

**IMPLEMENTASI *ALGORITMA RANDOM FOREST* DALAM
MENGKLASIFIKASIAN WILAYAH RAWAN PANGAN KABUPATEN
BERAU**

SKRIPSI

Diajukan oleh:

Bobi Zinaidin Zidan

2011102441155



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
JULI 2024**

**IMPLEMENTASI *ALGORITMA RANDOM FOREST* DALAM
MENGKLASIFIKASIAN WILAYAH RAWAN PANGAN KABUPATEN
BERAU**

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar
Sarjana Komputer Fakultas Sain dan Teknologi Universitas
Muhammadiyah Kalimantan Timur

Diajukan oleh:

Bobi Zinaidin Zidan

2011102441155



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
JULI 2024**

LEMBAR PERSETUJUAN

**IMPLEMENTASI *ALGORITMA RANDOM FOREST* DALAM
MENGKLASIFIKASIAN WILAYAH RAWAN PANGAN KABUPATEN
BERAU**

SKRIPSI

Diajukan oleh:

Bobi Zinaidin Zidan

2011102441155

Disetujui untuk diujikan

Pada tanggal 28 Juni 2024

Pembimbing



Abdul Rahim, S.Kom., M. Cs

NIDN. 0009047901

Mengetahui

Koordinator Tugas Akhir



Abdul Rahim, S.Kom., M. Cs

NIDN. 0009047901

LEMBAR PENGESAHAN

**IMPLEMENTASI *ALGORITMA RANDOM FOREST* DALAM
MENGKLASIFIKASIAN WILAYAH RAWAN PANGAN KABUPATEN
BERAU**

SKRIPSI

Diajukan oleh:

Bobi Zinaidin Zidan

2011102441155

Diseminarkan dan Diujikan

Pada tanggal 17 Juli 2024

Penguji I	Penguji II
 <u>Wawan Joko Pranoto, S.Kom, M.Ti</u> NIDN. 1102057701	 <u>Abdul Rahim, S.Kom., M. Cs</u> NIDN. 0009047901

Mengetahui,

Ketua

Program Studi Teknik Informatika



Arbansyah, S.Kom., M. TI
NIDN. 1118019203

PERNYATAAN KEASLIAN PENULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Bobi Zinaidin Zidan
NIM : 2011102441155
Program Studi : Teknik Informatika
Judul Penelitian : Implementasi *Algoritma Random Forest* Dalam Mengklasifikasian Wilayah Rawan Pangan Kabupaten Berau

Menyatakan bahwa skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, dan bukan merupakan hasil plagiasi/falsifikasi/fabrikasi baik Sebagian atau seluruhnya.

Atas pernyataan ini, saya siap menanggung risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya saya ini.

Samarinda, 22 Juli 2024

Yang membuat pernyataan



Bobi Zinaidin Zidan

NIM: 2011102441155

ABSTRAK

Kerawanan pangan terjadi ketika wilayah atau individu tidak dapat memenuhi kebutuhan pangan minimum untuk hidup berkelanjutan. Pemerintah memantau data dan informasi pangan melalui Food Security and Vulnerability Atlas (FSVA), yang disusun oleh Badan Ketahanan Pangan (BKP) dan Program Pangan Dunia (WFP) sejak 2002. Pada tahun 2019, FSVA disusun di tingkat nasional, provinsi, dan kabupaten, mencakup analisis kabupaten, kecamatan, dan desa. Dinas Pangan Kabupaten Berau telah menyusun peta FSVA untuk melaporkan indikator rawan pangan. Namun, proses ini memerlukan waktu lama, sehingga dibutuhkan teknik machine learning untuk mempercepat pengelompokan wilayah rawan pangan menggunakan algoritma Random Forest. Penelitian ini menggunakan data dari 110 desa di Kabupaten Berau dengan 14 kolom data. Analisis dilakukan dengan Python dan library scikit-learn. Model dievaluasi menggunakan Confusion Matrix dengan hasil akurasi 95%, presisi 96%, recall 95%, dan F1-Score 95% dengan pembagian data latih dan data uji sebesar 80:20.

Kata Kunci: Rawan Pangan, FSVA, *Algoritma Random Forest*, Klasifikasi

ABSTRACT

Food insecurity occurs when a region or individual is unable to meet the minimum food needs for sustainable living. The government monitors food data and information through the Food Security and Vulnerability Atlas (FSVA), which has been compiled by the Food Security Agency (BKP) and the World Food Program (WFP) since 2002. In 2019, the FSVA was compiled at the national, provincial, and district levels, covering district, sub-district, and village analyses. The Berau Regency Food Service has compiled an FSVA map to report food insecurity indicators. However, this process takes a long time, so machine learning techniques are needed to accelerate the grouping of food-prone areas using the Random Forest algorithm. This study uses data from 110 villages in Berau Regency with 14 data columns. The analysis was done with Python and the scikit-learn library. The model was evaluated using the Confusion Matrix with 95% accuracy, 96% precision, 95% recall, and 95% F1-Score with 80:20 division of training data and testing data.

Keywords: *Food Vulnerable, FSVA, Random Forest Algorithm, Classification*

PRAKATA

Puji syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi ini yang berjudul "Implementasi Algoritma Random Forest dalam Mengklasifikasi Wilayah Rawan Pangan Kabupaten Berau". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Program Studi Teknik Informatika Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur.

Penulis menyadari bahwa skripsi ini tidak akan terselesaikan tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Sugiantoro, Ibu Chusnul Khotimah, dan kakak tercinta yang selalu memberikan doa, dukungan moral, dan materiil tanpa henti.
2. Bapak Abdul Rahim, S.Kom., M. Cs., selaku dosen pembimbing yang telah memberikan bimbingan, masukan, dan dorongan yang berharga selama proses penyusunan skripsi ini.
3. Bapak. Wawan Joko Pranoto, S.Kom, M.Ti., selaku dosen penguji yang juga memberikan arahan dan dukungan dalam penelitian ini.
4. Seluruh Bapak dan Ibu dosen Prodi Teknik Informatika, yang telah memberikan ilmu dan wawasan yang sangat berguna selama masa perkuliahan.
5. Serta rekan-rekan mahasiswa Fakultas Sains dan Teknologi, khususnya teman-teman di Prodi Teknik Informatika, yang telah memberikan dukungan, kebersamaan, dan semangat selama proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan skripsi ini di masa yang akan datang. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca dan semua pihak yang berkepentingan, serta dapat memberikan kontribusi positif dalam bidang ketahanan pangan dan teknologi informasi.

