

BAB III

METODOLOGI PENELITIAN

3.1. Subjek dan Objek Penelitian

Subjek pada penelitian ini merupakan Provinsi Jawa Tengah. Sedangkan objek dalam penelitian ini berlandaskan dari tujuan penelitian yaitu dataset desa wisata menggunakan teknik data mining dengan *desicon tree classifier* algoritma C4.5.

3.2. Alat dan Bahan Penelitian

3.2.1. Alat Penelitian

Penelitian ini didukung dengan alat dan bahan yang terdiri dari perangkat keras dan perangkat lunak, yaitu:

Tabel 3.1 Alat Penelitian

Jenis Perangkat	Nama Alat
Perangkat Keras	Laptop Asus dengan spesifikasi <i>Operating System</i> Windows 11 Home Single Language 64-bit, digunakan untuk menyusun laporan proposal, mencari referensi penelitian terdahulu, serta mengolah data.
Perangkat Lunak	Microsoft Excel, digunakan untuk seleksi dan transformasi data ke dalam bentuk yang sesuai dengan format pengolahan data.
	<i>Google Colab</i> , digunakan untuk membantu user dalam membuat pemrograman dan pengolahan data
	<i>Python</i> , digunakan untuk melakukan pengujian terhadap dataset penelitian.

3.2.2. Bahan Penelitian

Pada penelitian ini, bahan yang digunakan adalah dataset desa wisata di Provinsi Jawa Tengah tahun 2019 yang diperoleh dari website portal data Provinsi Jawa Tengah, yaitu data.jatengprov.go.id dan data berupa kriteria-kriteria dalam menentukan kategori desa wisata di Jawa Tengah berdasarkan Peraturan Gubernur Nomor 53 Tahun 2019. Kriteria yang dibutuhkan dalam penelitian ini antara lain:[3]

Tabel 3.2 Kriteria Desa Wisata

No	Atraksi wisata yang menarik dan atraktif pada Desa Wisata
1	Mempunyai paket wisata yang menjadi keunikan daerah di Desa seperti sosial budaya, kesenian, <i>heritage</i> , aksesoris khas daerah, sejarah, batik, lingkungan, budaya, kuliner, <i>ecotourism</i> .
2	Mempunyai makanan lokal di Desa Wisata
3	Mempunyai kerajinan lokal di Desa Wisata
4	Mempunyai daya tarik wisatawan
5	Mempunyai <i>event</i> Desa Wisata
6	Mempunyai kesenian lokal Desa Wisata

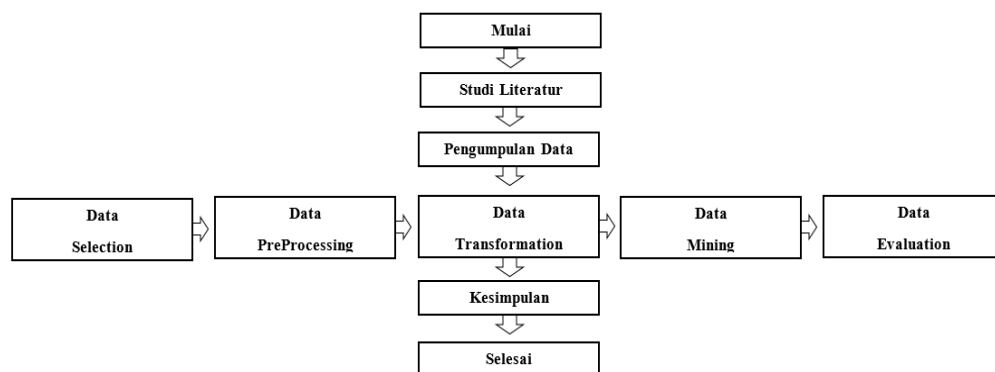
Berdasarkan skoring penilaian Desa Wisata, maka terdapat pengelompokan desa wisata sesuai dengan keputusan kepala Disporapar (Dinas Kepemudaan, Olahraga dan Pariwisata) yang meliputi:

Tabel 3.3 Kategori Desa Wisata

No	Status Desa Wisata
1	Desa Wisata Rintisan
2	Desa Wisata Berkembang
3	Desa Wisata Maju

3.3. Diagram Alir Penelitian

Penelitian ini dibagi menjadi beberapa bagian, yaitu pengumpulan data, dan beberapa tahapan pengolahan data dengan menerapkan tahapan-tahapan *knowledge discovery* pada database [15].



