

BAB II

METODE PENELITIAN

2.1 Objek Penelitian

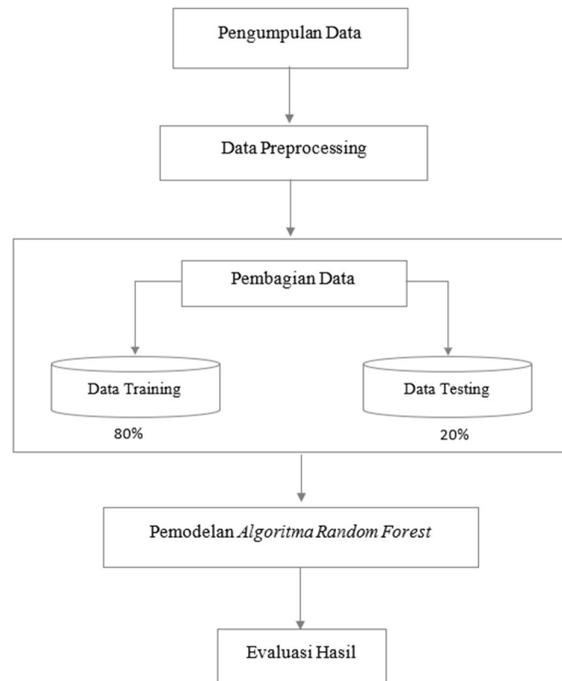
Penelitian ini dilaksanakan di Dinas Pangan Kab. Berau yang berlokasi di Jl. Dr. Murjani I No.80, Karang Ambun, Kec. Tanjung Redeb, Kabupaten Berau, Kalimantan Timur dari tanggal 1 Maret 2024 sd tanggal 30 Mei 2024 dengan menggunakan data *Food Security and Vulnerability Atlas* (FSVA) Kab. Berau tahun 2023. Data FSVA Kab. Berau terdiri dari keseluruhan desa/kelurahan yang ada di Kabupaten Berau yang dikelompokkan berdasarkan nilai komposit dan indikator-indikator penilaian seberapa rawan pangan desa/kelurahan.

2.2 Alat dan Bahan

Pada penelitian ini membutuhkan alat dan bahan dalam melakukan penelitian. Adapun alat dan bahan yang digunakan berupa (i) Laptop Lenovo Ideapad Slim 3, (ii) Processor AMD Ryzen 3 3250U, (iii) RAM 4GB DDR4, (iv) Mouse, (v) Windows 10 Home 64-bit, (vi) Bahasa Pemrograman *Python*, (vii) Microsoft Home Student 2019, (viii) Website *Google Collabs*, (ix) Data FSVA Kabupaten Berau tahun 2023.

2.3 Prosedur Penelitian

Pada penelitian ini berfokus pada klasifikasi daerah rawan pangan pada desa/kelurahan menggunakan *algoritma Random Forest*. Prosedur penelitian dimulai dengan pengumpulan data, data *preprocessing*, pemodelan *algoritma random forest*, hingga evaluasi akhir. Berikut adalah prosedur penelitian yang akan dilakukan:



Gambar 2. 1 Prosedur Penelitian

2.3.1 Pengumpulan data

Tahapan awal pada penelitian ini yaitu pengumpulan data, data yang dikumpulkan adalah data desa/kelurahan di kabupaten berau pada tahun 2023. Data yang didapatkan dari Dinas Pangan Kabupaten Berau sebanyak 110 total baris dengan 14 atribut. Pada Tabel 2.1 ditunjukkan untuk dapat memudahkan dalam memahami dan mengetahui berapa banyak atribut pada data *FSVA (Food Security and Vulnerability Atlas)* Kabupaten Berau, maka dibentuk tabel sebagai berikut.

Tabel 2. 1 Keterangan Atribut Data FSVA (*Food Security and Vulnerability Atlas*)

Kabupaten Berau			
NO	Atribut	Type Data	Keterangan
1	No	Integer	Nomor urut data
2	Nama Kec	String	Nama Kecamatan dari desa/kelurahan
3	Kode Kec	Integer	Kode Kecamatan dari desa/kelurahan
4	Kode Desa	Integer	Kode Desa
5	Nama Desa	String	Nama Desa
6	Rasio Lahan	Integer	Rasio Lahan Pertanian
7	Rasio Sarana	Integer	Rasio Sarana dan Prasarana Penyedia Pangan
8	Rasio Pddk Tidak Sejahtera	Integer	Rasio Penduduk Tingkat Kesejahteraan Terendah
9	Akses Jalan	Integer	Rasio Desa Tanpa Akses Penghubung Memadai
10	Rasio Tanpa Air Bersih	Integer	Rasio Jumlah Rumah Tangga Tanpa Akses Air Bersih
11	Rasio Pddk per Tenkes per <i>Density</i>	Integer	Rasio Jumlah Tenaga Kesehatan
12	INDEKS KOM	Numeric	Skor dari kondisi ketahanan pangan di suatu wilayah
13	PERINGKAT	Integer	Pemeringkatan relative antara wilayah berdasarkan indeks komposit
14	PRIO KOMP	Integer	Tingkat kerentanan pangan suatu daerah (desa/kelurahan) terhadap kerawanan pangan

