

BAB III

HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Pada penelitian ini, peneliti akan menguji *algoritma Random Forest* untuk mengklasifikasikan wilayah rawan pangan di Kabupaten Berau. *Algoritma* ini akan mengklasifikasikan wilayah rawan pangan berdasarkan 7 fitur hasil dari seleksi fitur yang telah dilakukan. Data tersebut penulis dapatkan dari Peta Ketahanan dan Kerentanan Pangan atau *Food Security and Vulnerability Atlas (FSVA)* Kabupaten Berau 2023 yang disusun oleh Dinas Pangan Kabupaten Berau. Jumlah unit analisis yang akan digunakan mencakup 110 desa yang tersebar di 13 kecamatan.

3.1.1 Seleksi Data

Pada Tabel 3.1 menunjukkan atribut yang telah dipilih secara manual terdapat 7 kolom yang terpilih dan 1 atribut PRIO KOMP sebagai target atau kelas.

Tabel 3. 1 Seleksi Data

NO	Nama Kec	Kode Kec	Kode Desa	Nama Desa	Rasio Lahan	Rasio Sarana	Rasio Pddk Tidak Sejahtera	Akses Jalan	Rasio Tanpa Air Bersih	Rasio Pddk per Tenkes per Density	INDEKS KOM	PERINGKAT	PRIO KOMP
1	Kelay	6403012	6403012001	Merabu	2	6	1	4	1	1	49.94	91	3
2	Kelay	6403012	6403012002	Panaan	2	2	6	4	6	1	51.71	87	4
3	Kelay	6403012	6403012003	Merapun	2	1	6	4	6	4	64.01	40	4
4	Kelay	6405010	6405010006	Muara Lesan	2	3	2	4	3	1	42.73	101	3
5	Kelay	6405010	6405010005	Merasa	5	3	1	4	1	1	40.03	103	3
...
106	Biatan	6405110	6405110005	Biatan Lempake	3	5	5	4	4	6	69.67	15	5
107	Biatan	6405110	6405110004	Manunggal Jaya	2	5	2	4	1	6	56.89	68	4
108	Biatan	6405110	6405110001	Biatan Bapinang	2	5	1	4	4	3	52.00	85	4
109	Biatan	6405110	6405110002	Biatan Baru	2	3	4	4	3	6	62.76	45	4
110	Biatan	6405110	6405110003	Bukit Makmur Jaya	5	1	3	4	1	6	60.21	55	4

3.1.2 Pembersihan Data

Pada Tabel 3.2 menunjukkan tampilan dataset sebelum pembersihan data, yang melibatkan penghapusan baris dengan nilai yang hilang dari dataset rawan pangan, proses ini memastikan bahwa data bersih dan siap untuk dianalisis.

Tabel 3. 2 Dataset sebelum pembersihan data

NO	Nama Kec	Kode Kec	Kode Desa	Nama Desa	Rasio Lahan	Rasio Sarana	Rasio Pddk Tidak Sejahtera	Akses Jalan	Rasio Tanpa Air Bersih	Rasio Pddk per Tenkes per Density	INDEKS KOM	PERINGKAT	PRIO KOMP
1	Kelay	6403012	6403012001	Merabu	2	6	1	4	1	1	49.94	91	3

2	Kelay	6403012	6403012002	Panaan	2	2	6	4	6	1	51.71	87	4
3	Kelay	6403012	6403012003	Merapun	2	1	6	4	6	4	64.01	40	4
4	Kelay	6405010	6405010006	Muara Lesan	2	3	2	4	3	1	42.73	101	3
5	Kelay	6405010	6405010005	Merasa	5	3	1	4	1	1	40.03	103	3
6	Kelay	6405010	6405010007	Lesan Dayak	2	6	6	4	2	1	53.86	77	4
...
105	Biatan	6405110	6405110007	karangan	4	6	1	4	1	1	49.16	95	3
106	Biatan	6405110	6405110005	Biatan Lempake	3	5	5	4	4	6	69.67	15	5
107	Biatan	6405110	6405110004	Manunggal Jaya	2	5	2	4	1	6	56.89	68	4
108	Biatan	6405110	6405110001	Biatan Bapinang	2	5	1	4	4	3	52.00	85	4
109	Biatan	6405110	6405110002	Biatan Baru	2	3	4	4	3	6	62.76	45	4
110	Biatan	6405110	6405110003	Bukit Makmur Jaya	5	1	3	4	1	6	60.21	55	4

