

Classification of Village Status in Landak Regency Using C5.0 Algorithm

Mutiara Cindy Nur Fitria^{1*}, Naomi Nessyana Debatara¹, Setyo Wira Rizki¹

¹ Program Studi Statistika FMIPA Universitas Tanjungpura, Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, Indonesia.

*Email: cindyfitriaa@student.untan.ac.id

Manuscript submitted : Maret 2022;

Accepted for publication : April 2022.

doi : <https://doi.org/10.30598/tensorvol3iss1pp33-42>

Abstract: Village development is an effort to improve the welfare and quality of life of rural communities. The development of a village, one of which can be measured by the Village Building Index (VBI). VBI is formed from three indices that are expected to cover all areas of village life. The lower the VBI value of a village, the more backward the village is. Classification of village status is very important for making policies that are in accordance with village conditions. This study used village status data based on the Village Building Index (VBI) of Landak Regency in 2020 obtained from the website of One Data West Kalimantan. The data used consists of Village Status variable (Y) which is the dependent variable and 12 independent variables, namely Health (X_1), Education (X_2), Social Capital (X_3), Settlement (X_4), Production Diversity (X_5), Trade (X_6), Distribution Access (X_7), Credit Access (X_8), Economic Institutions (X_9), Regional Openness (X_{10}), Environmental Quality (X_{11}), and Potential and Disaster Response (X_{12}). The purpose of this study is to classify the status of villages in Landak Regency using C5.0 Algorithm. Classification begins with data collection, then the data is divided into training and testing data in the 90:10 proportion. Next is the formation of a classification model using training data, after that, testing the classification model using data testing. Then the evaluation of the classification model and based on the results of the study obtained an accuracy value of 82.35%, which means the quality of the model is good and can be used, with the variables Health (X_1), Education (X_2), Settlement (X_4), and Credit Access (X_8), not too influential and the variable Potential and Disaster Response (X_{12}) the most influential in the classifying village status in Landak Regency.

2010 Mathematical Subject Classification: 97M10.

Keywords: Village development, VBI, C5.0 Algorithm

1. Pendahuluan

Indeks Desa Membangun (IDM) adalah suatu konsep sarana dalam meningkatkan pembangunan dan pengembangan potensi desa yang diharapkan bisa mencakup semua bidang kehidupan desa agar karakteristik yang dimiliki suatu desa dapat dilakukan pembangunan yang berkelanjutan [1]. Perangkat-perangkat indikator yang dikembangkan dalam IDM didasarkan pada persepsi bahwa untuk menuju desa yang maju dan mandiri diperlukan kerangka kerja pembangunan yang berkelanjutan. Adapun aspek ekonomi, sosial, dan lingkungan menjadi kekuatan yang saling mengisi dan menjaga potensi serta kemampuan desa untuk mensejahterakan

kehidupan desa [2].

Berdasarkan Indeks Desa Membangun, klasifikasi status desa dibagi menjadi 5 yaitu desa sangat tertinggal, tertinggal, berkembang, maju, dan mandiri. Pengklasifikasian desa bertujuan untuk mengindikasikan keragaman karakteristik setiap desa yang juga dapat digunakan dalam penajaman penentuan status perkembangan desa. Selain itu, sebagai rekomendasi interferensi kebijakan karena desa dengan status sebagai desa tertinggal dan sangat tertinggal memerlukan pendekatan dan intervensi kebijakan yang berbeda [3].

Berdasarkan data dari Kemendes PDTT pada tahun 2020 Kabupaten Landak menempati posisi ke 358 dari total 434 kabupaten se-Indonesia sebagai kabupaten tertinggal dengan nilai IDM sebesar 0,5802. Kabupaten Landak merupakan kabupaten dengan nilai IDM terendah dibandingkan kabupaten lain di Kalimantan Barat serta menjadi kabupaten dengan persentase tertinggi desa berstatus sangat tertinggal yaitu sebesar 4,49% atau sebanyak 7 desa dari total 156 desa. Besarnya persentase desa sangat tertinggal di Kabupaten Landak menjadi penyebab rendahnya nilai IDM Kabupaten Landak. Untuk meningkatkan IDM Kabupaten Landak agar tidak menimbulkan kesenjangan antar desa maupun kabupaten, diperlukan analisis untuk memprediksi dimensi utama yang berperan dalam pengklasifikasian status desa di Kabupaten Landak.

