk	C	B	B	T	T	TL
36	Prima	P	C	B	B	Y	T	L

37	Devi	P	B	B	B	Y	S	L
38	Cindy	P	B	C	B	Y	T	L
39	Rizka	P	K	C	K	Y	R	TL
40	Elya	P	K	C	C	Y	R	TL
41	Dita	P	C	B	C	Y	R	L
42	Nanda	P	K	K	C	Y	R	TL
43	Ilma	P	C	K	K	T	R	TL
44	Iis	P	B	C	B	Y	T	L
45	Wiji	P	B	B	B	T	S	TL
46	Dias	P	B	B	B	T	S	TL
47	Yohanes	Lk	B	B	B	T	T	TL
48	Agung	Lk	B	B	B	Y	T	L
49	Sugeng	Lk	B	C	B	Y	T	L
50	Mada	P	B	B	B	Y	T	L
51	Sandra	P	B	B	C	T	T	TL
52	Gista	P	B	C	B	T	T	TL
53	Wulan	P	B	B	C	Y	S	L
54	Gunawan	Lk	B	C	B	Y	S	L
55	Haris	Lk	B	B	C	T	S	TL
56	Dimas	Lk	C	B	K	T	T	TL
57	Soleh	Lk	C	B	K	T	R	TL
58	Priska	P	B	B	C	T	T	TL
59	Nuri	P	C	K	C	Y	R	TL
60	Villa	P	K	C	K	Y	S	TL

Sumber: Tim Pembina Paduan Suara Harmoni Nusantara, 2018.

Untuk memperjelas dan mempermudah dalam perhitungan, maka data pembelajaran dapat direkap seperti tabel 2, berikut:

Tabel 2. Rekap Data Pembelajaran

Atribut	Jumlah	L	TL
Gender			
Lk	24	15	9
P	36	18	18
Teknik			
B	36	25	11
C	20	8	12
K	4	0	4
Interpretasi			
B	40	27	13
C	17	6	11
K	3	0	3
Penampilan			
B	35	27	8
C	20	6	14
K	5	0	4
Komitmen			
Y	38	33	5
T	22	0	22
Oktaf			
R	16	5	11
S	20	15	5
T	24	13	11
Hasil	60	33	27

Untuk memulai langkah perhitungan algoritma *Decision Tree*, hal yang harus diingat adalah bahwa konstruksi *Decision Tree* dilakukan secara *top-down* untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar (*root*) sampai *leaf* (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu.

Langkah terpenting diawali dengan pertanyaan: "Attribute mana yang harus diperiksa pada root dari decision Tree?". [6]

Dalam penelitian ini formula yang digunakan adalah formula (5) dan (6). Berikut penyelesaian *Decision Tree* menggunakan konsep *C. Error* untuk sistem seleksi penerimaan anggota baru tim paduan suara Harmoni Nusantara UN PGRI Kediri.

- Examples (S), adalah training examples/ data pembelajaran yang ditunjukkan oleh tabel 1.
- Target atribut adalah Hasil dengan value L (lulus) atau TL (tidak_lulus) bagi peserta seleksi.
- Atribut input adalah gender, teknik, interprestasi, penampilan, komitmen, oktaf.

Penentuan atribut sebagai root dalam *Tree* dilakukan dengan mencari nilai *Gain* dari seluruh atribut pada data pembelajaran, dengan terlebih dahulu harus mencari nilai *c. error* dari data pembelajaran dan seluruh atribut. Dengan bantuan tabel 1 diketahui dari 60 data pembelajaran, diketahui clas hasil untuk L= 33 (+), TL = 27 (-), maka nilai *c. error* untuk data pembelajaran adalah:

$$\begin{aligned}
 C. Error (t) &= 1 - \max[(33/60), (27/60)] \\
 &= 1 - 33/60 \\
 &= 27/60 \\
 &= 0.45
 \end{aligned}$$

Untuk selanjutnya dapat dilakukan pencarian nilai *Gain* dari semua atribut input. Sebagai contoh adalah pencarian nilai *gain* untuk atribut gender:

Value Gender = (Lk, P)

$$\begin{aligned}
 S (Lk) &= (15+,9-) \\
 C. Error (Lk) &= 1 - \max((15/24), (9/24)) \\
 &= 1 - (15/24) \\
 &= \mathbf{0.375}
 \end{aligned}$$

$$\begin{aligned}
 S (P) &= (18+,18) \\
 C. Error (P) &= 1 - \max((18/36),(18/36)) \\
 &= 1 - (18/36) \\
 &= \mathbf{0.5}
 \end{aligned}$$

Maka nilai *Gain* dari atribut Gender adalah:

$$\begin{aligned}
 Gain(S, Gender) &= Error(S) - 24/60. Error(Lk) - \\
 &\quad 36/60.Error(P) \\
 &= 0.45 - 0.4*0.375 - 0.6*0.5 \\
 &= 0.45 - 0.15 - 0.3 \\
 &= \mathbf{0}
 \end{aligned}$$

Dengan cara yang sama dilakukan pencarian nilai *C.error* dan nilai *Gain* untuk seluruh atribut input, dengan hasil seperti pada tabel 3.

