



Klasifikasi Desa Wisata di Provinsi Jawa Tengah Menggunakan Algoritma C4.5

Nimas Leniawati¹, Sena Wijayanto²

^{1,2}Institut Teknologi Telkom Purwokerto, Indonesia

e-mail: ¹19103100@ittelkom-pwt.ac.id, ²sena@ittelkom-pwt.ac.id

Abstract

Central Java is part of a province on the island of Java that develops the tourism industry. Based on Central Java provincial regulation No. 2 of 2019, regarding the empowerment of tourism villages in Central Java Province, it is stated that tourism villages play an important role in advancing public welfare, equitable distribution of trade and employment, optimization of economic potential and regional characteristics, as well as promoting and protecting cultural, religious, and national values and natural preservation. This relates to the importance of identifying tourist villages that can be developed as tourist destinations. For this reason, classification in data mining can be done using the C4.5 algorithm. The C4.5 algorithm can be used in obtaining a decision tree that is easy to understand in classifying tourist villages. This research aims to classify tourist villages by implementing the decision tree classifier C4.5 algorithm. The research data was obtained from the Central Java Province data portal dataset in 2019 and the problem solving in this study raised the Knowledge Discovery Database process. The results of this study are in the form of testing the C4.5 algorithm using confusion matrix and obtaining accuracy of 86%, recall of 88% and precision of 86%.

Keywords: Data Mining, Decision Tree, Classification, Algoritma C4.5, Phyton

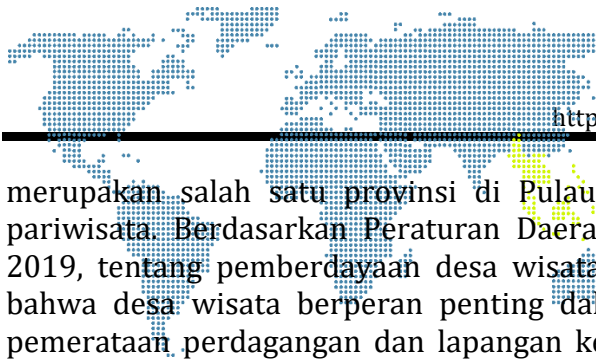
Abstrak

Jawa Tengah merupakan bagian provinsi di Pulau Jawa yang mengembangkan industri pariwisata. Berdasarkan peraturan daerah provinsi Jawa Tengah No. 2 tahun 2019, mengenai pemberdayaan desa wisata di Provinsi Jawa Tengah, disebutkan bahwa desa wisata berperan penting dalam memajukan kesejahteraan umum, pemerataan perdagangan dan ligan kerja optimalisasi potensi ekonomi dan ciri khas daerah, serta memajukan dan melindungi nilai-nilai budaya, agama, dan kebangsaan serta kelestarian alam. Hal ini berkaitan dengan pentingnya mengidentifikasi desa-desa wisata yang dapat dikembangkan sebagai tujuan wisata. Untuk itu, klasifikasi dalam data mining dapat dilakukan dengan menggunakan algoritma C4.5. Algoritma C4.5 dapat digunakan dalam mendapatkan sebuah pohon keputusan yang mudah dipahami dalam mengklasifikasikan desa wisata. Penelitian ini bertujuan untuk mengklasifikasikan desa wisata dengan mengimplementasikan decision tree classifier algoritma C4.5. Data penelitian diperoleh dari dataset portal data Provinsi Jawa Tengah tahun 2019 dan penyelesaian permasalahan pada penelitian ini mengangkat proses Knowledge Discovery Database. Hasil penelitian ini berupa pengujian algoritma C4.5 dengan menggunakan confusion matrix dan memperoleh accuracy sebesar 86%, recall sebesar 88% dan precision sebesar 86%.

Kata kunci: Data Mining, Pohon Keputusan, Klasifikasi, Algoritma C4.5, Phyton

1. PENDAHULUAN

Pariwisata merupakan bagian penting dari setiap daerah pada masing-masing provinsi. Selain sebagai tempat hiburan, pariwisata juga dianggap sebagai industri yang berperan penting dalam perekonomian dunia. Salah satu provinsi dengan diversifikasi pariwisata yaitu provinsi Jawa Tengah yang terletak di wilayah strategis karena letaknya di sebuah benua di pulau Jawa yang terbagi 2 (dua) provinsi utama yaitu Jawa Barat dan Jawa Timur yang merupakan Daerah Istimewa Yogyakarta yang terbatas dan Laut Jawa di utara. Jawa Tengah



merupakan salah satu provinsi di Pulau Jawa yang mengembangkan industri pariwisata. Berdasarkan Peraturan Daerah Provinsi Jawa Tengah No. 2 Tahun 2019, tentang pemberdayaan desa wisata di Provinsi Jawa Tengah, disebutkan bahwa desa wisata berperan penting dalam memajukan kesejahteraan umum, pemerataan perdagangan dan lapangan kerja, optimalisasi potensi ekonomi dan ciri khas daerah, serta memajukan dan melindungi nilai-nilai budaya, agama, dan kebangsaan serta kelestarian alam[1].