Sebelum pembersihan data, pengecekan dilakukan terlebih dahulu dimana proses ini akan menggunakan bahasa pemrograman *Python* serta *library* *Pandas*, dengan fungsi `isna()` untuk mendeteksi nilai yang hilang di mana entri adalah *True* jika entri asli adalah nilai yang hilang (NaN, None, atau NaT), dan *False* jika tidak.

```
#Melakukan pengecekan apakah terdapat data kosong
dc.isna().values.any()
```

False

Gambar 3. 1 Pengecekan Data Hilang

Setelah dilakukan pengecekan data hasil pengecekan mendapat hasil *False* dapat diartikan dataset tidak memiliki nilai kosong, setelah proses ini data yang berjumlah 110 data dilanjutkan ke tahap berikutnya.

3.1.3 Pembagian data

Pembagian data adalah tahap selanjutnya setelah pemilihan data dan pembersihan data. Data *training* dan data *testing* adalah dua subset utama tempat data akan dianalisis. Data *testing* digunakan untuk mengevaluasi kinerja model, sedangkan data *training* digunakan untuk membangun model.

```
# Mendefinisikan fitur dan variabel target

data_feature = cleanData.drop(columns=['Nama Desa', 'PRIO KOMP'])
data_class = cleanData['PRIO KOMP']
```

Gambar 3. 2 Pembagian Fitur dan Variabel Target

Pembagian data dimulai dengan mendefinisikan fitur dan variabel target. Fitur adalah atribut-atribut yang digunakan untuk membangun model prediksi, sedangkan variabel target adalah atribut yang ingin diprediksi. Pada gambar 3.2 menunjukkan keseluruhan atribut akan digunakan sebagai fitur kecuali atribut “Nama Desa” karena hanya berisi informasi identitas, dan atribut “PRIO KOMP” digunakan sebagai variabel target.

```
#Melakukan pembagian data dengan proporsi 80% data latih dan 20% data uji
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(data_feature,
                                              data_class,
                                              train_size=0.8,
                                              random_state=42)
```

Gambar 3. 3 Pembagian *Data Training* dan *Data Testing*

Setelah mengidentifikasi fitur dan variabel target, dilakukan proses pembagian data menggunakan *split validation* diimplementasikan melalui *library sklearn.model_selection* dengan fungsi *train_test_split* pada *python*. Pada penelitian ini, data dibagi menjadi 80% data *training* dan 20% data *testing*.

3.1.4 Pemodelan *Algoritma Random Forest*

Pada tahap ini, akan melakukan analisa pemodelan *algoritma random forest* dengan menggunakan *Python*. Pemodelan dianalisa berdasarkan hasil pengujian *random forest* berdasarkan data *training*, data *testing* dan Analisa pengaruh fitur pada pemodelan yang dibangun. Tujuan dari tahap ini adalah untuk menguji kinerja model *random forest* dalam mengklasifikasikan data.