Samarinda, 22 Juli 2024

Penulis



Bobo Zinaidin Zidan
NIM: 2011102441155

DAFTAR ISI

	Halaman
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN PENULISAN.....	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
PRAKATA	viii
DAFTAR ISI	ix
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Masalah	2
BAB II METODE PENELITIAN	4
2.1 Objek Penelitian	4
2.2 Alat dan Bahan	4
2.3 Prosedur Penelitian.....	4
2.3.1 Pengumpulan data	5
2.3.2 <i>Data Pre-Processing</i>	6
2.3.3 Pembagian Data.....	6
2.3.4 <i>Pemodelan Algoritma Random Forest</i>	7
2.3.5 Evaluasi	7
BAB III HASIL DAN PEMBAHASAN	9
3.1 Hasil Penelitian.....	9
3.1.1 Seleksi Data	9
3.1.2 Pembersihan Data	9
3.1.3 Pembagian data.....	10
3.1.4 <i>Pemodelan Algoritma Random Forest</i>	11
3.2 Pembahasan	13
BAB IV PENUTUP	14
4.1 Kesimpulan.....	14
4.2 Implikasi	14
DAFTAR RUJUKAN	15
RIWAYAT HIDUP	17

DAFTAR TABEL

Tabel	Halaman
Tabel 2. 1 Keterangan Atribut Data FSVA (<i>Food Security and Vulnerability Atlas</i>)	5
Tabel 2. 2 Target Rawan Pangan.....	5
Tabel 2. 3 Seleksi Data.....	6
Tabel 2. 4 Metode <i>Confusion Matrix</i> 3x3.....	7
Tabel 3. 1 Seleksi Data.....	9
Tabel 3. 2 Dataset sebelum pembersihan data.....	9
Tabel 3. 3 <i>Confusion Matrix</i>	11
Tabel 3. 4 Perbandingan Hasil Evaluasi <i>Confusion Matrix</i>	12

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2. 1 Prosedur Penelitian	4
Gambar 2. 2 Struktur <i>Random Forest</i>	7
Gambar 3. 1 Pengecekan Data Hilang	10
Gambar 3. 2 Pembagian Fitur dan Variabel Target	10
Gambar 3. 3 Pembagian Data <i>Training</i> dan Data <i>Testing</i>	11
Gambar 3. 4 Tampilan Grafik <i>Confusion Matrix</i>	12
Gambar 3. 5 Penggunaan Fitur	13

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 Surat Pengantar Pengambilan Data.....	18
Lampiran 2 Surat Balasan Instansi.....	19
Lampiran 3 Tampilan Dataset <i>Food Security and Vulnerability Atlas</i> (FSVA) Kabupaten Berau.....	20
Lampiran 4 Kartu Bimbingan.....	21
Lampiran 5 Seluruh Kode Python.....	23

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Wilayah rawan pangan merupakan wilayah yang memiliki karakteristik yang dapat menyebabkan potensi terjadinya kerawanan pangan. Kerawanan pangan terjadi ketika negara atau individu tidak mampu memenuhi kebutuhan pangan minimum untuk hidup berkelanjutan dalam kondisi ekosistem setempat.

Kondisi kerawanan pangan dibagi menjadi dua faktor: kronis dan transien. Faktor kronis mencakup elemen-elemen struktural yang tidak berubah dengan cepat, seperti iklim lokal, jenis tanah, sistem pemerintahan daerah, infrastruktur publik, sistem kepemilikan lahan, distribusi pendapatan dan mata pencaharian, hubungan antar suku, tingkat pendidikan, serta aspek sosial budaya dan adat istiadat. Sebaliknya, faktor transien mencakup elemen-elemen dinamis yang dapat berubah dengan cepat, seperti penyakit menular, bencana alam, pengungsian, perubahan fungsi pasar, tingkat hutang, dan migrasi. (Dinas Pangan Kabupaten Berau, 2023). Kondisi di wilayah rawan pangan dipengaruhi oleh faktor kronis maupun transien, sehingga diperlukan pengawasan dari pemerintah agar wilayah tersebut tidak mengalami kerawanan pangan. Pemerintah melakukan inovasi dalam pengawasan ini dengan menyediakan Food Security and Vulnerability Atlas (FSVA), yang berfungsi sebagai alat untuk menyediakan data dan informasi di bidang pangan.

Food Security and Vulnerability Atlas (FSVA) disusun sebagai instrumen untuk memantau dan menganalisis ketahanan pangan wilayah. FSVA merupakan hasil kolaborasi antara Badan Ketahanan Pangan (BKP) dan Program Pangan Dunia (*World Food Programme - WFP*) yang dimulai sejak tahun 2002. Kolaborasi ini menghasilkan Peta Kerawanan Pangan (*Food Insecurity Atlas - FIA*) pada tahun 2005. Kemudian, pada tahun 2009, diluncurkan Peta Ketahanan Pangan yang lebih dikenal sebagai *Food Security and Vulnerability Atlas* (FSVA). Pada tahun 2019, FSVA disusun di tingkat nasional, provinsi, dan kabupaten. FSVA Nasional disusun dengan unit analisis kabupaten, FSVA Provinsi dengan unit kecamatan, dan FSVA Kabupaten menggunakan unit analisis desa. (Badan Ketahanan Pangan Kementerian Pertanian, 2019). Dalam *Food Security and Vulnerability Atlas* (FSVA), wilayah yang termasuk daerah rentan rawan pangan adalah wilayah yang masuk dalam prioritas 1 hingga 3. Karakteristik wilayah rawan pangan ini dicirikan oleh tingginya rasio konsumsi per kapita terhadap ketersediaan air bersih, tingginya prevalensi balita stunting, tingginya jumlah rumah tangga tanpa akses ke air bersih, serta tingginya persentase penduduk yang hidup di bawah garis kemiskinan (Badan Pangan Nasional, 2023).

Dinas Pangan Kabupaten Berau telah menyusun peta FSVA yang memuat laporan mengenai indikator-indikator yang mempengaruhi kerawanan pangan wilayah. Sebagai alat penyedia informasi wilayah, FSVA dengan karakteristik daerah yang rentan terhadap kerawanan pangan diharapkan dapat membantu menentukan langkah yang tepat dalam menangani isu-isu ketahanan pangan yang relevan di wilayah tersebut. Namun, karena banyaknya tantangan di lapangan, proses pengumpulan data membutuhkan waktu yang cukup lama. Oleh karena itu, data yang diperoleh harus segera dianalisis dan hasilnya disampaikan oleh Dinas Pangan Kabupaten Berau.