Berdasarkan data Disporapar (Dinas Kepemudaan, Olahraga dan Pariwisata) Provinsi Jawa Tengah tahun 2019, terdapat sejumlah 35 kabupaten, 291 nama desa wisata serta 409 objek wisata beserta kekayaan alam, budaya dan kearifan lokal di desa masing-masing. Data tersebut memiliki banyak destinasi desa wisata dari berbagai daerah/provinsi. Sehingga perlu dilakukan klasifikasi terhadap destinasi desa wisata sesuai dengan kriteria-kriteria dalam penentuan kategori desa wisata[2]. Klasifikasi desa wisata adalah proses mengkategorikan desa-desa ke dalam beberapa kelompok berdasarkan kriteria tertentu seperti fasilitas, potensi wisata, dan tingkat perkembangannya. Metode algoritma C4.5 merupakan decision tree classifier yang banyak digunakan untuk mengklasifikasikan data. Algoritma ini membuat model berdasarkan aturan-aturan yang dipilih dari data training dan diterapkan pada data baru untuk melakukan klasifikasi. Penggunaan algoritma C4.5 dalam klasifikasi desa wisata dapat membantu dalam menentukan desa mana yang memiliki potensi wisata terbaik dan memerlukan pengembangan lebih lanjut. Hal ini berkaitan dengan pentingnya mengidentifikasi desa-desa wisata yang dapat dikembangkan sebagai tujuan wisata.

Untuk itu, klasifikasi desa wisata membutuhkan analisis data yang tepat dan akurat untuk membantu menentukan desa mana yang memiliki potensi wisata tertinggi. Untuk mengurangi resiko dalam klasifikasi tersebut, dibutuhkan metode yang dapat memudahkan peringkasan data mentah menjadi data yang mudah dikelola dengan mengkategorikan masing-masing data yang dipilih, sehingga data tersebut sesuai dengan proses penerapan data mining yang telah dilakukan dalam klasifikasi menggunakan decision tree classifier algoritma C4.5. Penelitian ini juga merujuk pada penelitian sejenis yang mengatakan bahwa algoritma C4.5 adalah algoritma yang mudah diterapkan dalam melakukan suatu pengelompokan dan menghasilkan akurasi prediksi terbaik. Maka, hasil dari penelitian ini diharapkan mampu meningkatkan efisiensi serta meminimalisir kesalahan dalam mengklasifikasi kategori desa wisata untuk perkembangan destinasi wisata di Provinsi Jawa Tengah.

2. METODOLOGI PENELITIAN

2.1. Klasifikasi

Klasifikasi merupakan suatu proses untuk menemukan beberapa model atau fungsi yang menghasilkan gambaran *class* atau konsep dari suatu data. Konsep dasar dari klasifikasi adalah beberapa data yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula [1].

2.2. Desa Wisata

Desa Wisata adalah sebuah konsep pengembangan pariwisata yang berfokus pada pengembangan potensi dan kearifan lokal masyarakat desa. Desa wisata memanfaatkan kekayaan alam, budaya, dan kearifan lokal masyarakat sebagai daya tarik utama untuk menarik wisatawan. Potensi desa wisata meliputi kekayaan alam seperti pantai, danau, pegunungan, hutan, dan flora dan fauna yang kaya, serta kearifan lokal masyarakat seperti budaya, tradisi, dan kuliner yang unik. Oleh karena itu, desa wisata mempunyai potensi besar sebagai salah satu destinasi wisata yang unik dan menarik untuk berkelanjutan dalam pengembangan destinasi wisata yang menjadi sebuah “*trend*” dalam dasawarsa terakhir ini[2].

2.3. Knowledge Discovery in Database (KDD)

Penambangan data atau data mining merupakan suatu metode pengolahan data dalam pencarian informasi yang sebelumnya tidak diketahui dari kumpulan data besar. Secara umum data mining juga disebut sebagai sebuah cara atau tehnik dalam menyusun data besar dalam melakukan pencarian sebuah informasi yang bermanfaat bagi pengguna [3]. Adapun beberapa tahapan dari pengolahan data mining yaitu sebagai berikut [4]:

2.2.1. Selection

Selection atau pemilihan, mempunyai tujuan utama yaitu membuat data target dari data asli dengan memilih subset variabel atau sampel data.

2.2.2. PreProcessing

Preprocessing memiliki tujuan utama yaitu membersihkan data dengan melakukan berbagai operasi seperti pemodelan dan penghapusan *noise*, menentukan strategi yang tepat untuk menangani bidang data yang hilang, menghitung informasi data yang dibutuhkan.

2.2.3. Transformation

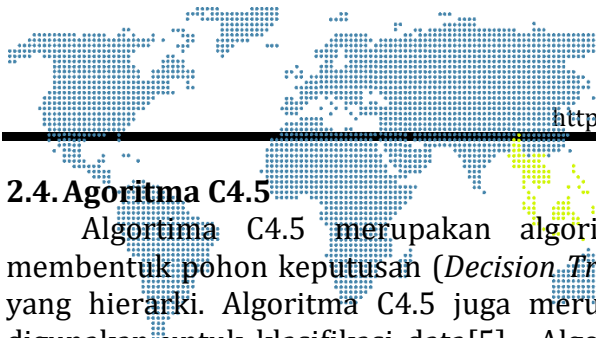
Transformation memiliki tugas yaitu memetakan seluruh rangkaian nilai atribut yang diberikan ke rangkaian nilai pengganti yang baru. Setiap nilai lama dapat diidentifikasi dengan salah satu nilai baru. Data *transformation* memiliki beberapa metode yaitu *smoothing*, *attribute/feature construction*, *aggregation*, *normalization* dan *discretization*.

2.2.4. Data Mining

Tahapan ini berhubungan dengan penggalian pola yang unik dengan memilih metode atau data mining tertentu, (misalnya *summarization*, *classification*, *clustering*, *regression*, dan sebagainya), algoritma yang tepat, dan representasi yang sesuai dari hasil keluaran.

2.2.5. Evaluation

Evaluation atau *Interpretation* digunakan untuk menafsirkan dan mengekstrak pengetahuan dari pola yang dihasilkan dengan memvisualisasikan pola. Interpretasi ini biasanya dilakukan dengan memvisualisasikan pola, model, atau data yang diberikan model tersebut.