Satu diantara analisis yang dapat digunakan untuk memprediksi indikator utama dalam pengklasifikasian status desa berdasarkan IDM di Kabupaten Landak adalah Algoritma C5.0. Algoritma C5.0 merupakan salah satu algoritma klasifikasi yang mengklasifikasikan datanya menjadi pohon bercabang dengan beberapa *node*. Penggunaan Algoritma C5.0 telah banyak digunakan pada penelitian terdahulu untuk mengklasifikasikan dan memprediksi suatu studi kasus. Akan tetapi, penggunaan Algoritma C5.0 untuk memprediksi indikator utama dalam pengklasifikasian status desa berdasarkan IDM belum pernah digunakan. Algoritma C5.0 adalah penyempurna dari metode sebelumnya yang sering digunakan yaitu Algoritma C4.5 dan Algoritma *Iterative Dichotomiser 3* (ID3) yang dibentuk pada tahun 1987 oleh Ross Quinlan [4].

Oleh sebab itu, dalam penelitian ini digunakan Algoritma C5.0 yang diharapkan dapat mengklasifikasikan status desa di Kabupaten Landak untuk memprediksi dimensi utama yang menjadi penyebab rendahnya IDM Kabupaten Landak pada tahun 2020. Sehingga pendekatan dan intervensi kebijakan yang diberikan akan sesuai dengan kondisi dan keadaan desa tersebut agar tidak terjadinya kesenjangan antar desa.

1.1. Indeks Desa Membangun (IDM)

Desa merupakan unit administrasi pemerintahan terkecil yang mempunyai kontribusi penting dalam pembangunan sendiri desa. Indeks Desa Membangun atau yang disingkat dengan IDM dicetuskan guna memperkuat upaya pencapaian tujuan pembangunan suatu desa serta kawasan perdesaan, tujuan pembangunan tersebut membutuhkan kepastian posisi desa atau status perkembangan desa. Status perkembangan desa dapat diukur melalui IDM dengan tujuan untuk melihat dampak berkelanjutan bagi kemajuan dan kesejahteraan desa tersebut. Indeks Desa Membangun (IDM) mengklasifikasikan status desa menjadi lima, yaitu desa mandiri, maju, berkembang, tertinggal, dan sangat tertinggal [5].

IDM diperoleh dari rata-rata Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE) dan Indeks Ketahanan Lingkungan (IKL). Adapun rumus untuk mencari IDM adalah sebagai berikut:

$$IDM = \frac{IKS+IKE+IKL}{3} \quad (1)$$

dengan:

- IDM* : Indeks Desa Membangun
- IKS* : Indeks Ketahanan Sosial
- IKE* : Indeks Ketahanan Ekonomi
- IKL* : Indeks Ketahanan Lingkungan/Ekologi

1.2. Dimensi Pembentuk IDM

Indeks Desa Membangun merupakan gabungan dari 3 indeks, yaitu indeks ketahanan sosial, ekonomi dan lingkungan. Ketiga indeks tersebut terdiri dari 12 dimensi yang kemudian diturunkan menjadi 50 indikator. Semua indikator mempunyai skor 0–5 dengan indikasi semakin besar skor maka semakin besar pula tingkat keberartiannya. Penetapan skor tersebut berdasarkan hasil FGD *Analytical Hierarchy Process* (AHP) [2]. Skor indikator yang dikelompokkan ke dalam dimensi menghasilkan skor dimensi. Adapun rumus yang digunakan dalam mencari skor dimensi adalah sebagai berikut:

$$D = \frac{\sum_{i=1}^n I_i}{n \times 5} \quad (2)$$

dengan:

D : Nilai dimensi

I_i : Nilai indikator ke- i

n : Banyaknya indikator

Adapun 12 dimensi pembentuk IDM yang digunakan pada penelitian ini, yaitu:

- Dimensi Kesehatan (X_1), Pendidikan (X_2), Modal Sosial (X_3), dan Permukiman (X_4) merupakan dimensi yang membentuk Indeks Ketahanan Sosial (IKS).
- Dimensi Keragaman Produksi (X_5), Perdagangan (X_6), Akses Distribusi (X_7), Akses Kredit (X_8), Lembaga Ekonomi (X_9), dan Keterbukaan Wilayah (X_{10}) merupakan dimensi yang membentuk Indeks Ketahanan Ekonomi (IKE).
- Dimensi Kualitas Lingkungan (X_{11}) serta Potensi dan Tanggap Bencana (X_{12}) merupakan dimensi yang membentuk Indeks Ketahanan Lingkungan (IKL).