Gambar 3.1 Diagram Alir

3.3.1. Studi Literatur

Tahapan studi literatur ini mengenai pengumpulan informasi yang berkaitan dengan penelitian yang akan dilakukan. Informasi yang dikumpulkan dari studi literatur ini diperoleh dari jurnal, artikel, buku, dan situs web resmi untuk mempelajari penggunaan metode algoritma C4.5 pada penelitian sebelumnya sebagai panduan dalam menyusun penelitian ini.

3.3.2. Pengumpulan Data

Tahapan pengumpulan data ini dilakukan dengan mengumpulkan data-data yang diperlukan untuk proses algoritma klasifikasi C4.5. Dataset desa wisata merupakan data yang didapatkan langsung dari situs portal data disporapar (Dinas Kepemudaan, Olahraga dan Pariwisata) provinsi Jawa Tengah yaitu <https://data.jatengprov.go.id/> sejumlah 129 desa wisata pada tahun 2019.

3.3.3. Data Selection

Pemilihan data dari dataset operasional dilakukan pada KDD (*Knowledge Discovery in Database*) sebelum dimulainya fase penambangan data. Data hasil seleksi kemudian diproses dan disimpan dalam file terpisah dari *database* operasional. Seleksi data dapat berupa pemilihan atribut/variabel dari data asli. Atribut yang terpilih yaitu kriteria paket wisata, daya tarik wisata, kesenian, kerajinan, makanan, *event* dan status sebagai label/data target yang selanjutnya akan diolah menjadi data siap proses data mining.

3.3.4. Pre-Processing Data

Tahap *pre-processing* ini dilakukan untuk mengolah data yang tidak diperlukan. *Pre-processing* merupakan proses data *mining* yang akan digunakan dalam melakukan *cleaning* atau pembersihan pada dataset. *Cleaning* data dilakukan dengan menghilangkan data yang memiliki *noise* atau data yang tidak konsisten (*missing value*) dan memperbaiki kesalahan data. Oleh karena itu tahapan ini melakukan identifikasi data, menghapus data yang tidak relevan, menghapus data kosong, melakukan pengisian data kosong, serta menentukan data target(label). Tahapan *preprocessing* dapat dilihat pada Gambar 3.2 berikut:

NAMA KABUPATEN	NAMA DESA	DAYA TARIK			STATUS	EVENT	MARIKINA	KERAMIS	KESUNDA	
		Alam	Budaya	Buatan						Lainnya
Kota Tengah Banjarbaru	Desa Wisata Karang Tengah	Telaga Merdada	Makam Mbah Husayyid	Home Screen (Agro Wisata)	Berkemban	Festival Merdada	Keripik Kentang Jemp Carica		Sendo Tari Merdada	
	Desa Wisata (Desa Dwaraban)	Tubing Grogga		Kolam Reuang	Berkemban		Nasi Jagung dan Keripik Buntal: Kuchan/ur	Kerajinan Renda Kapang. Lili-lukan		
	Desa Gita (Desa Gintara)	Air Terjun dan Air Panas			Wisata Agro (wisata petikanyor)	Berkemban		Masakan waluh. Soto Jerdelai. Keripik Indan	Kerajinan hamba dan hata	
	Tuk Kumala (Desa Kallianjer)	Bukit Amara Smk (Klub Budaya)	Boyang Oyad Genggong	Rumah Pakan dll		Berkemban	Event Oyad Genggong	Keripik Saka	Penbuatan Tadi tradisional. Tak. Insi.	
	Candradimuka (Desa Pakstran)	Sumur Kalimanda Kawah Candradimuka Telaga D'Elgan				Berkemban		Demplo (jagung dari alat pemakan padi)	Kerajinan Pencak silat Japin	
	Dong Pundara (Dong Kulen)	Landscape Candi dan Derasmasia. Kembang. Bukit Selent. Bukit Pangman	Kongkole Candi dan Derasmasia. Bukit Pemasulita	Museum Kallian	Kerajinan Kupu-kupu	Berkemban	DCF (Dong Agrowisata a hingga jala lilin dan carica	Carica. Parwaw & Sate donda dng	Misiatir Batik Kayu	Leggan. Bungk. Yakan

k_pakwis	k_dawis	k_mak	k_keraj	event	k_kesen
9	2	1	4	Tidak	1
9	6	1	1	Tidak	1
6	2	2	0	Tidak	1
7	2	0	0	Ada	3
7	1	1	0	Tidak	1
5	1	2	0	Tidak	0
8	8	2	0	Tidak	7
8	5	2	0	Tidak	2
8	4	1	1	Tidak	1
8	6	4	0	Tidak	3
7	9	0	0	Tidak	1
5	3	0	0	Ada	0
2	2	0	0	Tidak	0
2	3	0	0	Tidak	0
2	1	0	0	Tidak	0
2	1	0	0	Tidak	0
2	1	0	0	Tidak	0
9	5	8	3	Ada	1
9	9	1	1	Ada	3
9	9	6	1	Ada	6
9	6	1	2	Ada	5
8	8	1	1	Tidak	1
9	5	1	4	Tidak	2