2.4. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang sering digunakan untuk membentuk pohon keputusan (*Decision Tree*) untuk memprediksi model struktur yang hierarki. Algoritma C4.5 juga merupakan salah satu metode yang biasa digunakan untuk klasifikasi data[5]. Algoritma C4.5 merupakan pengembangan dari ID3 yaitu *Iterative Dichotomiser 3*. ID3 dikembangkan oleh J. Ross Quinlan. Prosedur dari algoritma ID3 mengambil input berupa contoh pelatihan, label pelatihan dan atribut lainnya. Beberapa proyek pengembangan yang diselesaikan di C4.5 mencakup passing nilai yang hilang, penerusan data berkelanjutan, dan pemangkasan [6].

Algoritma ini dapat menyelesaikan masalah dengan membangun pohon keputusan secara sistematis dengan tahapan-tahapan berikut [7]:

- a) Mempersiapkan data latih (*data training*), yaitu bisa dengan mengambil dari data histori yang sudah terjadi sebelumnya dan telah diklasifikasikan ke kelas-kelas tertentu.
- b) Menghitung akar dari pohon, yaitu dengan menghitung root (akar) yang akan diambil dari atribut yang terpilih. Sebelum menghitung nilai *gain* dari masing-masing atribut yang dipilih, langkah yang dilakukan dahulu yaitu menghitung nilai *entropy*. Hitung nilai *entropy* menggunakan persamaan rumus berikut:

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

Keterangan:

S = Himpunan (dataset) kasus

A = Atribut

N = Jumlah partisi S

$P(x_i)$ = Proporsi dari S_i terhadap S

- c) Menghitung nilai *gain*, yaitu saat membangun pohon keputusan, pemilihan atribut *root* didasarkan pada nilai validasi tertinggi dari atribut yang ada. Hitung keuntungan menggunakan rumus dalam persamaan di bawah ini.

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

Keterangan:

S = Himpunan (dataset) kasus

A = Atribut

N = Jumlah partisi atribut A

$|S_i|$ = Jumlah kasus pada a partisi ke-i

$|S|$ = Jumlah kasus pada S

- d) Ulangi tahapan kedua dan tahapan ketiga hingga *record* terpartisi.
- e) Proses partisi pohon keputusan akan berhenti ketika semua *record* pada node N memiliki kelas yang sama, tidak ada lagi atribut pada *record* yang dipartisi dan tidak ada *record* pada cabang yang kosong.

2.5. Confusion Matrix

Confusion matrix adalah metode yang umum digunakan untuk melakukan perhitungan yang akurat untuk konsep-konsep dalam data mining. Evaluasi



menurut metode matriks konfusi menghasilkan nilai *accuracy*, *precision* dan *recall*. Untuk menggambarkan analisis kinerja pengklasifikasi C4.5, penelitian ini menerapkan metode *confusion matrix* 3x3 [8]. Pada tabel 2.1 *confusion matrix* terdiri dari tiga kelas yaitu kelas A, B, dan C, sedangkan hasil prediksi adalah hasil identifikasi yang dinilai oleh pengklasifikasi C4.5 yang juga terdiri dari tiga kelas. Kasus-kasus tersebut dibagi menjadi sembilan nilai yaitu TA, FA1, FA2, FB1, TB, FB2, FC1, FC2, dan TC [9]. Berikut tabel *confusion matrix* 3x3 yang ditunjukkan pada Tabel 2.1 berikut:

Tabel 1. Confusion Matrix 3x3

Predicted	Actual		
	A	B	C
A	TA	FA1	FA2
B	FB1	TB	FB2
C	FC1	FC2	TC

Oleh karena itu, berdasarkan metode *Confusion Matrix* yang ditunjukkan pada Tabel 1, pengukuran akurasi pada penelitian ini dilakukan dengan menggunakan formulasi berikut:

$$Accuracy = \frac{T}{T+FA1+FA2+FB1+FB2+FC1+FC2} \times 100\% \quad (3)$$

Dimana $T = TA + TB + TC$. Dengan TA adalah kelas A yang diklasifikasikan dengan benar, TB adalah kelas B yang diklasifikasikan dengan benar, TC adalah kelas C yang diklasifikasikan dengan benar, FA1 adalah kelas B yang diklasifikasikan ke dalam kelas A, FA2 adalah kelas C yang diklasifikasikan ke dalam kelas A, FB1 adalah kelas A yang diklasifikasikan ke dalam kelas B, FB2 adalah kelas C yang diklasifikasikan ke dalam kelas B, FC1 adalah kelas A yang diklasifikasikan ke dalam kelas C, FC2 adalah kelas B yang diklasifikasikan ke dalam kelas C.

Berdasarkan Tabel 2 menunjukkan perubahan dari *extended confusion matrix* berukuran 3x3 menjadi ukuran 2x2, dengan kelas "A" didefinisikan sebagai kelas positif dan kelas "Not A" sebagai kelas negatif [10].

Tabel 2. Confusion Matrix 2x2

Actual	Predicted	
	A	Not A
A	TP	FN
Not A	FP	TN

Keterangan:

TP: *True Positive* yaitu jumlah data yang memiliki nilai positif dan diprediksi benar sebagai positif

TN: *True Negative* yaitu jumlah data yang memiliki nilai negatif dan diprediksi sebagai negatif

FP: *False Positive* yaitu jumlah data yang memiliki nilai negatif tetapi diprediksi sebagai positif

FN: *False Negative* jumlah data yang memiliki nilai positif tetapi diprediksi sebagai negatif

Beberapa persyaratan yang ditetapkan untuk matriks penilaian adalah sebagai berikut [11]:

a) *Accuracy* adalah rasio jumlah prediksi yang benar dengan menggunakan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{4}$$

b) *Precision* adalah proporsi kasus positif yang diprediksi dengan benar, yang dihitung menggunakan rumus berikut:

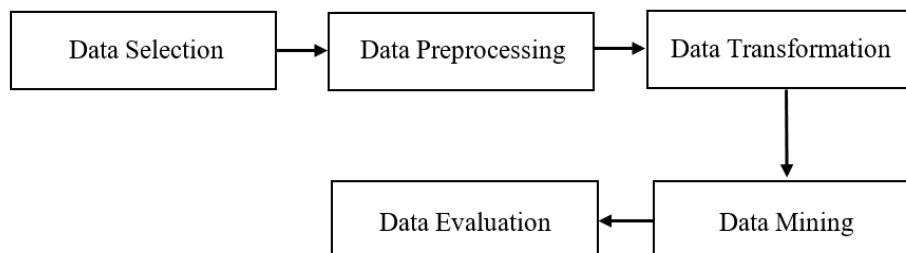
$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{5}$$

c) *Recall* adalah proporsi kasus positif yang teridentifikasi dengan benar, yang dapat dihitung dengan menggunakan rumus berikut:

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{6}$$

3. HASIL DAN PEMBAHASAN

Dalam analisis data mining, penelitian ini hanya memerlukan tahapan langkah-langkah dalam model *Knowledge Discovery in Databases* (KDD) sebagai berikut:



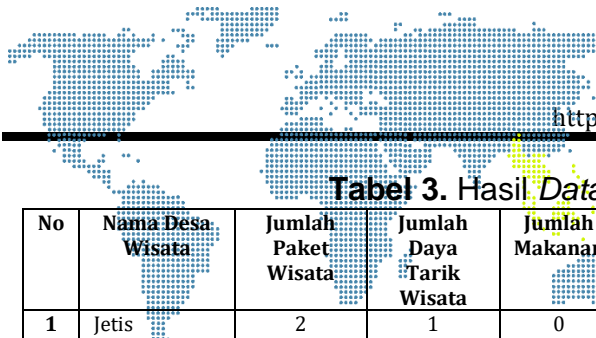
Gambar 1. Penerapan Metode *Knowledge Discovery in Databases*

3.1. Data Selection

Pemilihan data merupakan tahap memilih atribut data yang akan digunakan untuk pemodelan. Atribut yang digunakan dalam proses data mining adalah Paket Wisata, Daya Tarik, Makanan, Kerajinan, *Event*, Kesenian dan Status. Jumlah data wisata pada penelitian ini 329 namun hanya 129 yang dapat dikategorikan ke dalam label status rintisan, berkembang dan maju.

3.2. Hasil Data Preprocessing

Sebelum data diterapkan ke dalam algoritma data mining terhadap sebuah dataset, perlu dilakukan pengolahan awal yang bertujuan untuk mendapatkan dataset yang dapat diolah. Oleh karena itu, dataset penelitian ini diubah menjadi data numerik. Berikut ini *record* data yang diperoleh dari skor dataset Desa Wisata Provinsi Jawa Tengah.



Tabel 3. Hasil *Data Preprocessing*

No	Nama Desa Wisata	Jumlah Paket Wisata	Jumlah Daya Tarik Wisata	Jumlah Makanan	Jumlah Kerajinan	Event	Jumlah Kesenian	Status
1	Jetis	2	1	0	0	0	0	Rintisan
2	Ketawang	2	1	0	0	0	0	Rintisan
3	Sembungan	6	2	3	0	0	1	Maju
4	Benowo	8	6	4	0	0	3	Maju
5	Jogoboyo	5	3	0	0	1	0	Rintisan
6	Gedangan	2	2	0	0	0	0	Rintisan
7	Dewi Sri	6	2	2	0	0	2	Berkembang
8	Wana Pesona	6	2	3	3	1	1	Berkembang
9	Desa Pekasiran	9	3	2	1	0	3	Berkembang
10	Berta	9	15	10	5	0	1	Berkembang

3.3. Hasil *Data Transformation*

Pada tahap data *transformation* ini data diubah menjadi data yang disesuaikan untuk proses data mining. Terdapat beberapa atribut yang disederhanakan menjadi atribut baru yaitu atribut *k_pakwis*, *k_dawis*, *k_keraj*, *event* dan *k_kesen*. Kemudian dilakukan tehnik *discretization* yaitu proses yang mengubah variabel kontinu menjadi bentuk diskrit dan disimpan ke dalam kategori atau kelompok yang bermakna. Tahapan ini menggunakan tehnik *discretization* dengan membagi data menjadi interval yaitu dengan membuat rentang nilai. Proses ini digunakan untuk membantu data agar lebih mudah dipahami. Berikut kategori pada masing-masing atribut:

a) Paket Wisata

Atribut paket wisata memiliki nilai tipe integer dengan dibuat 3 kelas yaitu memiliki 1 hingga 4 paket wisata dengan kelas “Sedikit”, memiliki 5-7 paket wisata dengan kelas “Cukup” dan memiliki 8-10 paket wisata dengan kelas “Banyak”.

b) Daya Tarik Wisata

Atribut daya tarik memiliki nilai tipe integer dengan dibuat 3 kelas yaitu memiliki 0 hingga 3 daya tarik dengan kelas “Sedikit”, memiliki 4 hingga 7 daya tarik dengan kelas “Cukup” dan memiliki 9 hingga 15 daya tarik dengan kelas “Banyak”.

c) Makanan

Atribut makanan memiliki nilai tipe integer dengan dibuat 3 kelas yaitu memiliki 0 hingga 3 makanan dengan kelas “Sedikit”, memiliki 4 hingga 8 makanan dengan kelas “Cukup” dan memiliki 8 hingga kurang dari 35 makanan dengan kelas “Banyak”.

d) Event

Atribut *event* memiliki nilai tipe integer dengan dibuat 3 kelas yaitu memiliki 0 *event* dengan kelas “Tidak” dan memiliki 1 *event* dengan kelas “Ada”.

e) Kesenian

Atribut kesenian memiliki nilai tipe integer dengan dibuat 3 kelas yaitu memiliki 0 hingga 2 kesenian dengan kelas “Sedikit”, memiliki 3 hingga 4 kesenian dengan kelas “Cukup” dan memiliki 5 hingga 7 kesenian dengan kelas “Banyak”.



f) Kerajinan

Atribut kerajinan memiliki nilai tipe integer dengan dibuat 3 kelas yaitu memiliki 0 hingga 2 kerajinan dengan kelas “Sedikit”, memiliki 3 hingga 4 kerajinan dengan kelas “Cukup” dan memiliki 5 hingga 11 kerajinan dengan kelas “Banyak”.

