

**ANALISIS KORELASI LUAS PANEN DENGAN MEMBANDINGKAN
ALGORITMA *LINEAR REGRESSION* DAN *RANDOM FOREST*
UNTUK PREDIKSI PRODUKSI TANAMAN PANGAN
DI PULAU KALIMANTAN**

SKRIPSI

SALSABILA BALQIS

201401027



**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

**ANALISIS KORELASI LUAS PANEN DENGAN MEMBANDINGKAN
ALGORITMA *LINEAR REGRESSION* DAN *RANDOM FOREST*
UNTUK PREDIKSI PRODUKSI TANAMAN PANGAN
DI PULAU KALIMANTAN**

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Ilmu Komputer

SALSABILA BALQIS
201401027



**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

PERSETUJUAN

Judul : ANALISIS KORELASI LUAS PANEN
 DENGAN MEMBANDINGKAN ALGORITMA
LINEAR REGRESSION DAN RANDOM FOREST UNTUK PREDIKSI PRODUKSI
 TANAMAN PANGAN DI PULAU
 KALIMANTAN

Kategori :

Nama : SALSABILA BALQIS

Nomor Induk Mahasiswa : 201401027

Program Studi : SARJANA (S-1) ILMU KOMPUTER

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI
 INFORMASI UNIVERSITAS SUMATERA

Tanggal Sidang : 09 JANUARI 2025

Komisi Pembimbing :

Dosen Pembimbing II

Hayatunnufus, S.Kom, M.Cs
 NIP. 199207192024062001

Dosen Pembimbing I

Prof. Drs. Poltak Sihombing M.Kom., Ph.D
 NIP. 196203171991031001

Diketahui/disetujui oleh Program
 Studi S1 Ilmu Komputer

Ketua,

Dra. Amalia ST., M.T.

NIP. 197812212014042001



PERNYATAAN

**ANALISIS KORELASI LUAS PANEN DENGAN MEMBANDINGKAN
ALGORITMA *LINEAR REGRESSION* DAN *RANDOM FOREST*
UNTUK PREDIKSI PRODUKSI TANAMAN PANGAN
DI PULAU KALIMANTAN**

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil penelitian saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah dicantumkan sumbernya.

Medan, 09 Januari 2025



Salsabila Balqis
NIM. 201401027

PENGHARGAAN

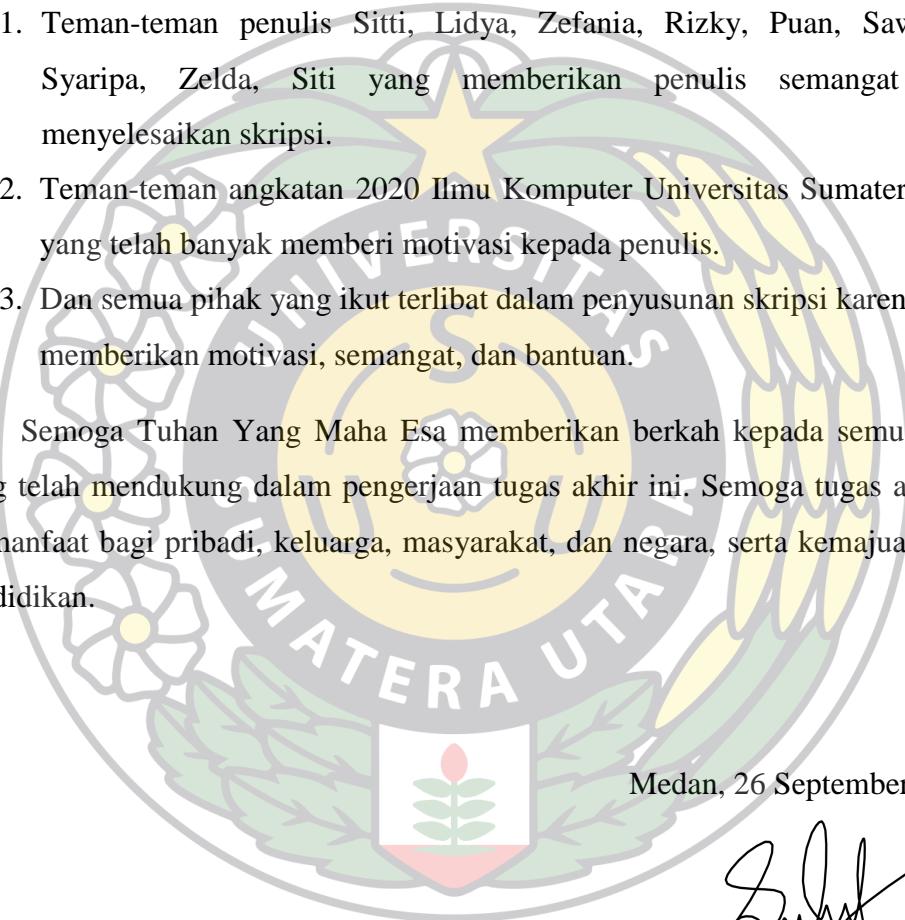
Puji syukur ke hadirat Tuhan Yang Maha Esa, karena atas berkat, rahmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini sebagai syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi S-1 Ilmu Komputer, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara.

Penulis ingin menyampaikan rasa hormat dan terima kasih yang sebesar-besarnya kepada pihak yang telah membantu penulis dalam bentuk doa, bimbingan, semangat, dan ilmu selama pembuatan dan penyelesaian skripsi ini.