Berikut adalah parafrasa dari kalimat tersebut: Berdasarkan permasalahan yang telah dipaparkan di atas, penting untuk melakukan analisis mendalam mengenai klasifikasi wilayah rawan pangan agar dapat meminimalisir kebijakan yang tidak sesuai terhadap wilayah-wilayah tersebut. Klasifikasi adalah proses untuk menciptakan model yang dapat menjelaskan dan membedakan konsep atau kelas data (Ikko Mulya Rizky et al., 2023). Klasifikasi digunakan untuk mempelajari berbagai fungsi yang dapat menetapkan data yang dipilih ke salah satu target kelas yang telah ditentukan (Suci Amaliah et al., 2022).

Ada banyak *algoritma* yang umum digunakan untuk klasifikasi data, seperti *Random Forest*, *K-Nearest Neighbors*, *Decision Tree*, *Naive Bayes*, *Support Vector Machines*, dan *Artificial Neural Network*. Berdasarkan penelitian (Suci Amaliah et al., 2022) mengenai klasifikasi Varian Minuman Kopi menggunakan Metode *Random Forest* menunjukkan nilai akurasi sebesar 94.12%. Kemudian mengenai penelitian yang dilakukan (Jaya Purnama & Rahayu, 2022) melakukan komparasi *Algoritma* diantaranya *Random Forest*, *Decision Tree*, *Naive Bayes* dan *Artificial Neural Network* mengenai klasifikasi konsumsi energi industri baja, bahwa algoritma *Random Forest* mendapatkan hasil yang sangat baik dengan nilai akurasi 91.13%. Pada penelitian mengenai klasifikasi kepuasan penumpang maskapai penerbangan menggunakan Algoritma *Random Forest*, *K-Nearest Neighbors*, *Decision Tree*, dan *Support Vector Machines*, Algoritma *Random Forest* mendapatkan hasil yang baik dengan nilai tingkat sebesar 96% (Setiono, 2022). Kemudian pada penelitian yang dilakukan oleh (Yoga Religia et al., 2021) melakukan klasifikasi data *Bank Marketing* menggunakan Algoritma *Random Forest* mendapatkan nilai akurasi sebesar 88,30%. Dari Penelitian terdahulu menunjukkan *algoritma Random Forest (RF)* terbukti efektif dalam melakukan klasifikasi pada berbagai topik yang berbeda, sehingga penulis tertarik untuk menerapkan metode *Random Forest* di dalam penelitian ini.

Penelitian mengenai klasifikasi wilayah rawan pangan dapat kita lihat pada penelitian terdahulu yang telah dilakukan oleh (Darnila, 2023), dengan judul “KLASIFIKASI WILAYAH RAWAN PANGAN DI KAB ACEH UTARA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR”. Penelitian ini melakukan pengklasifikasian data regional FSVA (*Food Security and Vulnerability Atlas*) Kabupaten Aceh Utara. Dalam penelitian tersebut, data FSVA dikumpulkan dan dikelompokkan ke dalam 6 prioritas, yaitu prioritas 1, prioritas 2, prioritas 3, prioritas 4, prioritas 5, dan prioritas 6. Setelah itu, data tersebut dibagi menjadi 70% untuk data *training* dan 30% untuk data *testing*. Untuk mengklasifikasikan data FSVA, *algoritma KNN* diterapkan dengan metode *Euclidean Distance*. Hasil klasifikasi menggunakan *algoritma KNN* menunjukkan prioritas 1 yang terdiri dari 7 desa, prioritas 2 yang terdiri dari 24 desa, dan prioritas 3 yang terdiri dari 4 desa. Penelitian ini mencapai tingkat akurasi 86%.

Berdasarkan penelitian sebelumnya tentang klasifikasi wilayah rawan pangan yang telah dilakukan menggunakan metode *K-Nearest Neighbor* peneliti tertarik melakukan penelitian menggunakan pendekatan yang berbeda menggunakan metode *Random Forest*, maka peneliti melakukan penelitian dengan judul “Implementasi *Algoritma Random Forest* dalam Mengklasifikasikan Wilayah Rawan Pangan Kabupaten Berau”.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah yang didapatkan oleh peneliti adalah (i) Bagaimana menerapkan *algoritma random forest* dalam mengklasifikasikan wilayah rawan pangan Kabupaten Berau. (ii) Bagaimana tingkat akurasi dari model *algoritma Random Forest* yang dibangun dalam klasifikasi wilayah rawan pangan di Kabupaten Berau.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah (i) Mengimplementasikan *algoritma random forest* dalam mengklasifikasikan wilayah rawan pangan Kabupaten Berau. (ii) Mengevaluasi kinerja hasil pemodelan *algoritma random forest* menggunakan metode *confusion matrix* untuk mendapatkan nilai akurasi, *presisi*, *recall* dan *F1-Score*.

1.4 Manfaat Masalah

Harapannya penelitian ini dapat membawa manfaat dan pengetahuan bagi beberapa pihak, khususnya (i) Menambah wawasan dalam penerapan metode klasifikasi menggunakan *algoritma*

Random Forest pada dataset wilayah rawan pangan di Kabupaten Berau. (ii) Penelitian ini diharapkan dapat memberi manfaat bagi peneliti lain yang tertarik dalam bidang data mining. Melalui penelitian ini diharapkan mendapatkan pemahaman mengenai klasifikasi wilayah rawan pangan menggunakan *algoritma Random Forest* dan dapat dijadikan referensi bagi penelitian selanjutnya. (iii) Diharapkan hasil penelitian ini dapat membantu pemerintah terkait mengambil kebijakan yang tepat dalam menangani isu-isu ketahanan pangan yang relevan di wilayahnya.