Table 3. Hasil Pencarian Nilai Gain untuk Root

Atribut	Value Atribut	Jum.	L	TL	C.	Gain
		Kasus			Error	
Gender						
	Lk	24	15	9	0,375	0,000
	P	36	18	18	0,500	
Teknik						
	B	36	25	11	0,306	0,133
	C	20	8	12	0,400	
	K	4	0	4	0,000	
Interpres trasi						
	B	40	27	13	0,325	0,133
	C	17	6	11	0,353	
	K	3	0	3	0,000	
Penampilan						
	B	35	27	8	0,229	0,200
	C	20	6	14	0,300	
	K	5	0	4	0,200	
Komitmen						
	Y	38	33	5	0,132	0,367
	T	22	0	22	0,000	
Oktaf						
	R	16	5	11	0,313	0,100
	S	20	15	5	0,250	
	T	24	13	11	0,458	

Dari tabel 3 diketahui nilai *gain* tertinggi didapat dari atribut *Komitmen*, maka dapat diputuskan bahwa atribut *komitmen* terpilih sebagai root dari *Tree* yang dibentuk. Selanjutnya mencari atribut yang akan menjadi cabang, dengan cara memilih nilai *c.error* terbesar dari atribut *komitmen*, yaitu *komitmen* dengan value=Y, dengan jumlah 38 data dari 60 data awal, seperti pada tabel 4. Maka data pada tabel 4 menjadi data pembelajaran baru.

Table 4. Data Hasil untuk Branch Node *Komitmen*=Y

No	Nama	G	T	Ip	P	K	Ot	Hasil
1	Wahyu	Lk	B	B	B	Y	S	L
3	Nadila	P	B	B	B	Y	T	L
6	Anita	P	B	B	B	Y	S	L
7	Christin	P	B	B	B	Y	T	L
8	Sukma	P	B	B	B	Y	T	L
9	Heni	P	B	B	B	Y	T	L
10	Galih	Lk	B	B	B	Y	S	L
11	Rudi	Lk	C	C	B	Y	S	L
12	Calvin	Lk	B	B	B	Y	T	L
13	Hardo	Lk	B	B	B	Y	S	L
16	Youanda	Lk	B	B	B	Y	T	L
17	Monic	P	B	B	B	Y	S	L
18	Alliffah	P	B	B	B	Y	T	L
19	Adi	Lk	C	B	B	Y	R	L
20	Eko	Lk	C	B	C	Y	R	L
25	Cici	P	B	B	B	Y	S	L
26	Bintang	P	B	B	C	Y	R	L
27	Irma	P	C	B	B	Y	S	L
28	Ira	P	B	C	B	Y	R	L
30	Arif	Lk	B	B	C	Y	S	L
31	Djoni	Lk	B	B	B	Y	S	L
32	Galang	Lk	C	B	B	Y	S	L
33	Danny	Lk	C	B	C	Y	S	L
36	Prima	P	C	B	B	Y	T	L
37	Devi	P	B	B	B	Y	S	L

38	Cindy	P	B	C	B	Y	T	L
39	Rizka	P	K	C	K	Y	R	TL
40	Elya	P	K	C	C	Y	R	TL
41	Dita	P	C	B	C	Y	R	L
42	Nanda	P	K	K	C	Y	R	TL
44	Iis	P	B	C	B	Y	T	L
48	Agung	Lk	B	B	B	Y	T	L
49	Sugeng	Lk	B	C	B	Y	T	L
50	Mada	P	B	B	B	Y	T	L
53	Wulan	P	B	B	C	Y	S	L
54	Gunawan	Lk	B	C	B	Y	S	L
59	Nuri	P	C	K	C	Y	R	TL
60	Villa	P	K	C	K	Y	S	TL