Tabel 4. Hasil *Data Transformation*

k_pakwis	k_dawis	k_mak	k_keraj	event	k_kesen	status
Banyak	Sedikit	Sedikit	Sedikit	Ada	Sedikit	Berkembang
Cukup	Sedikit	Cukup	Sedikit	Tidak	Sedikit	Berkembang
Cukup	Sedikit	Cukup	Sedikit	Tidak	Cukup	Berkembang
Banyak	Sedikit	Sedikit	Sedikit	Ada	Cukup	Berkembang
Cukup	Sedikit	Sedikit	Sedikit	Ada	Sedikit	Rintisan
Sedikit	Sedikit	Sedikit	Sedikit	Tidak	Sedikit	Rintisan
Sedikit	Sedikit	Sedikit	Sedikit	Tidak	Sedikit	Maju
Sedikit	Sedikit	Sedikit	Sedikit	Tidak	Sedikit	Rintisan
Sedikit	Sedikit	Sedikit	Sedikit	Tidak	Sedikit	Rintisan
Cukup	Sedikit	Cukup	Sedikit	Tidak	Sedikit	Maju

3.4. Hasil Data Mining

Selanjutnya data diproses dengan menggunakan metode algoritma C4.5 *classifier* untuk menghasilkan klasifikasi dari rintisan, berkembang, dan maju. Kemudian hasil perhitungan tersebut dapat dilihat pada Tabel 4 dimana nilai *gain* tertinggi akan menjadi akar dari pohon.

Tabel 5. Perhitungan *Entropy* dan *Gain Node 1*

Node	Atribut	Kelas	Jumlah kasus	Rintisan	Berkembang	Maju	Entropy	Gain
1	Total		129	4	119	6	0,664	0,346
	k_pakwis	Sedikit	10	3	6	1	0,913	
		Cukup	50	1	47	2	0,638	
		Banyak	69	0	66	3	0	
	k_dawis	Sedikit	70	4	63	3	0,694	0,288
		Cukup	42	0	39	3	0	
		Banyak	17	0	17	0	0	
	k_mak	Sedikit	86	4	79	3	0,670	0,218
		Cukup	39	0	36	3	0	
		Banyak	4	0	4	0	0	
	k_keraj	Sedikit	114	4	104	6	0,678	0,065
		Cukup	13	0	13	0	0	
		Banyak	2	0	2	0	0	
	event	Ada	85	1	82	2	0,599	0,010
		Tidak	44	3	37	4	0,760	
	k_kesen	Sedikit	101	4	93	4	0,667	0,142
		Cukup	23	0	22	1	0	
		Banyak	5	0	0	1	0	

Perhitungan manual menggunakan Microsoft Excel 2019 pada Tabel 5 menghasilkan node root dari pohon keputusan dengan nilai *gain* tertinggi 0,346 yaitu atribut k_pakwis. Selanjutnya dilakukan perhitungan kelas dengan nilai *entropy* tertinggi berdasarkan masing-masing label.

Tabel 6. Perhitungan *Entropy* dan *Gain* 2.1

Node	Atribut	Kelas	Jumlah kasus	Rintisan	Berkembang	Maju	Entropy	Gain	
2.1	k_pakwis (S)		10	3	6	1	0,913		
		k_dawis	Sedikit	9	3	5	1	0,930	0,076
			Cukup	0	0	0	0	0	
		Banyak	1	0	1	0	0		
	k_mak	Sedikit	10	3	6	1	0,913	0	
		Cukup	0	0	0	0	0		
		Banyak	0	0	0	0	0		
	k_keraj	Sedikit	10	3	6	1	0,913	0	
		Cukup	0	0	0	0	0		
		Banyak	0	0	0	0	0		
	event	Ada	3	0	3	0	0	0,242	
		Tidak	7	3	3	1	0,959		
	k_kesen	Sedikit	9	3	6	1	0,963	0,046	
		Cukup	1	0	0	0	0		
		Banyak	0	0	0	0	0		

Berdasarkan Tabel 6, pada atribut k_pakwis di kelas sedikit menghasilkan nilai gain tertinggi yaitu 0,242. Sehingga atribut event menjadi internal node yang memiliki leaf node pada klasifikasi “Ada” pada label Berkembang dan memiliki leaf node pada klasifikasi “Tidak” pada label Rintisan, Berkembang, dan Maju.