1. Orang tua penulis, Ayah dan Ibu yang telah memberikan dukungan, semangat dan doa kepada penulis.
2. Bapak Dr. Muryanto Amin, S.Sos, M.Si. selaku Rektor Universitas Sumatera Utara.
3. Ibu Dr. Maya Silvi Lydia, B.Sc., M.Sc. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
4. Bapak Dr. Mohammad Andri Budiman, S.T., M.Comp.Sc., M.E.M. selaku Wakil Dekan 1 Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
5. Ibu Dr. Amalia, S.T., M.T. selaku Ketua Program Studi S1 Ilmu Komputer Universitas Sumatera Utara.
6. Ibu Sri Melvani Hardi, S.Kom, M.Kom. selaku Sekretaris Program Studi S1 Ilmu Komputer Universitas Sumatera Utara.
7. Ibu Dr. Ir. Elviawaty Muisa Zamzami S.T., M.T., MM., IPU, selaku Dosen Pembimbing Akademik, yang telah memberikan arahan, nasihat, dan dukungan kepada penulis sebagai mahasiswa dalam menempuh berbagai tahapan akademik.
8. Bapak Prof. Dr. Poltak Sihombing M.Kom. selaku dosen Pembimbing I yang telah memberikan saran, kritik, dan motivasi yang membangun dalam penggerjaan skripsi.

9. Ibu Hayatunnufus, S.Kom, M.Cs selaku dosen Pembimbing II yang sudah memberikan arahan, ide dan saran, kritik, motivasi, dan bimbingan dalam penyelesaian skripsi ini.
10. Seluruh dosen, staf pengajar, dan pegawai di Program Studi S-1 Ilmu Komputer, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara yang telah membantu penulis dalam proses penyusunan skripsi.
11. Teman-teman penulis Sitti, Lidya, Zefania, Rizky, Puan, Sawaliyah, Syaripa, Zelda, Siti yang memberikan penulis semangat untuk menyelesaikan skripsi.
12. Teman-teman angkatan 2020 Ilmu Komputer Universitas Sumatera Utara yang telah banyak memberi motivasi kepada penulis.
13. Dan semua pihak yang ikut terlibat dalam penyusunan skripsi karena sudah memberikan motivasi, semangat, dan bantuan.

Semoga Tuhan Yang Maha Esa memberikan berkah kepada semua pihak yang telah mendukung dalam penggerjaan tugas akhir ini. Semoga tugas akhir ini bermanfaat bagi pribadi, keluarga, masyarakat, dan negara, serta kemajuan dunia pendidikan.



Medan, 26 September 2024



Salsabila Balqis

NIM. 201401027

ABSTRAK

Pangan selalu menjadi isu penting dalam pembangunan, baik di tingkat global maupun nasional. Berdasarkan data dari laporan tahunan Direktorat Jenderal Tanaman Pangan, produksi pangan di Pulau Kalimantan menunjukkan tren penurunan dalam beberapa tahun terakhir, yang berpotensi mempengaruhi ketersediaan pangan di wilayah Kalimantan di masa mendatang. Meskipun pemerintah Indonesia telah mengembangkan program *food estate*, implementasinya menghadapi berbagai konflik yang menyebabkan kegagalan. Oleh karena itu, diperlukan perencanaan yang tepat dan terstruktur untuk menghindari dampak buruk yang mungkin terjadi. Untuk meningkatkan produksi pangan di Kalimantan, penting untuk memahami faktor-faktor yang mempengaruhi jumlah produksinya, salah satunya adalah luas panen. Penelitian ini bertujuan untuk menganalisis korelasi antara luas panen dan produksi pangan serta membandingkan kinerja algoritma *Linear Regression* dan *Random Forest* dalam memprediksi produksi pangan di Kalimantan. Data penelitian diperoleh dari situs resmi Kementerian Pertanian Republik Indonesia yang berupa data *time-series* luas panen dan produksi tanaman padi serta palawija dari tahun 1970 hingga 2023. Hasil analisis korelasi menunjukkan bahwa luas area panen memiliki pengaruh signifikan terhadap produksi tanaman pangan, dan model *Random Forest* menghasilkan nilai MAE dan RMSE yang lebih rendah serta nilai R^2 yang lebih tinggi dibandingkan model *Linear Regression*. Hal ini menunjukkan bahwa model *Random Forest* memiliki kinerja yang lebih baik dalam memprediksi produksi tanaman pangan di Kalimantan. Model ini kemudian diintegrasikan ke dalam aplikasi berbasis *website* yang menampilkan hasil prediksi, sehingga diharapkan dapat memberikan wawasan bagi pemangku kepentingan dalam perencanaan strategis guna mencegah kerawanan pangan akibat ketidakstabilan ketersediaan pangan di Kalimantan.