BAB II

METODE PENELITIAN

2.1 Objek Penelitian

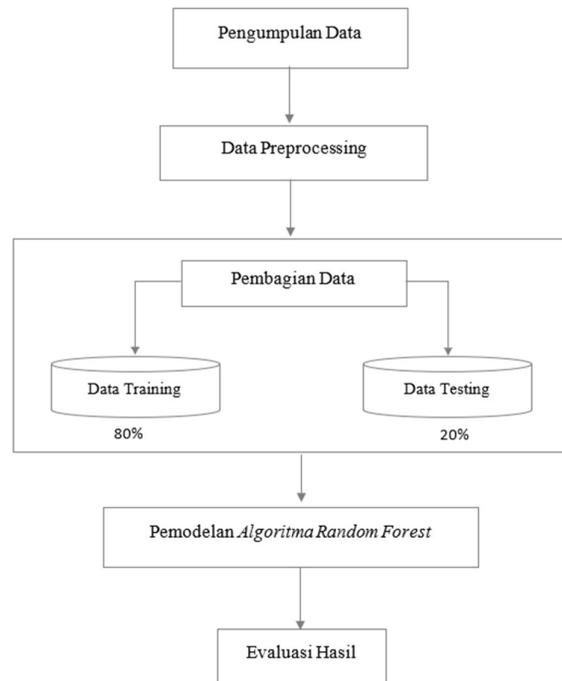
Penelitian ini dilaksanakan di Dinas Pangan Kab. Berau yang berlokasi di Jl. Dr. Murjani I No.80, Karang Ambun, Kec. Tanjung Redeb, Kabupaten Berau, Kalimantan Timur dari tanggal 1 Maret 2024 sd tanggal 30 Mei 2024 dengan menggunakan data *Food Security and Vulnerability Atlas* (FSVA) Kab. Berau tahun 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

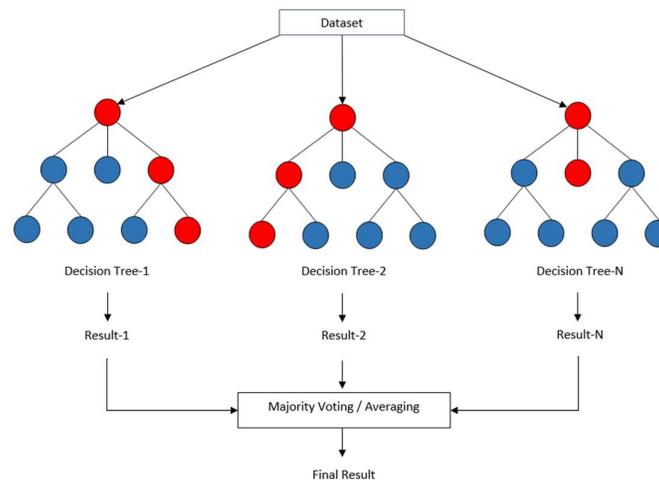
Proses pembagian data dilakukan sebelum melakukan pemodelan *algoritma random forest*, penelitian ini menggunakan teknik *split validation*. *Split validation* adalah teknik validasi yang secara acak membagi data dengan perbandingan antara data *training* dan *testing* dengan rasio 0,9 sampai

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).

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

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

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

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

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 Densiv	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 Densiv	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***

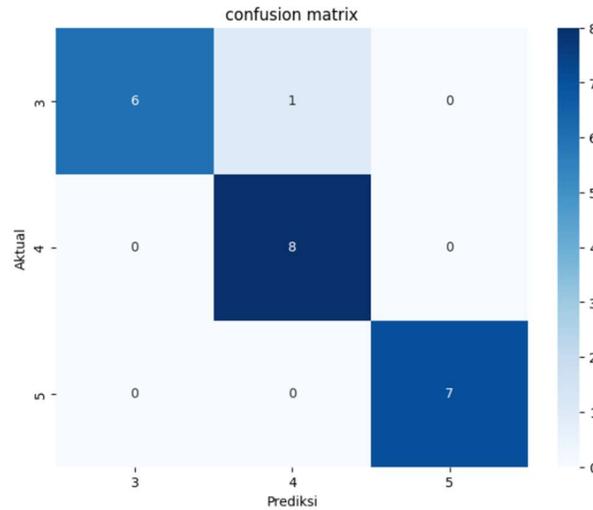
Hasil pengujian pada Tabel 3.3 menunjukkan bahwa model Random Forest melakukan klasifikasi dengan baik dalam memprediksi kelas pada dataset wilayah rawan pangan. Model ini melakukan klasifikasi pada 85 data *training* dengan rincian sebagai berikut: kelas 3 berhasil diprediksi benar sebanyak 15 kali, 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.

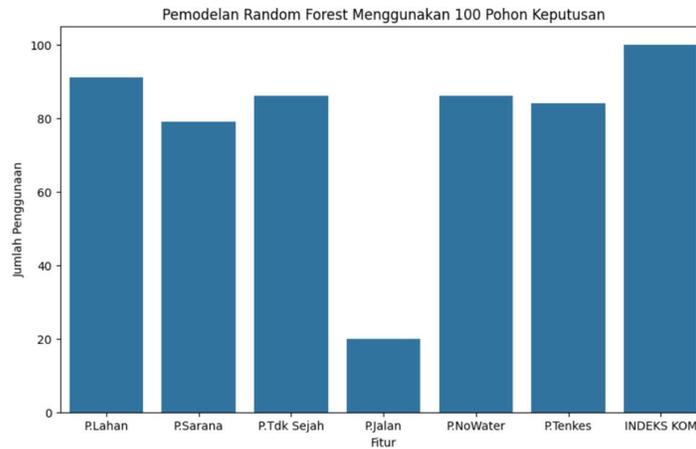
Tabel 3. 4 Perbandingan Hasil Evaluasi *Confusion Matrix*

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.

BAB IV

PENUTUP

4.1 Kesimpulan

Kesimpulan mengacu pada tujuan penelitian yang telah disusun sebelumnya, berhasil atau tidaknya dalam mencapai tujuan tersebut dibuktikan pada tahapan implementasi model dan evaluasi model. Kesimpulan yang dapat ditarik pada penelitian ini adalah:

1. Permasalahan Dinas Pangan Kabupaten Berau dalam mengidentifikasi daerah rawan pangan dapat diselesaikan dengan metode klasifikasi menggunakan *Algoritma Random Forest* dengan nilai rata-rata akurasi 95%.
2. Hasil dari evaluasi model *Random Forest* menggunakan metode *confusion matrix* menghasilkan nilai rata-rata *Accuracy* 95 %, *Presisi* 96%, *Recall* 95% dan nilai *F1-Score* 95%.

4.2 Implikasi

Berdasarkan kesimpulan yang telah diuraikan, maka ada beberapa implikasi yang dapat disampaikan yaitu:

1. Penelitian ini diharapkan memberikan kontribusi dalam pencegahan wilayah rawan pangan dengan menggabungkan teknologi *Machine Learning* dan analisis dataset *Food Security and Vulnerability Atlas* (FSVA) Kabupaten.
2. Pada penelitian yang akan datang, dapat juga menambahkan *algoritma machine learning* yang lain sebagai perbandingan atau komparasi *algoritma* sehingga akan menghasilkan hasil yang lebih variatif dan tentunya bernilai informasi yang sangat tinggi.
3. Hasil implementasi *algoritma Random Forest* dalam mengklasifikasikan wilayah rawan pangan dapat membantu pemerintah daerah dalam mengidentifikasi daerah-daerah yang memerlukan perhatian khusus. Hasil klasifikasi dapat digunakan untuk merencanakan dan mengalokasikan sumber daya dengan lebih efisien, serta menyusun kebijakan yang lebih tepat sasaran untuk mengurangi kerawanan pangan.