Tabel 7. Perhitungan *Entropy* dan *Gain* Node 2.2

Node	Atribut	Kelas	Jumlah kasus	Rintisan	Berkembang	Maju	Entropy	Gain	
2.2	k_pakwis (C)		50	1	47	2	0,638		
		k_dawis	Sedikit	39	1	36	2	0,662	0,122
			Cukup	9	0	9	0	0	
		Banyak	2	0	2	0	0		
	k_mak	Sedikit	36	1	35	0	0	0,638	
		Cukup	14	0	12	2	0		
		Banyak	0	0	0	0	0		
	k_keraj	Sedikit	48	1	45	2	0,642	0,022	
		Cukup	2	0	2	0	0		
		Banyak	0	0	0	0	0		
	event	Ada	30	1	28	1	0,649	0,248	
		Tidak	20	0	19	1	0		
	k_kesen	Sedikit	44	1	41	2	0,650	0,066	
		Cukup	6	0	6	0	0		
		Banyak	0	0	0	0	0		

Berdasarkan Tabel 7, internal node lanjutan dari atribut k_pakwis dengan kelas “Cukup”. Dengan nilai gain tertinggi terletak pada atribut k_mak yaitu sebesar 0,638. Sehingga atribut k_mak menjadi internal node dengan kelas “Sedikit” yang dan “Cukup” menjadi internal node dari pohon keputusan selanjutnya.

Tabel 8. Perhitungan *Entropy* dan *Gain* Node 2.3

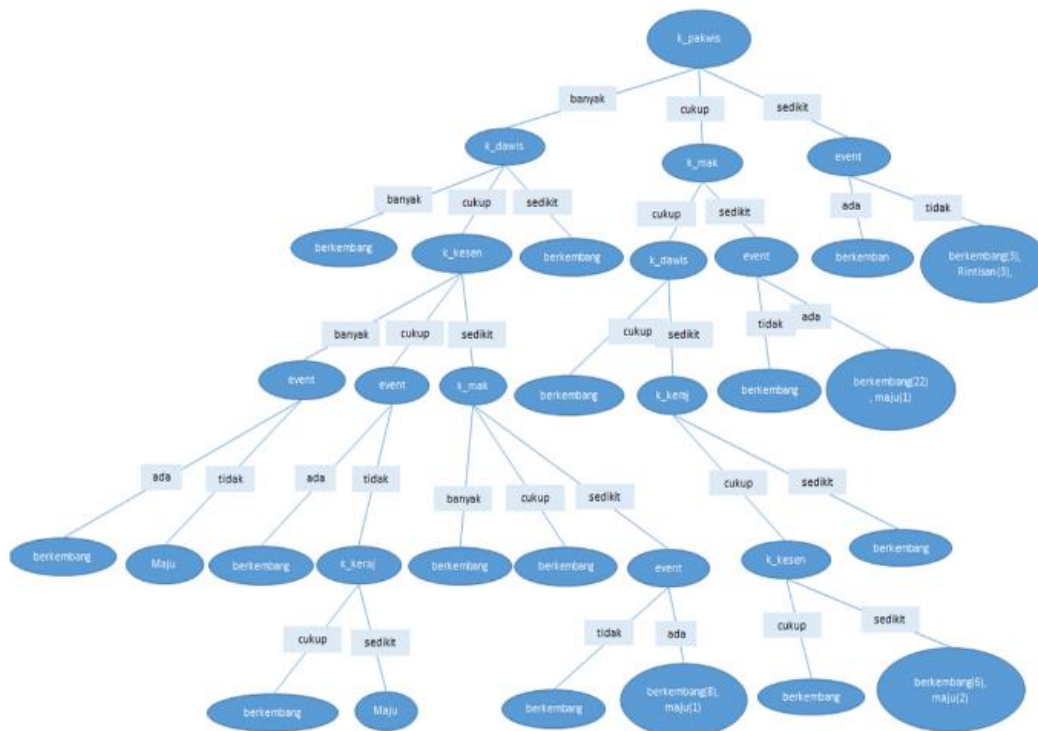
Node	Atribut	Kelas	Jumlah kasus	Rintisan	Berkembang	Maju	Entropy	Gain
2.3	k_pakwis (B)		69		66	3	0,258	
		k_dawis	Sedikit	22		22	0	0
		Cukup	33		30	3	0,439	
		Banyak	14		14	0	0	



Node	Atribut	Kelas	Jumlah kasus	Rintisan	Berkembang	Maju	Entropy	Gain
	k_mak	Sedikit	40		38	2	0,286	0,004
		Cukup	25		24	1	0,242	
		Banyak	4		4	0	0	
	k_keraj	Sedikit	56		53	3	0,301	0,013
		Cukup	11		11	0	0	
		Banyak	2		2	0	0	
event		Ada	52		51	1	0,137	0,026
		Tidak	17		15	2	0,523	
	k_kesen	Sedikit	47		46	1	0,149	0,025
		Cukup	17		16	1	0,323	
		Banyak	5		4	1	0,722	

Berdasarkan Tabel 8, nilai *gain* tertinggi terletak pada atribut *k_dawis* yaitu sebesar 0,048. Sehingga atribut *k_dawis* dengan kelas “Cukup” menjadi *internal node* dari pohon keputusan selanjutnya.