Adapun penjelasan target rawan pangan terdapat pada tabel 2.2 dibawah ini.

Tabel 2. 2 Target Rawan Pangan

Class	Keterangan
Prioritas 1	Sangat Rentan Pangan
Prioritas 2	Rentan Pangan
Prioritas 3	Agak Rentan Pangan
Prioritas 4	Agak Tahan Pangan

Prioritas 5	Tahan Pangan
Prioritas 6	Sangat Tahan Pangan

2.3.2 Data Pre-Processing

Persiapan data dilakukan untuk memperbaiki masalah yang terdapat pada data dengan melakukan seleksi data, pembersihan data, dan pembagian data.

1. Seleksi Data

Pada tahap ini, pemilihan data dilakukan dengan memilih fitur atau atribut yang relevan untuk digunakan pada penelitian sedangkan fitur yang dianggap tidak relevan akan dihilangkan. Tabel 2.3 menampilkan data awal yang telah didapatkan dari Dinas Pangan Kabupaten Berau yang memiliki 14 kolom. Setelah proses seleksi data didapatkan, 5 kolom dianggap tidak relevan untuk digunakan dalam mengklasifikasikan wilayah rawan pangan, setelah dilakukan proses seleksi secara manual, data yang digunakan menjadi 7 kolom yang dijadikan fitur/atribut dan 1 kolom yang akan dijadikan sebagai target atau kelas.

Tabel 2. 3 Seleksi Data

NO	Atribut	Type Data	Keterangan
1	Nama Desa	String	-
2	P.Lahan	Integer	Atribut
3	P.Sarana	Integer	Atribut
4	P.TdkSejahtera	Integer	Atribut
5	P.Jalan	Integer	Atribut
6	P.NoWater	Integer	Atribut
7	P.Tenkes	Integer	Atribut
8	INDEKS KOM	Numeric	Atribut
9	PRIO KOMP	Integer	Class/target

2. Pembersihan Data

Pada tahap ini dilakukan pembersihan data dengan cara menghilangkan data yang tidak lengkap atau kosong serta menghilangkan *noise* dan data duplikat sehingga data tersebut dapat diproses untuk melakukan proses klasifikasi data. Pembersihan data atau data *cleaning* yaitu proses mengidentifikasi, memperbaiki, menghapus kesalahan, dan ketidak akuratan, atau ketidak konsistenan dalam data untuk meningkatkan kualitas dan reliabilitas data (Triyanto et al., 2023). Setelah semua data yang diperlukan melewati tahap pembersihan data, maka data tersebut disimpan dalam dataset baru.

2.3.3 Pembagian Data

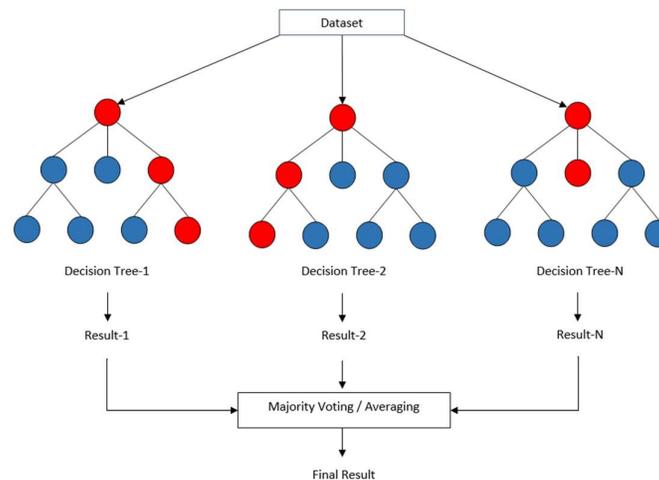
Proses pembagian data dilakukan sebelum melakukan pemodelan *algoritma random forest*, penelitian ini menggunakan teknik *split validation*. *Split validation* adalah teknik validasi yang secara acak membagi data dengan perbandingan antara data *training* dan *testing* dengan rasio 0,9 sampai

dengan 0,5 (Ziaul et al., 2023). Pembagian data *training* dan data *testing* merupakan salah satu faktor yang menentukan nilai akurasi, sehingga kesalahan dalam menentukan komposisi kedua tipe data tersebut akan mempengaruhi nilai akurasi yang diperoleh (Baiq Nurul Azmi et al., 2023). Pembagian data dilakukan dengan membagi dataset menjadi dua bagian untuk data *training* sebesar 80% digunakan dalam membangun model, kemudian untuk data *testing* 20% digunakan sebagai evaluasi model yang telah dibangun (Pratama et al., 2024).