1.3. Klasifikasi

Klasifikasi adalah suatu teknik yang dapat mengklasifikasikan data baru dengan cara melihat atribut atau kelakuan dari kelompok data yang telah didefinisikan atau dengan menggunakan sejumlah aturan yang telah diperoleh dari hasil klasifikasi [6]. Proses pengklasifikasian data dibagi menjadi 2 tahapan, yaitu *learning* dan *classification*. Pada tahapan *learning*, data *training* dianalisis dengan mengaplikasikan sebuah algoritma klasifikasi. Pada tahapan *classification*, *classification rules* yang telah diperoleh diestimasi ketepatannya menggunakan data *testing*. Jika keakuratan *classification rules* dapat diterima, maka model klasifikasi tersebut dapat diaplikasikan pada data baru [7].

2. Metode Penelitian

Pada penelitian ini digunakan data sekunder, yaitu data Status Desa di Kabupaten Landak berdasarkan Indeks Desa Membangun (IDM) tahun 2020 yang diperoleh dari website Satu Data Kalbar. Data yang digunakan dalam penelitian ini sebanyak 156 data dengan 12 variabel independen yaitu Kesehatan (X_1), Pendidikan (X_2), Modal Sosial (X_3), Permukiman (X_4), Keragaman Produksi (X_5), Perdagangan (X_6), Akses Distribusi (X_7), Akses Kredit (X_8), Lembaga Ekonomi (X_9), Keterbukaan Wilayah (X_{10}), Kualitas Lingkungan (X_{11}), serta Potensi dan Tanggap Bencana (X_{12}) yang merupakan dimensi pembentuk IDM. Adapun variabel dependen yang digunakan yaitu Status Desa. Adapun teknik analisis data yang digunakan adalah teknik analisis *data mining* dengan metode Algoritma C5.0. Proses analisis dilakukan dengan bantuan *software* R Studio.

Tabel 1 Data Status Desa di Kabupaten Landak berdasarkan Indeks Desa Membangun (IDM) tahun 2020

DESA	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	STATUS DESA
SABAKA	0,45 7	0,62 9	0,72 3	0,70 0	0,20 0	0,46 7	0,00 0	0,20 0	0,10 0	0,53 3	0,80 0	0,30 0	SANGAT TERTINGGAL
ANSOLOK	0,51 4	0,54 3	0,78 5	0,67 5	0,20 0	0,46 7	0,00 0	0,20 0	0,30 0	0,60 0	0,40 0	0,30 0	SANGAT TERTINGGAL
MORO BETUNG	0,51 4	0,60 0	0,61 5	0,55 0	0,20 0	0,33 3	0,00 0	0,20 0	0,60 0	0,33 3	0,40 0	0,50 0	SANGAT TERTINGGAL
AMPADI	0,57 1	0,42 9	0,80 0	0,40 0	0,20 0	0,46 7	0,00 0	0,10 0	0,30 0	0,53 3	0,80 0	0,30 0	SANGAT TERTINGGAL
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
KARANGAN	0,71 4	0,77 1	0,83 1	0,97 5	0,20 0	0,86 7	1,00 0	0,80 0	1,00 0	0,86 7	1,00 0	0,80 0	MANDIRI
MENJALIN	0,62 9	1,00 0	0,95 4	0,80 0	1,00 0	1,00 0	0,60 0	0,50 0	1,00 0	1,00 0	0,80 0	0,80 0	MANDIRI
DARIT	0,77 1	0,54 3	0,81 5	0,87 5	0,20 0	1,00 0	1,00 0	0,70 0	1,00 0	1,00 0	0,80 0	0,90 0	MANDIRI

Proses analisis yang dilakukan adalah sebagai berikut:

a. Pembagian data menjadi data *training* dan data *testing*

Dalam membuat model klasifikasi, data yang digunakan dibagi menjadi data *training* dan *testing*. Data *training* adalah data yang digunakan untuk membuat model klasifikasi, model mengevaluasi data berulang kali untuk mempelajari lebih lanjut tentang perilaku data dan kemudian menyesuaikan diri untuk memenuhi tujuan yang dimaksudkan. Sedangkan data *testing* adalah data yang digunakan untuk menguji model klasifikasi yang telah diperoleh.