a) **Klasifikasi Data *Training***

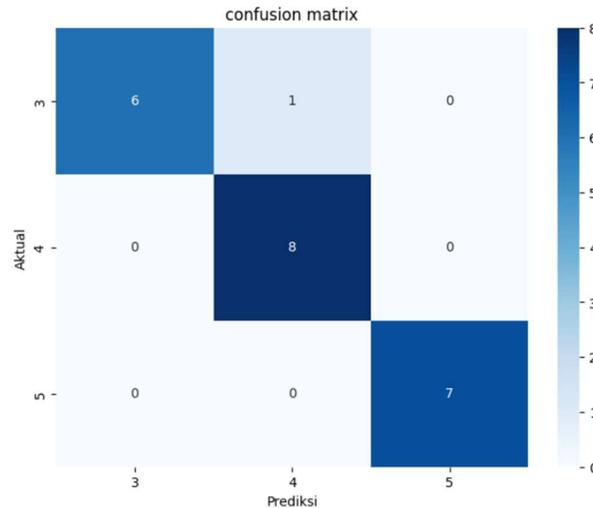
Hasil pengujian pada Tabel 3.3 menunjukkan bahwa model Random Forest melakukan klasifikasi dengan baik dalam memprediksi kelas pada dataset wilayah rawan pangan. Model ini melakukan klasifikasi pada 85 data *training* dengan rincian sebagai berikut: kelas 3 berhasil diprediksi benar sebanyak 15 kali, kelas 4 berhasil diprediksi benar sebanyak 44 kali, dan kelas 5 berhasil diprediksi benar sebanyak 26 kali. Dengan hasil klasifikasi pada data training yang baik ini, dapat disimpulkan bahwa model *Random Forest* yang dibangun mampu melakukan klasifikasi data dengan tingkat kesalahan yang rendah.

Tabel 3. 3 *Confusion Matrix*

		Class Prediksi		
		3	4	5
Nilai Aktual	3	15	0	0
	4	0	44	0
	5	0	0	26

b) **Pengujian Data *Testing***

Pengujian data *testing* dilakukan untuk menganalisis performa dan efektivitas model random forest yang dibangun dalam mengklasifikasikan data pada dataset yang digunakan. Untuk hasil klasifikasi terbagi ke dalam tiga kelas yaitu kelas 3(Agak Rentan Pangan), 4(Agak Tahan Pangan), dan 5(Tahan Pangan). Metode Evaluasi yang digunakan meliputi pengukuran akurasi, presisi, recall, dan F1-score, serta mengevaluasi kemampuan model dalam memprediksi dengan tepat antara kelas yang berbeda.



Gambar 3. 4 Tampilan Grafik *Confusion Matrix*

Pada grafik evaluasi menggunakan metode *confusion matrix* pada gambar 3.11, model berhasil melakukan klasifikasi dengan baik pada 22 data test dengan rincian kelas 5 diprediksi 7 kali benar, kelas 4 diprediksi 8 kali benar, dan kelas 3 diprediksi 6 kali benar akan tetapi untuk kelas 3 menunjukkan satu kesalahan klasifikasi. Secara keseluruhan model yang dibangun dapat mengklasifikasikan dengan cukup baik namun dengan kesalahan kecil. Untuk perbandingan hasil evaluasi seperti *Accuracy*, *presisi*, *recall* dan *F1-Score* dapat dijelaskan pada tabel 3.4.

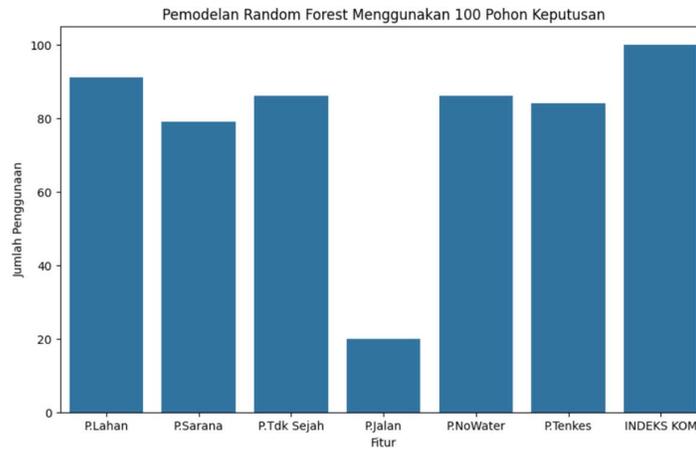
Tabel 3. 4 Perbandingan Hasil Evaluasi *Confusion Matrix*