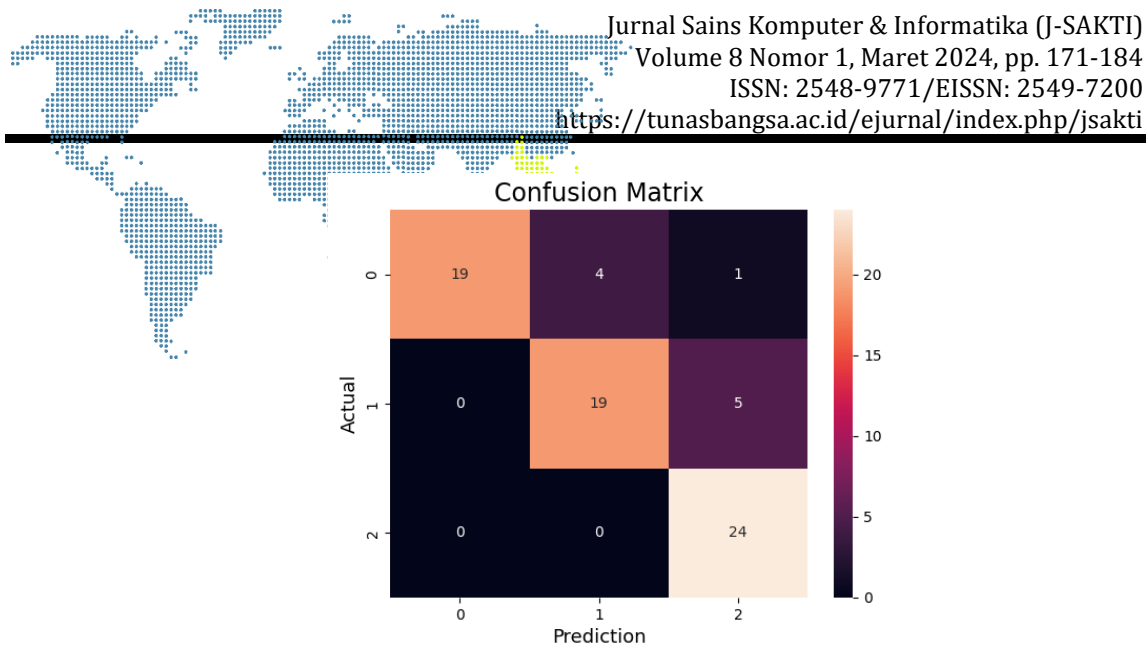
Setelah dilakukan perhitungan keseluruhan *entropy* dan *gain* secara manual menggunakan Microsoft Excel, maka pohon keputusan yang terbentuk dapat dilihat seperti gambar berikut:



Gambar 2. Hasil Decision Tree

3.5. Hasil Data Evaluation

Berikut adalah hasil pengujian *confusion matrix* menggunakan 80% data latih dan 20% data uji pada Gambar 3 dibawah ini:



Gambar 3. Visualisasi *Confusion Matrix 3x3*

Berdasarkan Gambar 3 mengenai metode algoritma C4.5 classifier, nilai actual atau *grand truth confusion matrix* diatas terbagi menjadi 3 bagian yaitu nilai prediksi rintisan dengan nilai *true* positif sebanyak 19 data, prediksi rintisan dengan nilai *true* negatif sebanyak 48 data, prediksi bukan rintisan dengan nilai *false* positif sebanyak 0 data, dan prediksi bukan rintisan dengan nilai *false* negatif sebanyak 5 data. Selanjutnya nilai prediksi berkembang dengan nilai *true* positif sebanyak 19 data, prediksi berkembang dengan nilai *true* negatif sebanyak 44 data, prediksi bukan berkembang dengan nilai *false* positif sebanyak 4 data, dan prediksi bukan berkembang dengan nilai *false* negatif sebanyak 5 data. Terakhir nilai prediksi maju dengan nilai *true* positif sebanyak 24 data, prediksi maju dengan nilai *true* negatif sebanyak 42 data, prediksi bukan maju dengan nilai *false* positif sebanyak 6 data, dan prediksi bukan maju dengan nilai *false* negatif sebanyak 0 data.