Metode proporsi merupakan metode yang paling sederhana dalam pembagian data *training* dan data *testing*. Pembagian proporsi data *training* dan *testing* biasanya bersifat diskrit. Adapun rumus yang digunakan untuk membagi data *training* dan data *testing* yaitu:

$$\text{Jumlah data training} = \text{Proporsi data training} \times N \tag{3}$$

$$\text{Jumlah data testing} = N - \text{Jumlah data training} \tag{4}$$

dengan:

: Banyaknya data penelitian

b. Pemodelan klasifikasi dengan mengimplementasikan Algoritma C5.0 menggunakan data *training*

Proses pembentukan *decision tree* pada Algoritma C5.0 serupa dengan pembentukan *decision tree* pada algoritma C4.5. Hal yang serupa diantara keduanya ialah pada perhitungan nilai *entropy* dan nilai *gain*. Pada algoritma C4.5 perhitungan berakhir pada nilai *gain*, sedangkan pada algoritma C5.0 perhitungan berlanjut dengan menghitung nilai *gain ratio* menggunakan nilai *entropy* dan *gain* yang telah diperoleh sebelumnya. Rumus yang digunakan untuk mencari nilai *entropy* sebagai berikut:

$$\text{Entropy}(S) = -\sum_{i=1}^k p_i \log_2 p_i \tag{5}$$

dengan:

- S : Himpunan kasus
- k : Banyaknya kelas
- p_i : Proporsi dari S_i dan S

Kemudian rumus yang digunakan untuk mencari nilai *gain* yaitu:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^j \frac{|S_i|}{|S|} \times Entropy(S_i) \tag{6}$$

dengan:

- $Gain(S, A)$: Nilai *gain* dari atribut suatu variabel
- A : Atribut suatu variabel
- j : Jumlah partisi himpunan atribut A
- S_i : Himpunan kasus pada partisi ke- i
- $Entropy(S_i)$: Nilai *entropy* pada partisi ke- i

Setelah memperoleh nilai *entropy* dan *gain*, langkah selanjutnya adalah menghitung nilai *gain ratio*. Adapun rumus dasar yang digunakan untuk mencari nilai *gain ratio* yaitu:

$$Gain Ratio = \frac{Gain(S,A)}{\sum_{i=1}^j Entropy(S_i)} \tag{7}$$

Nilai *gain ratio* yang tertinggi akan dipilih menjadi *node*. Proses perhitungan nilai *entropy*, *gain*, dan *gain ratio* dihitung kembali untuk setiap cabangnya hingga semua cabang mendapatkan kelasnya masing-masing.

c. Pengujian model klasifikasi menggunakan data *testing* untuk memperoleh tingkat akurasi

Confusion matrix merupakan salah satu alat bantu yang digunakan untuk menilai seberapa baik model klasifikasi. Format umum yang biasa digunakan dalam tabel *confusion matrix* adalah tabel biner, yaitu tabel yang terdiri dari dua kelas label saja. Akan tetapi ada beberapa kasus yang memiliki lebih dari dua kelas label, yaitu *multiclass*. Berikut adalah format *confusion matrix* untuk *multiclass*:

Tabel 2 Confusion Matrix

		<i>Predicted Class</i>				
		C1	C2	C3	...	Cm
<i>Actual Class</i>	C1	X ₁₁	X ₁₂	X ₁₃	...	X _{1m}
	C2	X ₂₁	X ₂₂	X ₂₃	...	X _{2m}
	C3	X ₃₁	X ₃₂	X ₃₃	...	X _{3m}
	⋮	⋮	⋮	⋮	...	⋮
	Cm	X _{m1}	X _{m2}	X _{m3}	...	X _{mm}

Dari Tabel 2 dapat dicari ketepatan model klasifikasi dengan cara mencari nilai akurasinya. Semakin tinggi akurasinya maka akan semakin baik pula model klasifikasinya. Adapun rumus yang digunakan yaitu sebagai berikut [8]:

$$\text{Nilai Akurasi } (a) = \frac{\sum_{i=1}^m x_{ii}}{\text{Jumlah data testing}} \times 100\% \tag{8}$$

d. Evaluasi model klasifikasi

Setelah diperoleh nilai akurasi menggunakan bantuan *confusion matrix*, maka dilakukan evaluasi model klasifikasi. Nilai minimum akurasi yang disarankan oleh Thunnissen dan Noordman adalah 0,7 atau 70% [9]. Hal tersebut sejalan dengan tabel klasifikasi kualitas model yang sering

digunakan untuk menguji nilai akurasi atau ketepatan suatu model, yaitu sebagai berikut [10]:

Tabel 3 Klasifikasi Kualitas Model

Akurasi	Kualitas Model
$0,9 \leq a < 1$	Sangat Baik
$0,8 \leq a < 0,9$	Baik
$0,7 \leq a < 0,8$	Cukup
$0,6 \leq a < 0,7$	Buruk
$a < 0,6$	Sangat Buruk

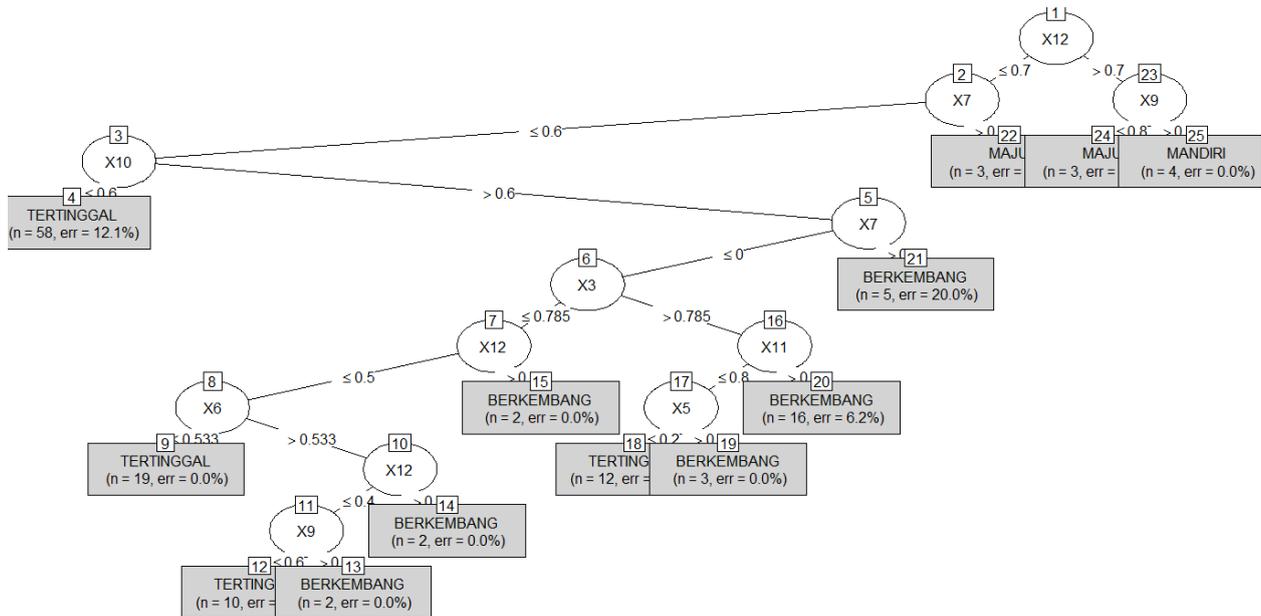
3. Hasil dan Pembahasan

3.1. Pembagian Data *Training* dan *Testing*

Pembagian data *training* dan *testing* dilakukan menggunakan bantuan *software* R Studio dengan proporsi 90:10. Diperoleh data yang termasuk ke dalam data *training* yaitu sebanyak 139 data sedangkan sisanya yaitu sebanyak 17 data termasuk kedalam data *testing*.

3.2. Pemodelan Klasifikasi

Pembentukan pohon klasifikasi Algoritma C5.0 dalam penelitian ini dilakukan menggunakan bantuan *software* R Studio. Adapun Gambar 1 merupakan hasil struktur model klasifikasi status desa di Kabupaten Landak menggunakan Algoritma C5.0.