Class	<i>Accuracy</i>	<i>Presisi</i>	<i>Recall</i>	<i>F1-Score</i>
3	0.95	1.0	0.86	0.92
4	0.95	0.89	1.0	0.94
5	0.95	1.0	1.0	1.0
Rata-rata	0.95	0.96	0.95	0.95

Pada tabel 3.4 menjelaskan perbandingan hasil evaluasi berdasarkan kelas klasifikasi, yang mendapatkan hasil yang sangat baik dengan rata-rata nilai *Accuracy* 0.95 atau 95 %, *presisi* 0.96 atau 96%, *recall* 0.95 atau 95% dan nilai *F1-Score* 0.95 atau 95%.

c) Analisis Pengaruh Fitur

Penggunaan fitur dalam pemodelan *Random Forest* dievaluasi berdasarkan seberapa sering setiap fitur digunakan oleh pohon keputusan dalam model untuk klasifikasi. Fitur yang sering digunakan oleh pohon keputusan dalam model mencerminkan relevansinya dalam membantu kemampuan model untuk menghasilkan prediksi yang benar.



Gambar 3. 5 Penggunaan Fitur

Dari grafik penggunaan fitur INDEKS KOM dengan nilai 100 dapat diartikan fitur digunakan oleh semua pohon dalam model, menunjukkan bahwa fitur ini adalah yang paling krusial dalam menentukan klasifikasi wilayah rawan pangan. Fitur lainnya seperti P.Lahan (91), P.Tdk Sejah (86), P.NoWater (86), dan P.Tenkes (84) juga menunjukkan penggunaan yang tinggi oleh banyak pohon, menandakan kontribusi signifikan mereka dalam pemodelan. Sebaliknya, fitur P.Jalan dengan nilai 20 digunakan oleh lebih sedikit pohon, menunjukkan pengaruh yang lebih rendah dalam proses klasifikasi. Pemilihan fitur yang digunakan secara luas oleh pohon-pohon dalam model *Random Forest* sangat penting untuk mencapai performa yang tinggi dan akurasi yang baik dalam klasifikasi dataset wilayah rawan pangan.

3.2 Pembahasan

Penelitian ini menggunakan data Wilayah rawan pangan Kabupaten Berau periode tahun 2023. Data yang didapat melalui beberapa tahapan pengolahan mulai dari seleksi data, pembersihan data, dan pembagian data.

- a) Tahapan Pre-processing dimulai melakukan seleksi data untuk memastikan hanya fitur-fitur yang relevan yang akan digunakan dalam pemodelan. Fitur-fitur yang dipilih mencakup berbagai indikator seperti P.Lahan, P.Sarana, P.Tdk Sejah, P.Jalan, P.NoWater, P.Tenkes, dan INDEKS KOM, yang diyakini memiliki pengaruh signifikan terhadap klasifikasi wilayah rawan pangan. Pembagian Data menggunakan *split validation*, data dibagi menjadi data latih dan data uji dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian.
- b) Penggunaan model *Random Forest* memberikan hasil yang sangat baik dengan nilai akurasi sebesar 0.95 atau 95%. Ini menunjukkan bahwa model dapat mengklasifikasikan dengan benar 95% dari data yang diuji. Selain itu, nilai presisi model adalah 96%, yang berarti bahwa 96% dari prediksi positif model adalah benar. Nilai recall sebesar 95% menunjukkan bahwa model dapat mengidentifikasi 95% dari semua kasus positif yang sebenarnya. Nilai F1-Score, yang merupakan harmoni antara presisi dan recall, adalah 0.95 atau 95%, menunjukkan keseimbangan yang baik antara kedua metrik. Analisis pengaruh fitur dilakukan pada pemodelan RF menggunakan 100 pohon keputusan menunjukkan bahwa fitur INDEKS KOM adalah yang paling relevan, digunakan oleh semua pohon keputusan dalam model. Fitur-fitur lain seperti P.Lahan, P.Tdk Sejah, P.NoWater, dan P.Tenkes juga menunjukkan pengaruh signifikan. Sebaliknya, fitur P.Jalan memiliki pengaruh yang lebih rendah tetapi tetap berkontribusi dalam model.