Model *Confusion Matrix* dengan menggunakan klasifikasi algoritma C4.5 menghasilkan *Accuracy* sebesar 86%, *Precision* 86%, *Recall* 88%, dan F1 sebesar 86%. Model pengujian dilakukan dalam bentuk pemrograman *python* menggunakan *Google Colab*, hasil pengujian ditunjukkan pada Gambar 4. berikut.

```
[ ] print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	1.00	0.79	0.88	24
1	0.83	0.79	0.81	24
2	0.80	1.00	0.89	24
accuracy			0.86	72
macro avg	0.88	0.86	0.86	72
weighted avg	0.88	0.86	0.86	72

Gambar 4. Hasil *Confusion Matrix* menggunakan *Google Colab*

Berdasarkan Tabel 9. pengujian juga dilakukan dengan menggunakan persamaan *confusion matrix* dengan bantuan *Microsoft Excel 2019*.

Tabel 9. Hasil perhitungan *Confusion Matrix*

Desa Wisata	Recall	Precision	F1
Rintisan	79%	100%	88%
Berkembang	79%	83%	81%
Maju	100%	80%	89%
Accuracy : 86%			

Hasil setiap item dievaluasi dengan menggunakan *confusion matrix*. Berikut ini adalah formula matematika untuk akurasi, presisi, *recall* untuk setiap skor item:

$$Accuracy = \frac{T}{T+FA1+FA2+FB1+FB2+FC1+FC2} \times 100\%$$

$$= \frac{52}{19+19+24} = \frac{52}{72} = 0,86 \times 100\% = 86\%$$

Berdasarkan Persamaan 3, dalam hal menghitung persamaan accuracy yang digunakan adalah TP/Total data, dimana nilai TP adalah setiap kelas yang pada kondisi kelas sebenarnya dapat diprediksi dengan benar. Sehingga nilai *confusion matrix* yang pertama yaitu *accuracy* memperoleh sebesar 86%.

$$Rintisan Recall = \frac{TP}{TP+FN} = \frac{19}{19+5} = \frac{19}{24} = 0,79 \times 100\% = 79\%$$

$$Berkembang Recall = \frac{TP}{TP+FN} = \frac{19}{19+5} = \frac{19}{24} = 0,79 \times 100\% = 79\%$$

$$Maju Recall = \frac{TP}{TP+FN} = \frac{24}{24+0} = \frac{24}{24} = 1 \times 100\% = 100\%$$

Berdasarkan Persamaan 6, nilai *confusion matrix* kedua adalah *recall* dengan menghasilkan 79% yang menunjukkan bahwa 5 sample tidak dapat diberi label sebagai rintisan. Kemudian terdapat 5 kejadian yang tidak dapat diberi label sebagai berkembang dengan menunjukkan *recall* berkembang sebesar 79%. Semua sampel yang diberi label dengan akurasi 100%.

$$Rintisan Precision = \frac{TP}{TP+FP} = \frac{19}{19+0} = \frac{19}{19} = 1 \times 100\% = 100\%$$

$$Berkembang Precision = \frac{TP}{TP+FP} = \frac{19}{19+4} = \frac{19}{23} = 0,83 \times 100\% = 83\%$$

$$Maju Precision = \frac{TP}{TP+FP} = \frac{24}{24+6} = \frac{24}{30} = 0,8 \times 100\% = 80\%$$

Berdasarkan Persamaan 2.5, nilai *confusion matrix* ketiga adalah *precision*. Semua sampel berlabel rintisan menunjukkan hasil klasifikasi sebesar 100%. Selanjutnya berkembang menunjukkan *precision* sebesar 83% terdapat 4 kejadian yang tidak diklasifikasikan sebagai label berkembang. Terakhir, nilai *precision* untuk label maju adalah 80% karena terdapat 6 kejadian yang tidak diklasifikasikan sebagai maju.

4. SIMPULAN

Kesimpulan penelitian terlihat bahwa implementasi Algoritma C4.5 dapat digunakan untuk klasifikasi desa wisata di Jawa Tengah. Pohon keputusan Algoritma C4.5 menunjukkan bahwa atribut *k_pakwis* atau kriteria paket wisata menjadi node akar pada pohon keputusan dengan nilai Gain tertinggi yaitu 0,346. Tingkat akurasi dengan menggunakan 7 atribut yaitu paket wisata, daya tarik, makanan, kerajinan, kesenian, *event* dan status yaitu cukup baik. Akurasi terbaik diperoleh dengan menggunakan persentase data training sebesar 80% dan data testing sebesar 20%. Hasil penelitian ini berupa pengujian algoritma C4.5 dengan menggunakan *confusion matrix* dan memperoleh *accuracy* sebesar 86%, *recall* sebesar 88% dan *precision* sebesar 86%, yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk klasifikasi desa wisata dan diharapkan bisa menjadi sumber informasi untuk penelitian selanjutnya mengenai klasifikasi menggunakan algoritma C4.5.

DAFTAR PUSTAKA

- [1] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [2] S. B. Helpiastuti, "Pengembangan Destinasi Pariwisata Kreatif Melalui Pasar Lumpur (Analisis Wacana Grand Opening 'Pasar Lumpur' Kawasan Wisata Lumpur, Kecamatan Ledokombo, Kabupaten Jember)," *J. Tour. Creat.*, vol. 2, no. 1, pp. 13–23, 2018, [Online]. Available: <https://jurnal.unej.ac.id/index.php/tourismjournal/article/download/13837/7204/>.
- [3] N. Nazeriandy, Y. Syahra, and M. Syaifudin, "Penerapan Data Mining Untuk Memprediksi Penggunaan Daya Listrik Pada PT.PLN (Persero) Rayon Medan Selatan Dengan Menggunakan Metode Regresi Linier Berganda," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 20, no. 1, p. 20, 2021, doi: 10.53513/jis.v20i1.2431.
- [4] A. S. Chan, "Prediksi Kedatangan Wisatawan Pada Pariwisata Kota Batam Dengan Menggunakan Teknik Knowledge Data Discovery," *J. Ilm. Inform.*, vol. 6, no. 01, p. 11, 2018, doi: 10.33884/jif.v6i01.432.
- [5] A. Alfiani Mahardhika, R. Saptono, and R. Anggrainingsih, "Sistem Klasifikasi Feedback Pelanggan Dan Rekomendasi Solusi Atas Keluhan Di UPT Puskom UNS Dengan Algoritma Naive Bayes Classifier Dan Cosine Similarity," *J. Teknol. Inf. ITSmart*, vol. 4, no. 1, p. 36, 2016, doi: 10.20961/its.v4i1.1806.
- [6] E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 3, no. July, p. 1, 2018, doi: 10.30645/jurasik.v3i0.60.
- [7] M. Ardiansyah Sembiring, M. Fitri Larasati Sibuea, A. Sapta, P. Studi Sistem Informasi, and S. Royal, "Analisa Kinerja Algoritma C.45 Dalam Memprediksi Hasil Belajar," *J. Sci. Soc. Res.*, vol. 1, no. February, pp. 73–79, 2018, [Online]. Available: <http://jurnal.goretanpena.com/index.php/jSSR>.
- [8] R. Nursyahfitri, A. N. Maharadja, R. A. Farissa, and Y. Umaidah, "Klasifikasi Penentuan Jenis Obat Menggunakan Algoritma Decision Tree," *J. Inform. Polinema*, vol. 7, no. 3, pp. 53–60, 2021, doi: 10.33795/jip.v7i3.629.
- [9] A. Faricha, Suwito, M. Rivai, M. A. Nanda, D. Purwanto, and R. P. R. Anhar, "Design of electronic nose system using gas chromatography principle and Surface Acoustic

- Wave sensor," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 16, no. 4, pp. 1458–1467, 2018, doi: 10.12928/TELKOMNIKA.v16i4.7127.
- [10] A. Sepharni, I. E. Hendrawan, and C. Rozikin, "Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 7, no. 2, p. 117, 2022, doi: 10.30998/string.v7i2.12012.
- [11] S. Tangirala, "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, pp. 612–619, 2020, doi: 10.14569/ijacsa.2020.0110277.
- [13] A. Sepharni, I. E. Hendrawan, and C. Rozikin, "Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 7, no. 2, p. 117, 2022, doi: 10.30998/string.v7i2.12012.