2.3.4 Pemodelan *Algoritma Random Forest*

Random Forest adalah pengembangan dari metode *Decision Tree* yang menggunakan beberapa *Decision Tree*, dimana setiap *Decision Tree* telah dilakukan pelatihan menggunakan sampel dan setiap atribut dibagi pada pohon yang dipilih diantara subset atribut yang bersifat acak (Supriyadi et al., 2020). Metode klasifikasi ini lebih stabil dan konsisten untuk data yang beragam dan kompleks, keuntungan lain dari *random forest* adalah dapat menangani kumpulan data yang sangat besar dengan karakteristik yang beragam dengan baik tanpa memerlukan pra-pemrosesan data yang rumit (Breiman, 2001). *Random Forest* memiliki beberapa kelebihan, dapat meningkatkan hasil akurasi jika terdapat data yang hilang, serta efisien untuk penyimpanan sebuah data (Chairunisa et al., 2020).

Adapun struktur dari *random forest* digambarkan pada Gambar 2.2:



Gambar 2. 2 Struktur *Random Forest* (Adnin Kamila et al., 2023)

2.3.5 Evaluasi

Evaluasi dilakukan untuk mendapatkan hasil yang sesuai terhadap model yang telah dibangun menggunakan teknik pengukuran *confusion matrix*. *Confusion matrix* adalah metode pengukuran untuk menilai konsep ide data mining, *confusion matrix* digambarkan dengan tabel yang menunjukkan persentase data uji yang diklasifikasikan dengan benar dan yang diberi label salah (Oon Wira Yuda et al., 2022). Tabel confusion matrix 3x3 ditunjukkan pada Tabel 2.4 (Fahmy Amin, 2023).

Tabel 2. 4 Metode *Confusion Matrix* 3x3

		Class Prediksi		
		Kelas A	Kelas B	Kelas C
Nilai Aktual	Kelas A	AA	AB	AC
	Kelas B	BA	BB	BC
	Kelas C	CA	CB	CC

Keterangan:

1. AA adalah kelas yang diklasifikasikan dengan benar sebagai kelas A.
2. AB adalah kelas yang diklasifikasikan dengan benar sebagai kelas A tetapi hasil prediksi menjadi kelas B.
3. AC adalah kelas yang diklasifikasikan dengan benar sebagai kelas A tetapi hasil prediksi menjadi kelas C.
4. BA adalah kelas yang diklasifikasikan dengan benar sebagai kelas B tetapi hasil prediksi menjadi kelas A.
5. BB adalah kelas yang diklasifikasikan dengan benar sebagai kelas B.
6. BC adalah kelas yang diklasifikasikan dengan benar sebagai kelas B tetapi hasil prediksi menjadi kelas C.
7. CA adalah kelas yang diklasifikasikan dengan benar sebagai kelas C tetapi hasil prediksi menjadi kelas A.
8. CB adalah kelas yang diklasifikasikan dengan benar sebagai kelas C tetapi hasil prediksi menjadi kelas B.
9. CC adalah kelas yang diklasifikasikan dengan benar sebagai kelas C.

Nilai akurasi melakukan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Pada persamaan (2.1) adalah perhitungan nilai akurasi kinerja model. Selain *Accuracy*, pengukuran evaluasi kinerja model klasifikasi lainnya seperti *presisi*, *recall* dan *F1-Score* ditunjukkan pada persamaan (2.2-2.4). (Suci Pania et al., 2024).

$$\text{Accuracy} = \frac{AA + BB + CC}{AA + AB + AC + BA + BB + BC + CA + CB + CC} \times 100\% \quad 2.1$$

$$\text{Presisi} = \frac{\frac{AA}{AA + BA + CA} + \frac{BB}{AB + BB + CB} + \frac{CC}{AC + BC + CC}}{3} \quad 2.2$$

$$\text{Recall} = \frac{\frac{AA}{AA + AB + AC} + \frac{BB}{BA + BB + BC} + \frac{CC}{CA + CB + CC}}{3} \quad 2.3$$

$$F1 - \text{Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\% \quad 2.4$$