Kata kunci: Produksi Tanaman Pangan, Data *Time-Series*, Korelasi, *Linear Regression*, *Random Forest*

**CORRELATION ANALYSIS OF HARVEST AREA AND COMPARISON
OF LINEAR REGRESSION AND RANDOM FOREST ALGORITHMS
FOR PREDICTING FOOD CROP PRODUCTION
IN KALIMANTAN ISLAND**

ABSTRACT

Food security remains a crucial issue in development, both globally and nationally. According to data from the annual report of the Directorate General of Food Crops, food production in Kalimantan has shown a declining trend in recent years, which may affect food availability in the Kalimantan region in the future. Although the Indonesian government has developed the food estate program, its implementation has encountered various conflicts that have led to failures. Therefore, precise and structured planning is needed to prevent potential negative impacts. To enhance food production in Kalimantan, it is essential to understand the factors that influence production volume, one of which is harvest area. This study aims to analyze the correlation between harvest area and food production, as well as to compare the performance of the Linear Regression and Random Forest algorithms in predicting food production in Kalimantan. The research data was obtained from the official website of the Ministry of Agriculture of the Republic of Indonesia, consisting of time-series data on harvest area and production of rice and secondary crops from 1970 to 2023. The correlation analysis results indicate that harvest area significantly affects food production, and the Random Forest model produces lower MAE and RMSE values and higher R² values compared to the Linear Regression model. This indicates that the Random Forest model performs better in predicting food production in Kalimantan. This model is then integrated into a web-based application that displays the prediction results, thereby providing insights for stakeholders in strategic planning to prevent food insecurity due to unstable food availability in Kalimantan.

Keywords: Food Crop Production, Time-Series Data, Correlation, Linear Regression, Random Forest

DAFTAR ISI

PERSETUJUAN	iii
PERNYATAAN.....	iv
PENGHARGAAN	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Metode Penelitian	7
1.7 Sistematika Penulisan	8
BAB II LANDASAN TEORI	10
2.1 Ketahanan Pangan.....	10
2.2 Pulau Kalimantan.....	11
2.3 Analisis Korelasi.....	12
2.4 <i>Data Mining</i>	13
2.5 <i>Data Time Series</i>	15
2.6 <i>Linear Regression</i>	16
2.7 <i>Random Forest</i>	17
2.8 Evaluasi Performa Model	18
2.8.1 <i>Mean Absolute Error (MAE)</i>	19
2.8.2 <i>Root Mean Square Error (RMSE)</i>	19
2.8.3 <i>R-Squared (R²)</i>	20
2.9 Penelitian Relevan	20
BAB III ANALISIS DAN PERANCANGAN.....	23
3.1 Alur Penelitian	23
3.2 Perancangan Sistem	23

3.2.1	Pengumpulan Data	24
3.2.2	Pra-pemrosesan Data.....	24
3.2.3	Analisis Korelasi	25
3.2.4	Pembangunan Model.....	26
3.2.5	Perbandingan Evaluasi Kinerja Model.....	27
3.2.6	Visualisasi Hasil Prediksi pada Aplikasi Berbasis <i>Website</i>	28
BAB IV IMPLEMENTASI DAN PENGUJIAN		29
4.1	Implementasi Sistem.....	29
4.1.1	Spesifikasi Perangkat Keras.....	29
4.1.2	Spesifikasi Perangkat Lunak	29
4.2	Implementasi Tahap Pengumpulan Data	29
4.3	Implementasi Tahap Pra-pemrosesan Data.....	31
4.3.1	<i>Data Loading</i>	31
4.3.2	Penanganan <i>Missing Values</i>	33
4.3.3	Pembuatan <i>Lag Features</i>	34
4.4	Visualisasi <i>Dataset</i>	36
4.5	Implementasi Tahap Analisis Korelasi	48
4.6	Implementasi Tahap Pembangunan Model.....	57
4.6.1	Pemisahan Fitur dan Target	58
4.6.2	Pelatihan dan Evaluasi Model <i>Linear Regression</i>	58
4.6.3	Pelatihan dan Evaluasi Model <i>Random Forest</i>	59
4.7	Hasil Perbandingan Evaluasi Kinerja Model	61
4.8	Implementasi Prediksi Menggunakan Algoritma <i>Random Forest</i>	65
4.8.1	Pembangunan Model Prediksi Luas Panen	66
4.8.2	Implementasi Prediksi Luas Panen	67
4.8.3	Implementasi Prediksi Produksi Tanaman Pangan	69
4.8.4	Hasil Prediksi Luas Panen dan Produksi Tanaman Pangan	70
4.9	Implementasi Visualisasi Hasil Prediksi pada Aplikasi <i>Website</i>	93
BAB V PENUTUP		96
5. 1	Kesimpulan	96
5. 2	Saran	97
DAFTAR PUSTAKA		98