Gambar 3.2 Sampel Proses Data Mentah menjadi Data Siap Proses

3.3.5. Transformation

Pada tahap ini, data ditransformasikan ke dalam format data yang sesuai untuk proses data mining. Salah satu teknik yang dapat digunakan dalam proses transformasi data adalah mengubah tipe data kontinu menjadi tipe data diskrit. Hal ini digunakan untuk mengurangi jumlah atribut numerik dengan cara membuat label *range* dan membagi *range* dari atribut tersebut. 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”.

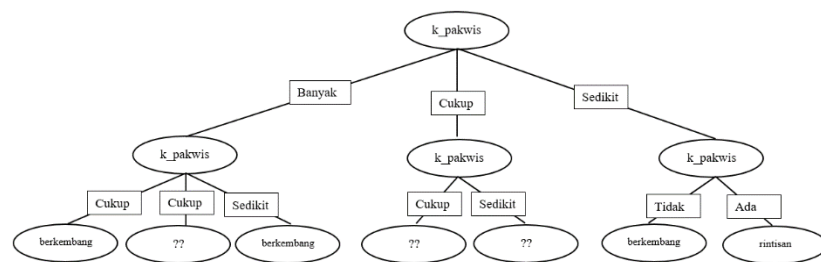
f. Kerajinan

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”.

3.3.6. **Data Mining**

Data yang sudah melalui tahapan *pre-processing* dan transformasi data, selanjutnya dilakukan pengolahan data dengan menggunakan metode atau algoritma data mining. Penelitian ini menggunakan algoritma C4.5 untuk mengklasifikasi desa wisata dengan menggunakan algoritma C4.5 membentuk *decision tree* berdasarkan aturan pohon keputusan tersebut.

Tahapan ini melakukan proses perhitungan awal sesuai dengan persamaan Algoritma C4.5 yaitu dataset penelitian, kemudian menghitung *entropy*, menghitung nilai *gain*, menentukan atribut sebagai pohon berdasarkan nilai *gain* tertinggi, lalu ulangi proses kedua hingga semua kelas terpartisi. Contoh pohon keputusan dapat dilihat pada Gambar 3.3 berikut:

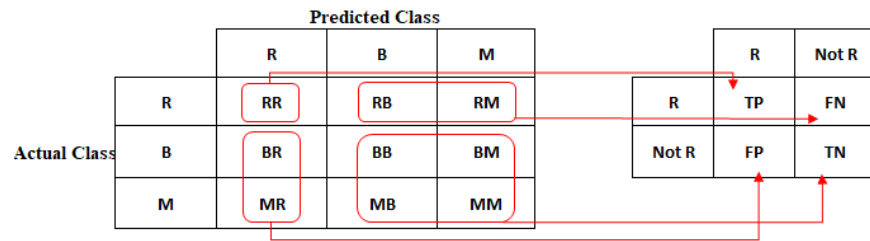


Gambar 3.3 Sampel Hasil dari Perhitungan *Entropy* dan *Gain*

3.3.7. Evaluation

Pada tahap ini, setelah mendapatkan model yang diperoleh dari algoritma C4.5, dilakukan pengujian menggunakan *confusion matrix* terhadap dataset yang ada. *Confusion Matrix* adalah *visualization tool* yang umum digunakan pada *supervised learning*. Setiap kolom dari matriks merupakan contoh dari kelas yang diprediksi, sementara setiap baris merepresentasikan kejadian pada kelas yang sesungguhnya. Hasil dari proses perhitungan *confusion matrix* berupa 3 *output* yaitu *recall*, *precision*, *accuracy*. Pada tahapan ini, implementasi dilakukan dalam bentuk pemrograman *Python* untuk mendapatkan hasil *forecast*.

Berdasarkan banyaknya kelas/label yang digunakan pada dataset, maka tahap evaluasi ini menggunakan *confusion matrix* prediksi klasifikasi multiclass. Berikut Gambar 3.4 yang menunjukkan perubahan dari *extended confusion matrix* 3x3 menjadi *confusion matrix* 2x2:



Gambar 3.4 *Extended Confusion Matrix 3x3* [31]

3.3.8. Hasil Penelitian

Pada tahap ini, hasil penelitian dilakukan untuk mengetahui bahwa hasil pengujian yang dilakukan sudah sesuai dengan tujuan penelitian yang memberikan keterangan tingkat akurasi prediksi dan menghasilkan pohon keputusan sebagai *output*.