DAFTAR RUJUKAN

- Adnin Kamila, S., Sri Sulistijowati, R., & Susanto, I. (2023). Klasifikasi Penyakit Jantung Menggunakan Decision Tree dan Random Forest. *Universitas Nusantara PGRI Kediri. Kediri*, 2, 7–12.
- Badan Ketahanan Pangan Kementerian Pertanian. (2019). *Peta Ketahanan dan Kerentanan Pangan (FSVA) Tahun 2019* (FSVA 2019, Vol. 2019).
<https://badanpangan.go.id/storage/app/media/Pusat%20Ketersediaan/FSVA%202019%20FINAL.pdf>
- Badan Pangan Nasional. (2023). *Peta Ketahanan dan Kerentanan Pangan (FSVA) Tahun 2023* (FSVA 2023, Vol. 2023). <https://badanpangan.go.id/buku-digital>
- Baiq Nurul Azmi, Arief Hermawan, & Donny Avianto. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290.
<https://doi.org/10.35746/jtim.v4i4.298>
- Breiman, L. (2001). *Random Forests* (R. E & S. Schapire, Eds.; Vol. 45, pp. 5–32). Kluwer Academic Publishers. Manufactured in The Netherlands.
<https://link.springer.com/article/10.1023/a:1010933404324>
- Chairunisa, R., Adiwijaya, A., & Astuti, W. (2020). Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 805–812.
- Darnila, E. (2023). KLASIFIKASI WILAYAH RAWAN PANGAN DI KAB ACEH UTARA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal TEKINKOM*, 6(2), 824–834. <https://doi.org/10.37600/tekinkom.v6i2.1075>
- Dinas Pangan Kabupaten Berau. (2023). *FSVA BERAU FSVA BERAU (FOOD SECURITY AND VULNERABILITY ATLAS-FSVA) KABUPATEN BERAU TAHUN 2023* (Vol. 2020).
- Fahmy Amin, M. (2023). Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 7(1), 0–0.
<https://doi.org/10.21608/erjeng.2023.296718>
- Ikko Mulya Rizky, I., Yusuf Irianto, S., & Sriyanto, S. (2023). Perbandingan Kinerja Algoritma Naive Bayes, Support Vector Machine dan Random forest untuk Prediksi Penyakit Ginjal Kronis. *Seminar Nasional Hasil Penelitian Dan Pengabdian Masyarakat*, 1, 139–151.
<https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/3832>
- Jaya Purnama, J., & Rahayu, S. (2022). KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING. *JURNAL TEKNOINFO*, 16(2), 395–407.
<https://ejournal.teknokrat.ac.id/index.php/teknoinfo/index>
- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, & Susanti. (2022). Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest. *SATIN - Sains Dan Teknologi Informasi*, 8(2), 122–131.
<https://doi.org/10.33372/stn.v8i2.885>
- Pratama, F., Hadryan Nst, Z., Khairi, Z., Rahmaddeni, & Efrizoni, L. (2024). PERBANDINGAN ALGORITMA RANDOM FOREST DAN K-NEAREST NEIGHBOR DALAM KLASIFIKASI KESEHATAN MENTAL MAHASISWA. *JURNAL ILMIAH BETRIK : BESEMAH*

TEKNOLOGI INFORMASI DAN KOMPUTER, 15(1), 31–37.
<https://ejournal.pppmitpa.or.id/index.php/betrik/article/view/224/>

- Setiono, M. H. (2022). KOMPARASI ALGORITMA DECISION TREE, RANDOM FOREST, SVM DAN K-NN DALAM KLASIFIKASI KEPUASAN PENUMPANG MASKAPAI PENERBANGAN. *INTI Nusa Mandiri*, 17(1), 32–39. <https://doi.org/10.33480/inti.v17i1.3420>
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127.
<https://doi.org/10.35580/variasiunm31>
- Suci Pania, T., Hidayati, R., & Kasliono. (2024). Klasifikasi Kecanduan Bermain Game online Pada Remaja Menggunakan Metode Naïve Bayes Classifier Berbasis Website. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(5), 2518–2526. <https://doi.org/10.30865/klik.v4i5.1782>
- Supriyadi, R., Gata, W., Maulidah, N., Fauzi, A., Komputer, I., & Nusa Mandiri Jalan Margonda Raya No, S. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *JURNAL ILMIAH EKONOMI DAN BISNIS*, 13(2), 67–75.
<http://journal.stekom.ac.id/index.php/E-Bisnis/page67>
- Triyanto, D., Sholeh, M., & Hasan, F. N. (2023). Implementasi Business Intelligence Menggunakan Tableau Untuk Visualisasi Data Dampak Bencana Banjir di Indonesia. *KLIK: Kajian Ilmiah Informatika Dan Komputer Media Online*, 3(6), 586–594. <https://doi.org/10.30865/klik.v3i6.769>
- Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187–192. <https://doi.org/10.29207/resti.v5i1.2813>
- Ziaul, D., Iskandar, H., & Ramdhani, Y. (2023). Optimasi Parameter Random Forest Menggunakan Grid Search Untuk Analisis Time Series. *Fountain of Informatics Journal*, 8(2), 2548–5113.
<https://doi.org/10.21111/fij.v8i2.10532>

RIWAYAT HIDUP



Nama saya Bobi Zinaidin Zidan, lahir di Berau pada tanggal 19 Juni 2002. Saya berdomisili di Jl. Mulawarman, Kalimantan Timur, Kabupaten Berau, 77352. Saya menempuh pendidikan di SDN 004 Tepian Buah (2008-2014), SMPN 17 Merancang Ulu (2014-2017), SMKN 06 Berau (2017-2020). Penulis saat ini tercatat sebagai mahasiswa pada perguruan tinggi swasta Universitas Muhammadiyah Kalimantan Timur pada Fakultas Sains dan Teknologi di Program Studi Teknik Informatika pada tahun 2020. Saya pernah magang di Unit Kerja Pengadaan Barang/Jasa Kabupaten Berau di bagian Pengelolaan Sistem Informasi (Juli-September 2023) yang dilaksanakan di semester 7. Demikian deskripsi riwayat hidup yang dapat penulis sampaikan jika terdapat kesalahan atau kekurangan mohon dimaafkan, maka penulis mengharapkan kritik dan saran yang membangun mengenai skripsi ini yang berjudul "*Implementasi Algoritma Random Forest dalam Mengklasifikasi Wilayah Rawan Pangan Kabupaten Berau*".