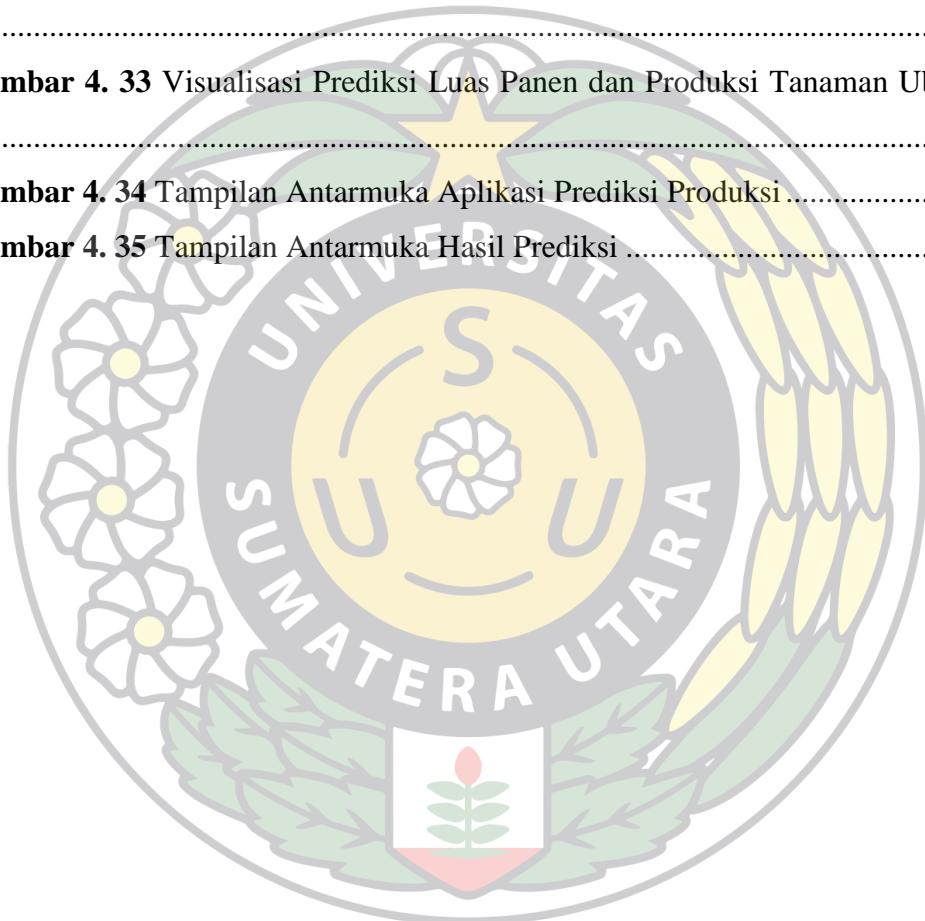
DAFTAR TABEL

Tabel 2. 1 Interpretasi Nilai Koefisien Korelasi.....	13
Tabel 4. 1 Gambaran isi Dataset yang Digunakan	30
Tabel 4. 2 Perhitungan Nilai Deviasi	49
Tabel 4. 3 Nilai Koefisien Korelasi Provinsi Kalimantan Barat	52
Tabel 4. 4 Nilai Koefisien Korelasi Provinsi Kalimantan Tengah	53
Tabel 4. 5 Nilai Koefisien Korelasi Provinsi Kalimantan Selatan	54
Tabel 4. 6 Nilai Koefisien Korelasi Provinsi Kalimantan Timur	55
Tabel 4. 7 Nilai Koefisien Korelasi Provinsi Kalimantan Utara	56
Tabel 4. 8 Evaluasi Kinerja Model untuk Tanaman Padi.....	62
Tabel 4. 9 Evaluasi Kinerja Model untuk Tanaman Jagung	62
Tabel 4. 10 Evaluasi Kinerja Model untuk Tanaman Kedelai	63
Tabel 4. 11 Evaluasi Kinerja Model untuk Tanaman Ubi Kayu	64
Tabel 4. 12 Evaluasi Kinerja Model untuk Tanaman Ubi Jalar	64
Tabel 4. 13 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Padi.....	71
Tabel 4. 14 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Jagung	75
Tabel 4. 15 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Kedelai ...	79
Tabel 4. 16 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Ubi Kayu	84
Tabel 4. 17 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Ubi Jalar .	88

DAFTAR GAMBAR

Gambar 2. 1 Peta Wilayah Pulau Kalimantan	11
Gambar 2. 2 Tahapan Knowledge Discovery in Database (KDD)	14
Gambar 3. 1 Alur Penelitian	23
Gambar 3. 2 Pengambilan Dataset dari Kementerian Pertanian	24
Gambar 4. 1 Import Library untuk Proses Pra-pemrosesan Data.....	31
Gambar 4. 2 Data Loading	32
Gambar 4. 3 Informasi Struktur Awal Dataset	32
Gambar 4. 4 Proses Konversi Tipe Data pada Dataset	33
Gambar 4. 5 Identifikasi Jumlah Nilai yang Hilang	34
Gambar 4. 6 Penanganan Data Hilang dengan Interpolasi Linear	34
Gambar 4. 7 Pembuatan Lag Features Produksi.....	35
Gambar 4. 8 Dataset Setelah Proses Pembuatan Lag Features.....	35
Gambar 4. 9 Penghapusan Nilai NaN.....	36
Gambar 4. 10 Proses Visualisasi Dataset	36
Gambar 4. 11 Hasil Visualisasi Dataset Tanaman Padi	37
Gambar 4. 12 Hasil Visualisasi Dataset Tanaman Jagung	39
Gambar 4. 13 Hasil Visualisasi Dataset Tanaman Kedelai	42
Gambar 4. 14 Hasil Visualisasi Dataset Tanaman Ubi Kayu.....	44
Gambar 4. 15 Hasil Visualisasi Dataset Tanaman Ubi Jalar.....	46
Gambar 4. 16 Proses Analisis Korelasi	51
Gambar 4. 17 Import Library untuk Proses Pembangunan Model.....	57
Gambar 4. 18 Pemisahan Fitur dan Target	58
Gambar 4. 19 Inisialisasi dan Pelatihan Model Linear Regression	58
Gambar 4. 20 Evaluasi Model Linear Regression	59
Gambar 4. 21 Inisialisasi dan Pelatihan Model Random Forest	60
Gambar 4. 22 Evaluasi Model Random Forest.....	61
Gambar 4. 23 Pembuatan Lag Features Luas Panen	66
Gambar 4. 24 Pemisahan Fitur dan Target Model Prediksi Luas Panen.....	66
Gambar 4. 25 Insialisasi dan Pelatihan Model Prediksi Luas Panen.....	67