Gambar 1 Pohon Klasifikasi Algoritma C5.0

Dari 12 variabel independen yang digunakan dalam penelitian ini, variabel Kesehatan (X_1), Pendidikan (X_2), Permukiman (X_4), dan Akses Kredit (X_8) tidak terlalu mempengaruhi hasil prediksi dalam mengklasifikasikan status desa di Kabupaten Landak. Sedangkan, variabel Potensi dan Tanggap Bencana (X_{12}) menjadi variabel yang memiliki pengaruh terbesar. Karena variabel Potensi dan Tanggap Bencana (X_{12}) menjadi *root node* yang artinya alasan paling dominan Kabupaten Landak memiliki nilai IDM terendah dari Kabupaten lainnya di Kalimantan Barat adalah besarnya potensi bencana di Kabupaten Landak dan kurangnya fasilitas mitigasi atau tanggap

bencana.

Berdasarkan Gambar 1 juga dapat disimpulkan bahwa:

- a. Rules 1 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} \leq 0,6)) \Rightarrow \text{Node 4}$
- b. Rules 2 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 \leq 0,785) \wedge (X_{12} \leq 0,5) \wedge (X_6 \leq 0,533)) \Rightarrow \text{Node 9}$
- c. Rules 3 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 \leq 0,785) \wedge (X_{12} \leq 0,5) \wedge (X_6 > 0,533) \wedge (X_{12} \leq 0,4) \wedge (X_9 \leq 0,6)) \Rightarrow \text{Node 12}$
- d. Rules 4 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 \leq 0,785) \wedge (X_{12} \leq 0,5) \wedge (X_6 > 0,533) \wedge (X_{12} \leq 0,4) \wedge (X_9 > 0,6)) \Rightarrow \text{Node 13}$
- e. Rules 5 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 \leq 0,785) \wedge (X_{12} \leq 0,5) \wedge (X_6 > 0,533) \wedge (X_{12} > 0,4)) \Rightarrow \text{Node 14}$
- f. Rules 6 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 \leq 0,785) \wedge (X_{12} > 0,5)) \Rightarrow \text{Node 15}$
- g. Rules 7 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 > 0,785) \wedge (X_{11} \leq 0,8) \wedge (X_5 \leq 0,2)) \Rightarrow \text{Node 18}$
- h. Rules 8 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 > 0,785) \wedge (X_{11} \leq 0,8) \wedge (X_5 > 0,2)) \Rightarrow \text{Node 19}$
- i. Rules 9 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 \leq 0) \wedge (X_3 > 0,785) \wedge (X_{11} > 0,8)) \Rightarrow \text{Node 20}$
- j. Rules 10 = $(IF(X_{12} \leq 0,7) \wedge (X_7 \leq 0,6) \wedge (X_{10} > 0,6) \wedge (X_7 > 0)) \Rightarrow \text{Node 21}$
- k. Rules 11 = $(IF(X_{12} \leq 0,7) \wedge (X_7 > 0,6)) \Rightarrow \text{Node 22}$
- l. Rules 12 = $(IF(X_{12} > 0,7) \wedge (X_9 \leq 0,8)) \Rightarrow \text{Node 24}$
- m. Rules 13 = $(IF(X_{12} > 0,7) \wedge (X_9 > 0,8)) \Rightarrow \text{Node 25}$

3.3. Pengujian Model Klasifikasi

Setelah diperoleh model pohon klasifikasi Algoritma C5.0, maka dilakukan pengujian model klasifikasi menggunakan bantuan *confusion matrix*. Pengujian model klasifikasi dilakukan menggunakan data *testing* sebanyak 17 data. Berikut adalah hasil pengujian model klasifikasi menggunakan bantuan tabel *confusion matrix*:

Tabel 4 *Confusion Matrix* Algoritma C5.0

		<i>Predicted Class</i>				
		Sangat Tertinggal	Tertinggal	Berkembang	Maju	Mandiri
<i>Actual Class</i>	Sangat Tertinggal	0	1	0	0	0
	Tertinggal	0	9	1	0	0
	Berkembang	0	0	3	1	0
	Maju	0	0	0	1	0
	Mandiri	0	0	0	0	1

Dari Tabel 4 dapat diketahui bahwa prediksi desa berstatus tertinggal yang sesuai dengan fakta sebanyak 9 desa, prediksi desa berstatus tertinggal yang salah terklasifikasi sebanyak 1 desa.