Lampiran 1 Surat Pengantar Pengambilan Data



UMKT
Program Studi
Teknik Informatika
Fakultas Sains dan Teknologi

Telp. 0541-748511 Fax. 0541-766832
Website <http://Informatika.umkt.ac.id>
email: Informatika@umkt.ac.id



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Nomor : 003-016/FST.1/D.3/C/2024
Lampiran : -
Perihal : Permohonan Pengambilan Data

Kepada Yth.
Kepala Dinas Pangan Kabupaten Berau
di -

Tempat

Assalamu'alaikum Warrahmatullahi Wabarrakatuh

Puji Syukur kepada Allah Subhanahu wa ta'ala yang senantiasa melimpahkan Rahmat-Nya kepada kita sekalian. Aamiin.

Sehubungan untuk memenuhi Tugas Akhir/Skripsi Tahun Akademik 2023/2024, maka dengan ini kami bermaksud untuk melakukan pengambilan data di Dinas Pangan Kabupaten Berau. Adapun data yang diminta yaitu data kerentanan wilayah rawan pangan di Kabupaten Berau, dengan nama mahasiswa sebagai berikut:

No	Nama	NIM
1	Farhan Akbar	2011102441147
2	Rilfa Dini	2011102441236
3	Bobi Zinaidin Zidan	2011102441155
4	Reny Novita Sari	2011102441160
5	Deni Iskandar	2011102441112

Demikian surat permohonan ini dibuat. Atas perhatiannya dan kerjasamanya kami mengucapkan terima kasih.

Wassalamu'alaikum Warrahmatullahi Wabarrakatuh

Samarinda, 20 Syawal 1445 H
29 April 2024 M



Program Studi S1 Teknik Informatika

Arhamsyah, S.Kom., M.TI
NIDN. 1118019203

Kampus 1 : Jl. Ir. H. Juanda, No.15, Samarinda
Kampus 2 : Jl. Pelita, Pesona Mahakam, Samarinda

Lampiran 2 Surat Balasan Instansi



PEMERINTAH KABUPATEN BERAU
DINAS PANGAN
Jalan Dr. Murjani I No. 080 Tanjung Redeb Kabupaten Berau - 77311
Email : panganberau@gmail.com

Tanjung Redeb, 30 April 2024

Nomor : 500.1.2.3/181/DP-KKP
Sifat : BIASA
Perihal : Kesiediaan Pemberian Data

Yth. Ketua Program Studi
S1 Teknik Informatika UMKT
di -
Samarinda

Menindaklanjuti Surat Ketua Program Studi S1 Teknik Informatika Universitas Muhammadiyah Kalimantan Timur nomor 003-016/FST.1/D.3/C/2024 tanggal 29 April 2024 perihal Permohonan Pengambilan Data di Dinas Pangan Kabupaten Berau dengan data dimaksud berupa data kerentanan wilayah rawan pangan di Kabupaten Berau.

Berkenaan hal tersebut, pada prinsipnya kami bersedia menerima dan memberikan data dimaksud sesuai dengan batasan/kewenangan yang dimiliki Dinas Pangan Kab. Berau dan peraturan perundangan yang berlaku,

Demikian disampaikan, atas perhatiannya diucapkan terima kasih.

Kepala,



Rakhmadi Pasarakan, S.Hut., M.Sc

Lampiran 3 Tampilan Dataset *Food Security and Vulnerability Atlas (FSVA)* Kabupaten Berau

No.	Nama Kecamatan	Kode Kecamatan	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
7	Kelay	6405010	6405010009	Long Beliu	6	5	6	4	6	4	84.17	1	6
8	Kelay	6405010	6405010010	Long Duhung	2	6	6	3	6	1	53.28	80	4
9	Kelay	6405010	6405010013	Long Lamcin	2	6	6	3	6	1	58.36	60	4
10	Kelay	6405010	6405010011	Long Keluh	2	5	6	3	6	4	62.74	46	4
...
100	Batu Putih	6405100	6405100006	Ampen Medang	5	5	2	4	2	3	53.45	79	4
101	Batu Putih	6405100	6405100007	Balikukup	2	6	5	4	6	4	72.57	9	5
102	Batu Putih	6405100	6405100002	Sumber Agung	2	5	2	4	3	5	61.28	51	4
103	Biatan	6405110	6405110006	Biatan Ulu	2	2	1	4	3	3	38.56	106	3
104	Biatan	6405110	6405110008	Biatan Ilir	5	4	3	4	5	1	56.33	70	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

Lampiran 4 Kartu Bimbingan

KARTU KENDALI BIMBINGAN LAPORAN KARYA ILMIAH

Nama Mahasiswa : Bobi Zinaidin Zidan
 NIM : 2011102441155
 Nama Dosen Pembimbing : Abdul Rahim, S.Kom., M.Cs
 Judul Penelitian : Implemntasi Algoritma Random Forest Dalam Mengklasifikasian Wilayah Rawan Pangan di kabupaten Berau

No	Tanggal	Uraian Pembimbingan	Paraf Dosen
1	07 Februari 2024	Pengajuan RTA, pembahasan tahapan dalam penelitian, dan arahan untuk mencari permasalahan yang ingin di gunakan.	
2	27 Februari 2024	Konsultasi mengenai data yang ingin digunakan dalam penelitian	
3	21 Maret 2024	Mencari jurnal/artikel penelitian terhadap klasifikasi terkait data science untuk menentukan gap penelitian saat membuat canvas penelitian.	
4	07 Mei 2024	Bimbingan online membahas proposal penelitian, revisi proposal: 1. Menyesuaikan margin keseluruhan mengikuti pedoman. 2. Melakukan perbaikan di bagian latar belakang permasalahan pada bagian objek/lokasi penelitian belum ditemukan. 3. Melakukan perbaikan bentuk kalimat di akhir paragraf latar belakang. 4. Penambahan kalimat di bab 2 bagian objek penelitian. 5. Menambahkan Referensi internasional. 6. perbaikan pengejaan kata,	
5	21 Mei 2024	Hasil review desk: Penguji pertama: penggunaan bahasa asing perlu diperhatikan karena dibuat font italic. penguji kedua: 1. Dalam paragraf dihindari penyajian dengan poin-poin, namun harus dijadikan sebagai bagian dari kalimat utuh dalam paragraf. Contoh: Tujuan penelitian ini adalah (I) Tujuan 1, (II) Tujuan 2, dan (III) Tujuan 3 (Tidak disajikan dalam baris yang terpisah). 2. Penomoran sebagai bagian dari narasi dalam suatu paragraf dinyatakan dalam (I), (II), (III) dan seterusnya (dapat dilihat pada panduan Hal 13).	
6	05 Juni 2024	Bimbingan secara daring membahas penulisan bab 3 dan 4.	
7	18 Juni 2024	Membahas hasil penulisan bab 3 dan 4.	
8	25 Juni 2024	Membahas hasil revisi bab 3 dan 4, Menambah jurnal inter dan lokal minimal berjumlah 20 buah.	
9	26 Juni 2024	Perbaiki format tabel, gambar, dan perhatikan jumlah halaman naskah.	