Gambar 4. 26 Source Code Prediksi Luas Panen	68
Gambar 4. 27 Source Code Prediksi Produksi	69
Gambar 4. 28 Source Code Visualisasi Hasil Prediksi.....	70
Gambar 4. 29 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Padi	72
Gambar 4. 30 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Jagung.	76
Gambar 4. 31 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Kedelai	81
Gambar 4. 32 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Ubi Kayu	85
Gambar 4. 33 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Ubi Jalar	89
Gambar 4. 34 Tampilan Antarmuka Aplikasi Prediksi Produksi	94
Gambar 4. 35 Tampilan Antarmuka Hasil Prediksi	95



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pangan selalu menjadi isu penting dalam pembangunan baik di tingkat global maupun nasional. Pangan adalah kebutuhan dasar bagi setiap manusia, dan pemenuhannya merupakan bagian dari hak asasi manusia. Negara memiliki kewajiban untuk menjamin ketersediaan, keterjangkauan, serta pemenuhan pangan yang cukup, aman, dan berkualitas bagi seluruh warganya (Hadi *et al.*, 2019). Undang-Undang No. 18 tahun 2012 mendefinisikan ketahanan pangan sebagai kondisi di mana kebutuhan pangan suatu negara, hingga ke tingkat individu, dapat terpenuhi. Kondisi ini tercermin dalam ketersediaan pangan yang cukup, baik dari segi jumlah maupun mutu, yang aman, beragam, bergizi, merata, terjangkau, serta sesuai dengan agama, keyakinan, dan budaya masyarakat, sehingga mendukung kehidupan yang sehat, aktif, dan produktif secara berkelanjutan. Mengacu pada hal tersebut, pencapaian ketahanan pangan harus dilakukan secara bertahap, dengan fokus tidak hanya pada tingkat nasional, tetapi juga pada tingkat provinsi serta kabupaten/kota.

Menurut data dari *Global Food Security Index* (GFSI), indeks ketahanan pangan Indonesia pada tahun 2022 mengalami peningkatan menjadi 60,2, yang menunjukkan kenaikan sebesar 1,69% dibandingkan tahun 2021. Peningkatan ini menempatkan Indonesia pada peringkat ke-69 dari 113 negara. Meskipun ada peningkatan dibandingkan tahun sebelumnya, indeks ketahanan pangan Indonesia masih di bawah rata-rata global yang mencapai 62,2 dan lebih rendah dibandingkan dengan rata-rata Asia Pasifik yang berada di angka 63,4. Indonesia perlu meningkatkan ketersediaan pangan dengan memproduksi pangan dalam jumlah besar dan berkualitas tinggi, terutama pada komoditas tanaman pangan padi dan palawija. Selain padi, palawija juga memiliki peranan signifikan dalam penyediaan pangan di

Indonesia. Palawija berfungsi sebagai sumber utama makanan bagi manusia dan ternak, bahan baku industri, serta sisa hijauannya yang dapat dimanfaatkan untuk menyuburkan tanah (Direktorat Statistik Tanaman Pangan, Hortikultura, dan Perkebunan, 2022).

Sebagian besar komoditas pangan dihasilkan di wilayah-wilayah sentra produksi pangan. Saat ini, produksi pangan masih terpusat di pulau Jawa dan Sumatera, menyebabkan wilayah-wilayah ini sering mengalami surplus pangan saat musim panen, sementara wilayah lainnya, yang merupakan wilayah non-sentra produksi, cenderung mengalami defisit pangan (Rumra *et al.*, 2023). Hal ini dapat terlihat dari data produktivitas pangan pada tahun 2023, di mana Kalimantan memiliki produktivitas yang relatif rendah dibandingkan dengan pulau-pulau lainnya. Produktivitas pertanian di Kalimantan untuk sebagian besar komoditas masih tertinggal dibandingkan dengan Jawa dan Sumatera.

Sebagai salah satu pulau terluas di Indonesia, Kalimantan memiliki potensi besar di sektor pertanian (Christiyanto dan Mayulu, 2021). Namun, dari laporan tahunan Direktorat Jenderal Tanaman Pangan menunjukkan bahwa produksi pangan di pulau ini menunjukkan tren penurunan dalam beberapa tahun terakhir. Pada tahun 2023, produksi berbagai komoditas pangan di Kalimantan mengalami penurunan yang signifikan dibandingkan dengan tahun sebelumnya. Produksi padi turun sebesar 7.331 ton, sementara jagung mengalami penurunan yang cukup besar dengan selisih 172.927 ton. Penurunan juga terjadi pada produksi kedelai, yang turun sebesar 8.097 ton. Selain itu, produksi ubi kayu dan ubi jalar juga mengalami penurun, yaitu sebesar 29.039 ton dan 7.190 ton dibandingkan tahun 2022. Kondisi ini tentu berdampak pada ketersediaan pangan di wilayah Kalimantan.

Untuk mengatasi permasalahan ini, pemerintah Indonesia telah mengembangkan program *food estate* di Kalimantan sebagai salah satu langkah strategis untuk meningkatkan produksi pangan di wilayah ini. *Food estate* merupakan area pertanian yang dikembangkan dalam skala besar dengan tujuan untuk menghasilkan bahan pangan yang dibutuhkan oleh

masyarakat di wilayah tersebut. Program ini diharapkan dapat meningkatkan kapasitas produksi lokal dan mengurangi ketergantungan pangan dari wilayah lain. Namun, dalam implementasinya, proyek *food estate* menghadapi banyak konflik yang menyebabkan kegagalan. Salah satu faktor utama kegagalan ini adalah kurangnya perencanaan yang matang (Rasman *et al.*, 2023). Oleh karena itu, perencanaan yang tepat dan terstruktur sangat diperlukan untuk mencegah dampak-dampak buruk yang dapat terjadi.