Tabel 5. Pencarian Gain untuk Branch Node Komitmen=Y

Atribut	Value Atribut	J.Kasus	L	TL	Error	Gain
Komitmen	Y	38	33	5	0,132	
Gender						
	Lk	15	15	0	0,000	0,000
	P	23	18	5	0,217	
Teknik						
	B	25	25	0	0,000	0,105
	C	9	8	1	0,111	
	K	4	0	4	0,000	
Interprestasi						
	B	27	27	0	0,000	0,053
	C	9	6	3	0,333	
	K	2	0	2	0,000	
Penampilan						
	B	27	27	0	0,229	0,053
	C	9	6	3	0,333	
	K	2	0	2	0,000	
Oktaf						
	R	9	5	4	0,444	-2,77
	S	16	15	1	0,063	
	T	13	13	0	0,000	

Dari tabel 5 diketahui nilai gain tertinggi didapat dari atribut Teknik=C, maka dapat diputuskan bahwa atribut Teknik=C terpilih sebagai branch node (cabang) dari batang komitmen=Y. Untuk branch node komitmen=T tidak perlu diproses lanjut, karena dari dataset sudah diketahui bahwa Komitmen=T tidak akan lulus seleksi. Langkah selanjutnya adalah pencarian branch node untuk Komitmen=Y and Teknik=C.

Dari seleksi data diperoleh dataset baru sejumlah 8 record seperti tabel 6.

Table 6. Data Branch Node Komitmen=Y and Teknik=C

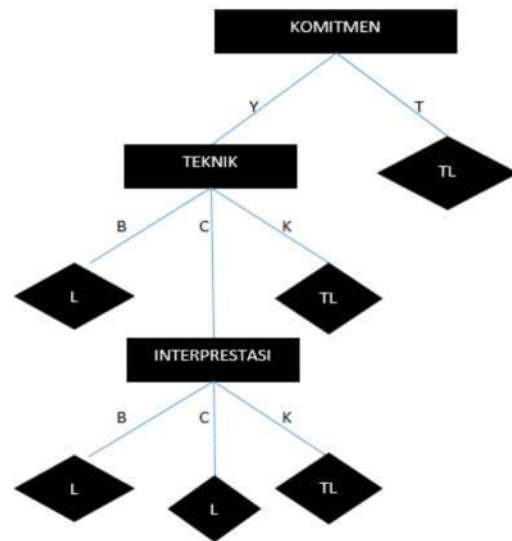
No	Nama	G	T	Ip	P	K	Ot	Hasil
11	Rudi	Lk	C	C	B	Y	S	L
19	Adi	Lk	C	B	B	Y	R	L
20	Eko	Lk	C	B	C	Y	R	L
27	Irma	P	C	B	B	Y	S	L
32	Galang	Lk	C	B	B	Y	S	L
33	Danny	Lk	C	B	C	Y	S	L
36	Prima	P	C	B	B	Y	T	L
41	Dita	P	C	B	C	Y	R	L
59	Nuri	P	C	K	C	Y	R	TL

Tabel 7. Pencarian Gain untuk Branch Node Komitmen=Y and Teknik=C

Atribut	Value Atribut	J.Kasus	L	TL	Error	Gain
Teknik	C	9	8	1	0,111	
Gender						
	Lk	5	5	0	0,000	0,000
	P	4	3	1	0,250	
Interprestasi						
	B	7	7	0	0,000	0,111
	C	1	1	0	0,000	
	K	1	0	1	0,000	
Penampilan						
	B	5	5	0	0,229	-5,55
	C	4	3	1	0,333	
	K	0	0	0	0,000	
Oktaf						
	R	4	3	1	0,250	-5,55
	S	4	4	0	0,000	
	T	1	1	0	0,000	

Dari tabel 7 diketahui nilai gain tertinggi didapat dari atribut Interprestasi=C, maka dapat diputuskan bahwa atribut Interprestasi=C terpilih sebagai branch node (cabang) dari batang komitmen=Y and Teknik=C.

Dari rangkaian proses perhitungan yang dilakukan, dapat dibentuk suatu pohon keputusan seperti yang terlihat pada gambar 2.



Gambar 2. Pohon Keputusan Pemodelan Decision Tree Untuk Seleksi Anggota Baru Tim Paduan Suara Harmoni Nusantara

Hasil akhir yang diperoleh dari proses pembentukan pohon keputusan ini adalah rule base yang dapat dibaca dari alur pohon keputusan mulai dari root sampai leaf yang merupakan label kelas dari kondisi yang disusun oleh atribut-atribut diatas leaf.