Prediksi desa berstatus berkembang yang sesuai dengan fakta sebanyak 3 desa, prediksi desa berstatus berkembang yang salah terklasifikasi sebanyak 1 desa. Prediksi desa berstatus maju yang sesuai dengan fakta sebanyak 1 desa, prediksi desa berstatus maju yang salah terklasifikasi sebanyak 1 desa. Prediksi desa berstatus mandiri yang sesuai dengan fakta sebanyak 1 desa.

3.4. Evaluasi Model

Setelah diperoleh hasil pengujian model menggunakan bantuan tabel *confusion matrix*, maka dilakukan evaluasi model dengan cara menghitung tingkat ketepatan akurasi menggunakan Persamaan 8 yaitu, sebagai berikut:

$$\begin{aligned} \text{Akurasi} &= \frac{0 + 9 + 3 + 1 + 1}{17} \times 100\% \\ &= \frac{14}{17} \times 100\% = 82,35\% \end{aligned}$$

Berdasarkan hasil perhitungan akurasi menggunakan bantuan *confusion matrix* diperoleh nilai akurasi sebesar 82,35%. Berdasarkan Tabel 3, nilai tersebut menunjukkan kualitas model baik karena $\geq 80\%$ sehingga model dapat digunakan untuk mengklasifikasikan status desa.

4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada skripsi ini, diperoleh kesimpulan sebagai berikut:

1. Implementasi Algoritma C5.0 dalam pengklasifikasian status desa di Kabupaten Landak menghasilkan nilai akurasi dengan kualitas model yang baik sehingga model dapat digunakan untuk mengklasifikasikan status desa.
2. Dari hasil analisis pengklasifikasian status desa di Kabupaten Landak yang telah dilakukan menggunakan Algoritma C5.0 dengan proporsi 90:10 diperoleh nilai akurasi sebesar 82,35% dengan variabel Kesehatan (X_1), Pendidikan (X_2), Permukiman (X_4), dan Akses Kredit (X_8) yang tidak terlalu berpengaruh serta variabel Potensi dan Tanggap Bencana (X_{12}) yang paling berpengaruh atau memiliki pengaruh paling besar dalam pengklasifikasian status desa di Kabupaten Landak.

Referensi

- [1] Sari, N., dan Oktavianor, T., (2020), Indeks Desa Membangun (IDM) di Kabupaten Barito Kuala, *Jurnal Administrasi Publik dan Pembangunan*, 2(1), 35-41.
- [2] Kementerian Desa, (2020), Pembangunan Daerah Tertinggal, dan Transmigrasi, 2020, *Standar Operasional Prosedur Pemutakhiran Status Perkembangan Desa Indeks Desa Membangun (IDM) Tahun 2020*, Available at: <https://idm.kemendes.go.id>.
- [3] Xavierus, F., (2021), Perkembangan Desa Berdasarkan Indeks Desa Membangu Di Kabupaten Mempawah, *Jurnal Borneo Akcaya*, 7(1), pp. 10–20.
- [4] Dewi, D.A.W., Cholissodin, I., dan Sutrisno, (2019), Klasifikasi Penyimpangan Tumbuh Kembang Anak Menggunakan Algoritma C5.0, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(10), 10258-10265.
- [5] Ramly, A.R., Wahyuddin, Mursyida, J., dan Mawardati, (2019), Analysis of Building Village Index (IDM) in Village Development in Kuala Sub-District, *Jurnal Ekonomi Pembangunan*, 17(01), 22-31.
- [6] Nofriansyah, D., dan Nurcahyo, G.W., (2019), *Algoritma Data Mining dan Pengujiannya*, Deepublish Publisher, Yogyakarta.
- [7] Vulandari, R.T., (2017), *Data Mining Teori Aplikasi dan Rapidminer*, Penerbit Gava Media, Yogyakarta.
- [8] Manliguez, C., (2016), *Generalized Confusion Matrix for Multiple Classes*, Available

at: <https://www.researchgate.net/publication/310799885>.

- [9] Kiptala, J.K., (2016), *Managing Basin Interdependencies in a Heterogeneous, Highly Utilized and Data Scarce River Basin in Semi-Arid Africa (the Case of the Pangani River Basin, Eastern Africa)*, CRC Press/Balkema, Leiden.
- [10] Moocarme, M., Abdolahnejad, M., dan Bhagwat, R., (2020), *The Deep Learning with Keras Workshop: learn how to define and train neural network models with just a few lines of code*, Packt Publishing Ltd., Birmingham.