10	27 Juni 2024	Perhatikan dalam pembuatan abstrak naskah.	
----	--------------	--	---

Dosen Pembimbing



Abdul Rahim, S.Kom., M. Cs
NIDN. 0009047901

Mengetahui



Arbansyah, S.Kom., M. TI
NIDN. 1118019203

Lampiran 5 Seluruh Kode Python

```
[47] from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
[48] #Menghubungkan Google Collab dengan google Drive
from google.colab import drive
drive.mount('/content/gdrive')

#Mengimport pustaka dan menampilkan dataset
import pandas as pd
df = pd.read_csv('/content/gdrive/MyDrive/Skripsi 2024/FSVA Desa 2023.csv')
df.head()
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call drive.mount("/content/gdrive", force_remount=True).

No.	Nama Kec	Kode Kec	Kode Desa	Nama Desa	P.Lahan	P.Sarana	P.Tdk Sejah	P.Jalan	P.NoWater	P.Tenkes	INDEKS KOM	PERINGKAT	PRIO KOMP	
0	1	Kelay	6403012	6403012001	Merabu	2	6	1	4	1	1	49.94	91	3
1	2	Kelay	6403012	6403012002	Panaan	2	2	6	4	6	1	51.71	87	4
2	3	Kelay	6403012	6403012003	Merapun	2	1	6	4	6	4	64.01	40	4
3	4	Kelay	6405010	6405010006	Muara Lesan	2	3	2	4	3	1	42.73	101	3
4	5	Kelay	6405010	6405010005	Merasa	5	3	1	4	1	1	40.03	103	3

```
[49] #Melakukan Pemeriksaan Tipe Data Atribut
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   No.         110 non-null    int64
 1   Nama Kec   110 non-null    object
 2   Kode Kec   110 non-null    int64
 3   Kode Desa   110 non-null    int64
 4   Nama Desa   110 non-null    object
 5   P.Lahan     110 non-null    int64
 6   P.Sarana    110 non-null    int64
 7   P.Tdk Sejah 110 non-null    int64
 8   P.Jalan     110 non-null    int64
 9   P.NoWater   110 non-null    int64
10   P.Tenkes    110 non-null    int64
11   INDEKS KOM  110 non-null    float64
12   PERINGKAT  110 non-null    int64
13   PRIO KOMP   110 non-null    int64
dtypes: float64(1), int64(11), object(2)
memory usage: 12.2+ KB
```

```
#Melakukan penghapusan atribut yang tidak digunakan
dc=df.drop(["No.", "Nama Kec", "Kode Kec", "Kode Desa", "PERINGKAT"], axis=1)
dc.head()
```

	Nama Desa	P.Lahan	P.Sarana	P.Tdk Sejah	P.Jalan	P.NoWater	P.Tenkes	INDEKS KOM	PRIO KOMP
0	Merabu	2	6	1	4	1	1	49.94	3
1	Panaan	2	2	6	4	6	1	51.71	4
2	Merapun	2	1	6	4	6	4	64.01	4
3	Muara Lesan	2	3	2	4	3	1	42.73	3
4	Merasa	5	3	1	4	1	1	40.03	3

Next steps: [Generate code with dc](#) [View recommended plots](#)

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

```
False
```

```
[52] #Melakukan pembersihan data kosong
cleanData = dc.dropna()

print(f"Jumlah Data sesudah pembersihan data kosong: {len(cleanData)}")
```

```
Jumlah Data sesudah pembersihan data kosong: 110
```

```
[53] # Mendefinisikan fitur dan variabel target

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

print(f'Dimensi feature: {data_feature.shape}')
print(f'Class: {set(data_class)}')
```

```
Dimensi feature: (110, 7)
Class: {3, 4, 5, 6}
```

```
[54] #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)
```

```

#Melakukan penerapan model algoritma Random Forest
from sklearn.ensemble import RandomForestClassifier
rf_klasifikasi = RandomForestClassifier(n_estimators=100,
                                     random_state=42)
rf_klasifikasi.fit(xtrain, ytrain)

ypred = rf_klasifikasi.predict(xtest)

# Menampilkan laporan Akurasi
print(classification_report(ytest, ypred))

```

	precision	recall	f1-score	support
3	1.00	0.86	0.92	7
4	0.89	1.00	0.94	8
5	1.00	1.00	1.00	7
accuracy			0.95	22
macro avg	0.96	0.95	0.95	22
weighted avg	0.96	0.95	0.95	22

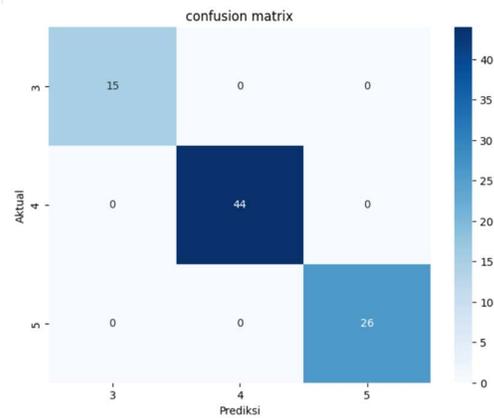
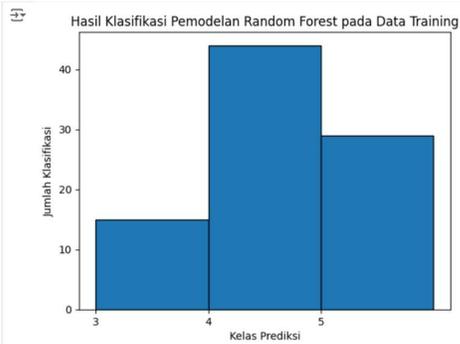
```