Rule base yang didapatkan adalah:

1. If Komitmen=Y and Teknik=K then Hasil=TL
2. If Komitmen=Y and Teknik=B then Hasil=L
3. If Komitmen=Y and Teknik=C and Interprestasi=B or Interprestasi=C then Hasil=L
4. If Komitmen=Y and Teknik=C and Interprestasi=K or Interprestasi=C then Hasil=TL
5. If Komitmen=T then Hasil=TL

Dari penelusuran *rule* base yang diperoleh dari pohon keputusan, dapat diketahui bahwa keseluruhan record data training (60 record) tercover oleh *rule* yang terbentuk, yaitu:

1. *If Komitmen=Y and Teknik=K then Hasil=TL*, mengcover 4 record, yaitu : 39,40,42,60.
2. *If Komitmen=Y and Teknik=B then Hasil=L*
Mengcover 25 record, yaitu no: yaitu recor no: 1, 3, 6, 7, 8, 9, 10, 12, 13, 16, 17, 18, 25, 26, 28, 30, 31, 37, 38, 44, 48, 49, 50, 53, 54.
3. *If Komitmen=Y and Teknik=C and Interpretasi=B or Interpretasi=C then Hasil=L*, mengcover 8 record, yaitu: 11, 19, 20, 17, 32, 33, 36, 41.
4. *If Komitmen=Y and Teknik=C and Interpretasi=K or Interpretasi=C then Hasil=TL*, mengcover satu record, no: 59.
5. *If Komitmen=T then Hasil=TL*, mengcover 22 record: 2, 4, 5, 14, 15, 21, 22, 23, 24, 29, 34, 35, 43, 45, 46, 47, 51, 52, 55, 56, 57, 58.

Dilihat dari pohon keputusan yang terbentuk, tidak semua atribut input dalam data training digunakan. Hal ini bisa terjadi karena hasil seleksi dari nilai gain selama proses perhitungan, sehingga atribut tercover oleh atribut yang lain. Selain itu atribut-atribut tersebut telah mempunyai kejelasan alur dan kejelasan hasil jika dilihat secara langsung pada data training, sehingga dapat membentuk *rule* base yang sudah jelas.

5. KESIMPULAN

Berdasarkan hasil pemodelan dan proses perhitungan *Decision Tree* dengan konsep *Classification Error* untuk algoritma bantu dalam seleksi anggota baru tim paduan suara Harmoni Nusantara, diperoleh kesimpulan sebagai berikut:

1. Pemodelan *Decision Tree* menggunakan konsep *Classification Error* dilakukan dengan 60 record data training dengan 6 atribut input (Gender, Teknik, Interpretasi, Penampilan, Komitmen, Oktaf) telah berhasil membentuk sebuah *Decision Tree* (pohon keputusan).
2. Dari pohon keputusan yang terbentuk diperoleh 5 *rule's base* yang mengcover keseluruhan record (60 record) data training yang berarti bahwa algoritma mempunyai tingkat akurasi 100% karena telah berhasil mengcover seluruh record data training dan memberikan hasil yang sangat baik berdasarkan kondisi value atribut.
3. Berdasarkan rangkaian proses dan hasil pemodelan yang telah dilakukan dapat disimpulkan bahwa *Decision Tree* dengan konsep *Classification Error* dapat digunakan sebagai algoritma bantu yang akan diimplementasikan dalam sistem bantu (aplikasi bantu) dalam seleksi anggota baru Tim Paduan Suara Harmoni Nusantara pada Universitas Nusantara PGRI Kediri yang direncanakan.

DAFTAR PUSTAKA

- [1] Niswatin, Ratih Kumalasri (2017). Penerapan Algoritma *Decision Tree* Pada Penentuan

Keberhasilan Akademik Mahasiswa, Semnas Teknomedia, AMIKON, Yogyakarta, 2017.

- [2] Pramudiono, I, 2006, Apa itu Data Mining? dalam <http://datamining.japati.net/cgi-bin/indodm.cgi?bacaarsip%115552761&arti%kel>, diakses tanggal Juli 2019.
- [3] Larose D, T., 2005, Discovering knowledge in data : an introduction to data mining, Jhon Wiley & Sons Inc.
- [4] Han, J dan Kamber, M., "Data Mining Concept and Technique", Morgan Kaufmann, 2001.
- [5] Zalilia, L. (2007). Penerapan data Mining untuk IDS. Tugas Akhir Mata Kuliah EC7010 Program Studi Teknik Elektro Insitut Teknologi Bandung.
- [6] Tom M. Mitchell, McGraw Hill, 1997, Decision Tree Learning, Lecture slides for textbook Machine Learning.

BIODATA PENULIS



Patmi Kasih, M.Kom

Lahir di Kota Nganjuk, tinggal di Kota Kediri. Menempuh pendidikan S2 di Sekolah Tinggi Teknik Surabaya, 2014. Dosen Tetap pada Universitas Nusantara PGRI Kediri pada Program Studi Teknik Informatika.