Dalam upaya meningkatkan perencanaan dan pengelolaan produksi pangan di Pulau Kalimantan, penting untuk memahami keterkaitan antara berbagai faktor yang memengaruhi jumlah produksi pangan, salah satunya adalah luas panen. Luas panen merupakan indikator utama yang sering digunakan untuk memperkirakan potensi produksi suatu wilayah. Analisis korelasi antara luas panen dan produksi pangan dapat memberikan wawasan penting mengenai seberapa besar pengaruh luas panen terhadap jumlah produksi yang dihasilkan. Dengan mengetahui korelasi ini, strategi yang lebih efektif dapat dirancang untuk mengoptimalkan penggunaan lahan pertanian dan meningkatkan produksi pangan secara keseluruhan. Pemahaman ini kemudian dapat menjadi dasar untuk membuat prediksi jumlah produksi pangan di masa depan. Prediksi yang akurat sangat diperlukan agar kebijakan dan keputusan yang diambil bisa lebih tepat sasaran, terutama dalam menjaga ketahanan pangan di Kalimantan. Selain itu, prediksi ini juga dapat memberikan wawasan terkait dengan pengembangan proyek *food estate* di Kalimantan ke depannya. Dengan prediksi yang baik, para pemangku kepentingan dapat menyusun rencana yang lebih matang serta meminimalkan risiko kekurangan pangan.

Prediksi produksi pangan di Pulau Kalimantan dapat dilakukan dengan menggunakan metode *data mining*. *Data mining* merupakan proses pengumpulan dan pengolahan data untuk memperoleh pengetahuan atau informasi yang berguna dari sekumpulan data (Jassim dan Abdulwahid, 2020). Dalam *data mining*, terdapat berbagai teknik yang dapat diterapkan, salah satunya adalah *forecasting* (Hussein *et al.*, 2020). *Forecasting* adalah

metode yang digunakan untuk memprediksi nilai data di masa yang akan datang berdasarkan data yang diperoleh secara berkelanjutan dari waktu ke waktu (*time series*) (Kolambe dan Arora, 2024).

Penelitian ini akan menggunakan algoritma *Linear Regression* dan *Random Forest* untuk melakukan *forecasting*. *Linear Regression* adalah algoritma *machine learning* yang paling dasar dan sering diterapkan untuk memprediksi nilai suatu variabel target berdasarkan satu atau lebih variabel independen (Acharya *et al.*, 2019). Sementara itu, *Random Forest* adalah algoritma *ensemble learning* yang menggabungkan beberapa pohon keputusan (*decision trees*) dan menggabungkan hasil prediksi dari pohon-pohon tersebut untuk meningkatkan akurasi prediksi. *Random Forest* memperbaiki kelemahan dari *decision tree* yang cenderung mengalami *overfitting* terhadap data latih (Fu *et al.*, 2023).

Berdasarkan penjelasan di atas, penelitian ini akan membandingkan algoritma *Linear Regression* dan *Random Forest* untuk melakukan prediksi produksi pangan di Pulau Kalimantan. Perbandingan dari algoritma ini bertujuan untuk menentukan model prediksi yang paling akurat dan efektif berdasarkan tingkat kesalahan yang dihasilkan dari kedua algoritma. Dengan begitu, penelitian ini diharapkan dapat bermanfaat bagi para pemangku kepentingan dalam merencanakan langkah ke depan guna menghindari terjadinya kerawanan pangan akibat tidak stabilnya ketersediaan pangan yang terus melanda wilayah Kalimantan.

1.2 Rumusan Masalah

Berdasarkan data dari laporan tahunan Direktorat Jenderal Tanaman Pangan, produksi pangan di Pulau Kalimantan menunjukkan tren penurunan dalam beberapa tahun terakhir, yang dapat mempengaruhi ketersediaan pangan di wilayah Kalimantan pada masa mendatang. Meskipun pemerintah Indonesia telah mengembangkan program *food estate*, implementasinya menghadapi berbagai konflik yang menyebabkan kegagalan. Oleh karena itu, diperlukan perencanaan yang tepat dan terstruktur untuk menghindari

dampak buruk yang mungkin terjadi. Untuk meningkatkan produksi pangan di Pulau Kalimantan, penting untuk memahami keterkaitan antara faktor-faktor yang mempengaruhi produksi pangan, salah satunya adalah luas panen. Analisis korelasi antara luas panen dan produksi pangan dapat memberikan wawasan penting mengenai seberapa besar pengaruh luas panen terhadap jumlah produksi yang dihasilkan. Pemahaman ini dapat menjadi dasar dalam membuat prediksi produksi pangan di masa depan dan memberikan wawasan yang berguna bagi pengembangan proyek *food estate* di Kalimantan. Penelitian ini akan membandingkan algoritma *Linear Regression* dan *Random Forest* dalam melakukan prediksi dengan tujuan untuk menentukan model prediksi yang akurat berdasarkan tingkat kesalahan yang dihasilkan dari kedua algoritma.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Penelitian ini fokus pada wilayah Pulau Kalimantan dan tidak mencakup wilayah lain di Indonesia.
- 2) Penelitian berfokus pada dua variabel utama, yaitu luas panen dan jumlah produksi pangan, tanpa mempertimbangkan faktor-faktor lain seperti perubahan iklim, teknologi pertanian, atau kebijakan pemerintah yang juga dapat mempengaruhi produksi pangan.
- 3) Jenis tanaman pangan yang akan diteliti ialah kelompok komoditas pangan utama, yaitu padi dan beragam jenis tanaman palawija, yaitu jagung, kedelai, ubi kayu, dan ubi jalar.
- 4) Data yang digunakan dalam penelitian ini merupakan data *time series* yang bersumber dari Kementerian Pertanian Republik Indonesia pada periode 1970-2023.
- 5) Penelitian akan membandingkan algoritma *Linear Regression* dan *Random Forest* dalam melakukan prediksi.