[56] # Menampilkan histogram kelas prediksi
pred_train = rf_klasifikasi.predict(xtrain)
plt.hist(pred_train, bins=np.unique(pred_train), edgecolor='black')
plt.xlabel('Kelas Prediksi')
plt.ylabel('Jumlah Klasifikasi')
plt.title('Hasil Klasifikasi Pemodelan Random Forest pada Data Training')
plt.xticks([3, 4, 5])
plt.show()

conf_matrix_train = confusion_matrix(ytrain, pred_train)

plt.figure(figsize=(8,6))
sns.heatmap(conf_matrix_train[[0, 1, 2], :][:, [0, 1, 2]], annot=True, fmt='d', cmap='Blues', xticklabels=[3, 4, 5], yticklabels=[3, 4, 5])
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('confusion matrix')
plt.show()

```



```

namaFitur = data.feature.columns
feature_importances1 = []
for pohon in rf.klasifikasi.estimators_:
    feature_importances1.append(pohon.feature_importances_)

# Calculate the total usage count for each feature
jumlahFitur = {}
for feature_importance in feature_importances1:
    for i, importance in enumerate(feature_importance):
        fitur_names = namaFitur[i]
        if fitur_names not in jumlahFitur:
            jumlahFitur[fitur_names] = 0
        if importance > 0:
            jumlahFitur[fitur_names] += 1

# Print the feature usage count
print("Feature Usage Count:")
for fitur, jumlah in jumlahFitur.items():
    print(f"{fitur}: {jumlah}")

```

```

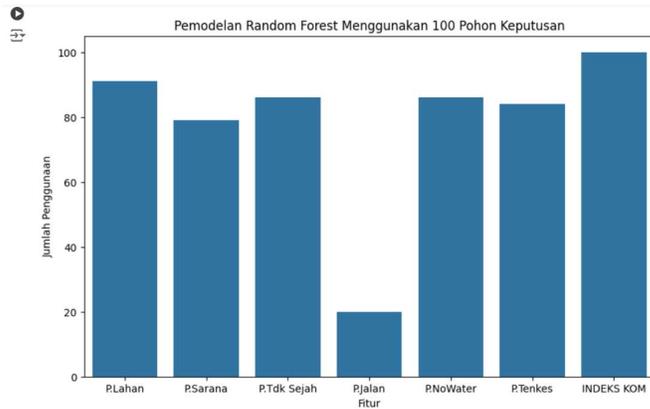
[58] Feature Usage Count:
P.Lahan: 91
P.Sarana: 79
P.Tdk Sejah: 86
P.Jalan: 20
P.NoWater: 86
P.Tenkes: 84
INDEKS KOM: 100

```

```

# Create a bar chart to display the feature usage count
plt.figure(figsize=(10, 6))
sns.barplot(x=list(jumlahFitur.keys()), y=list(jumlahFitur.values()))
plt.xlabel('Fitur')
plt.ylabel('Jumlah Penggunaan')
plt.title('Pemodelan Random Forest Menggunakan 100 Pohon Keputusan')
plt.show()

```

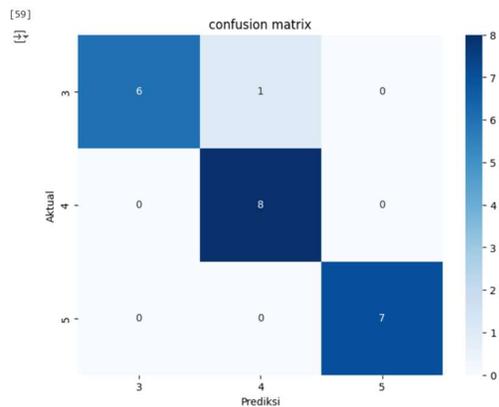


```

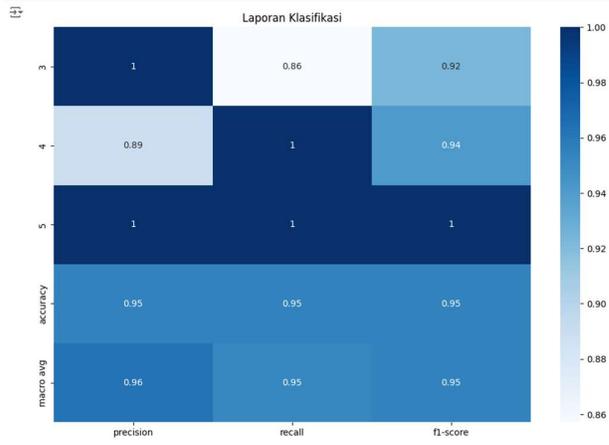
# Visualisasi Confusion Matrix
conf_matrix = confusion_matrix(ytest, ypred)

plt.figure(figsize=(8,6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=[3, 4, 5], yticklabels=[3, 4, 5])
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('confusion matrix')
plt.show()

```

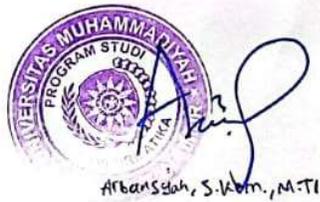


```
# Visualisasi Laporan Klasifikasi
class_report = classification_report(ytest, ypred, output_dict=True)
report_df = pd.DataFrame(class_report).transpose()
plt.figure(figsize=(12,8))
sns.heatmap(report_df.iloc[:-1, :-1], annot=True, cmap='Blues')
plt.title('Laporan Klasifikasi')
plt.show()
```



SKRIPSI BOBI ZINAIDIN ZIDAN

by Teknik Informatika Universitas Muhammadiyah Kalimantan Timur



Submission date: 24-Jul-2024 02:30PM (UTC+0800)

Submission ID: 2421692648

File name: SKRIPSI_BOBI_ZINAIDIN_ZIDAN.docx (437.61K)

Word count: 4417

Character count: 26450

SKRIPSI BOBI ZINAIDIN ZIDAN

Arbansyah, S.kom, MTI

ORIGINALITY REPORT

27% SIMILARITY INDEX	23% INTERNET SOURCES	11% PUBLICATIONS	8% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	-----------------------------

PRIMARY SOURCES

1	jurnal.murnisadar.ac.id Internet Source	3%
2	data.langsakota.go.id Internet Source	3%
3	bkp.pertanian.go.id Internet Source	1%
4	rama.unimal.ac.id Internet Source	1%
5	Ari Ahmad Dhani, Taghfirul Azhima Yoga Siswa, Wawan Joko Pranoto. "Perbaikan Akurasi Random Forest Dengan ANOVA Dan SMOTE Pada Klasifikasi Data Stunting", Teknika, 2024 Publication	1%
6	pkm.tunasbangsa.ac.id Internet Source	1%
7	tahuko.com Internet Source	1%
8	Submitted to Sriwijaya University Student Paper	