- 6) Penelitian ini menggunakan analisis korelasi dengan metode statistik sederhana untuk memahami hubungan antara luas panen dan produksi pangan.
- 7) Evaluasi model prediksi akan dilakukan berdasarkan metrik evaluasi, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2) untuk menilai tingkat kesalahan serta akurasi masing-masing algoritma.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis korelasi antara luas panen dan jumlah produksi pangan di Pulau Kalimantan guna memahami sejauh mana luas panen memengaruhi tingkat produksi pangan. Berdasarkan pemahaman ini, penelitian akan melakukan prediksi produksi pangan di Kalimantan menggunakan algoritma *Linear Regression* dan *Random Forest*. Selanjutnya, hasil prediksi dari kedua algoritma akan dibandingkan berdasarkan tingkat kesalahan yang dihasilkan untuk menentukan model prediksi yang paling akurat. Berdasarkan perbandingan ini, penelitian ini juga akan merekomendasikan model prediksi yang paling sesuai untuk perencanaan ketahanan pangan di Kalimantan. Selain itu, hasil prediksi yang diperoleh diharapkan dapat memberikan wawasan bagi pengembangan proyek *food estate* di Kalimantan ke depannya.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini antara lain:

- 1) Bagi Peneliti

Memberikan kesempatan bagi peneliti untuk mengembangkan keahlian dalam melakukan analisis korelasi serta menerapkan algoritma *machine learning*, yaitu algoritma *Linear Regression* dan *Random Forest* untuk membuat model prediksi.

2) Bagi Akademik

Menyediakan referensi bagi akademisi terkait dengan prediksi produksi pangan yang dapat dijadikan perbandingan dan acuan untuk masalah serupa, sehingga mendukung peningkatan kualitas pendidikan.

3) Bagi Pemerintah

Memberikan dukungan bagi pemerintah dalam perumusan kebijakan pangan yang lebih efektif di Kalimantan berdasarkan hasil prediksi yang akurat dan pemahaman korelasi luas panen, sehingga pemerintah dapat mengoptimalkan pengelolaan lahan dan memastikan ketahanan pangan di masa depan.

1.6 Metode Penelitian

Berikut ini adalah metodologi penelitian yang dilakukan dalam penelitian ini:

1) Studi Pustaka

Tahap awal penelitian dimulai dengan melakukan kajian literatur melalui buku, jurnal, *e-book*, artikel ilmiah, dan sumber-sumber dari situs *internet* yang relevan dan mendukung topik penelitian, yaitu terkait dengan ketahanan pangan, analisis korelasi, *data mining*, data *time-series*, *forecasting*, *Linear Regression*, dan *Random Forest*. Hasil dari studi pustaka ini akan memberikan landasan teoritis yang kuat bagi penelitian.

2) Analisis dan Perancangan

Selanjutnya, melakukan analisis terhadap kebutuhan penelitian dan perancangan kerangka kerja penelitian. Perancangan meliputi identifikasi variabel yang akan dianalisis, pemilihan metode statistik dan algoritma prediksi yang akan digunakan, serta diagram alir proses penelitian dari pengumpulan data hingga evaluasi model.

3) Implementasi

Tahapan implementasi meliputi proses membangun model menggunakan algoritma *Linear Regression* dan *Random Forest* menggunakan bahasa pemrograman *Python* sesuai dengan diagram alir yang telah dirancang.

4) Evaluasi

Pada tahap ini akan dilakukan evaluasi model menggunakan nilai metrik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2) untuk mengukur performa dan akurasi prediksi. Hasil dari evaluasi ini akan menentukan model dengan kinerja terbaik dalam melakukan prediksi.

5) Dokumentasi

Seluruh proses penelitian didokumentasikan secara sistematis ke dalam bentuk penyusunan laporan akhir. Dokumentasi mencakup metodologi yang digunakan, hasil analisis, pembangunan model, serta hasil prediksi. Dokumentasi ini penting untuk memastikan transparansi penelitian dan memudahkan pemahaman serta replikasi oleh peneliti lain di masa depan.

1.7 Sistematika Penulisan

Berikut merupakan sistematika penulisan dari skripsi ini, yang terbagi ke dalam beberapa bagian utama, yaitu:

BAB I

PENDAHULUAN

Bagian ini mengulas latar belakang permasalahan yang akan diteliti, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, serta sistematika penulisan dalam penelitian ini.

BAB II

LANDASAN TEORI

Bagian ini membahas topik-topik terkait ketahanan pangan, analisis korelasi, serta algoritma *Linear Regression* dan *Random Forest*.

BAB III**ANALISIS DAN PERANCANGAN SISTEM**

Bagian ini menguraikan analisis permasalahan penelitian serta memberikan gambaran umum mengenai perancangan sistem yang akan diterapkan dalam penelitian ini.

BAB IV**IMPLEMENTASI DAN PENGUJIAN SISTEM**

Bagian ini menjelaskan hasil dari implementasi analisis dan perancangan yang telah dilakukan sebelumnya, serta menyajikan hasil analisis yang diperoleh dari proses implementasi tersebut.

BAB V**PENUTUP**

Bagian ini menyajikan kesimpulan yang diambil dari hasil penelitian serta memberikan saran untuk pengembangan penelitian di masa yang akan datang.



BAB II

LANDASAN TEORI

2.1 Ketahanan Pangan

Pangan merupakan kebutuhan dasar yang paling penting bagi manusia untuk mempertahankan hidup dan kesejahteraan. Pembangunan di bidang pangan dan gizi di Indonesia sangat erat kaitannya dengan pencapaian ketahanan pangan, sebagaimana tercantum dalam Undang-Undang Nomor 7 Tahun 1996 tentang Pangan dan Peraturan Pemerintah Nomor 68 Tahun 2002 tentang Ketahanan Pangan. Ketahanan pangan memegang peranan sangat penting bagi suatu negara, khususnya negara dengan populasi besar seperti Indonesia. Populasi Indonesia diperkirakan mencapai 220 juta jiwa pada tahun 2020 dan diproyeksikan akan mencapai 270 juta jiwa pada tahun 2025 (Chaireni *et al.*, 2020). Indonesia terus berupaya mencapai ketahanan pangan dan gizi secara berkelanjutan (Ariani dan Suryana, 2023). Upaya ini sejalan dengan tujuan Pembangunan Pertanian Berkelanjutan (SDGs), khususnya SDG-1 yang bertujuan untuk mengentaskan kemiskinan, serta SDG-2 yang bertujuan untuk mengakhiri kelaparan melalui ketahanan pangan, perbaikan gizi, dan sistem pertanian yang berkelanjutan (Deputi Bidang Kerawanan Pangan dan Gizi Badan Pangan Nasional, 2023).

Menurut *Food and Agriculture Organization* (FAO), ketahanan pangan didefinisikan sebagai kondisi di mana semua orang, setiap saat, memiliki akses fisik, sosial, dan ekonomi yang cukup terhadap pangan yang aman dan bergizi untuk memenuhi kebutuhan dan preferensi makanan mereka guna hidup sehat dan aktif. Terdapat empat pilar dalam ketahanan pangan, yaitu ketersediaan, akses, pemanfaatan, dan stabilitas (Alonso *et al.*, 2018). Ketersediaan pangan memastikan adanya pasokan pangan yang cukup, baik melalui produksi lokal, impor, maupun cadangan. Aksesibilitas pangan mengacu pada kemampuan individu untuk memperoleh pangan secara ekonomi dan fisik, yang dipengaruhi oleh pendapatan, harga pangan, dan

distribusi. Pemanfaatan pangan mencakup aspek kualitas dan keamanan pangan serta pengetahuan gizi untuk memastikan konsumsi yang tepat. Stabilitas pangan menekankan pada konsistensi ketersediaan dan akses pangan dari waktu ke waktu, mengatasi potensi gangguan seperti bencana alam atau fluktuasi ekonomi (Peng dan Berry, 2019).

2.2 Pulau Kalimantan

Berdasarkan Peraturan Menteri Dalam Negeri Nomor 137 Tahun 2017 tanggal 29 Desember 2017, luas wilayah Pulau Kalimantan yang termasuk dalam wilayah Indonesia mencapai 544.150,07 km², atau sekitar 28,39 persen dari total luas wilayah Indonesia yang mencapai 1.916.862,20 km². Secara administratif, Pulau Kalimantan dibagi menjadi lima provinsi, yaitu Kalimantan Utara dengan ibu kota Tanjung Selor, Kalimantan Timur dengan ibu kota Samarinda, Kalimantan Selatan dengan ibu kota Banjarmasin, Kalimantan Tengah dengan ibu kota Palangkaraya, dan Kalimantan Barat dengan ibu kota Pontianak. Provinsi dengan luas terbesar adalah Kalimantan Tengah, yang memiliki luas wilayah sebesar 153.564,50 km², sementara provinsi dengan luas wilayah terkecil di Kalimantan adalah Kalimantan Selatan, dengan luas hanya sekitar 38.744,23 km² (Pemerintah Provinsi Kalimantan Timur, 2019).



Gambar 2. 1 Peta Wilayah Pulau Kalimantan

Sumber: tribunnews.com

2.3 Analisis Korelasi

Analisis korelasi adalah metode statistik yang digunakan untuk mengidentifikasi hubungan antara dua variabel. Tujuan utama dari analisis korelasi adalah untuk mengetahui sejauh mana satu variabel berhubungan dengan variabel lainnya, serta apakah hubungan tersebut bersifat positif, negatif, atau tidak ada hubungan sama sekali (Sofian *et al.*, 2022). Ada beberapa jenis metode analisis korelasi, salah satunya adalah korelasi *Pearson*. Korelasi *Pearson* adalah metode statistik yang digunakan untuk mengukur hubungan linear antara dua variabel. Koefisien korelasi *Pearson* (r) memiliki rentang nilai antara -1 hingga 1. Nilai korelasi 1 menunjukkan adanya hubungan linear positif yang sempurna antara kedua variabel, sementara nilai -1 menunjukkan hubungan linear negatif yang sempurna. Semakin mendekati 0 nilai absolut koefisien korelasi, semakin lemah hubungan linier antara kedua variabel. Sebaliknya, semakin mendekati 1 atau -1, semakin kuat hubungan linier antara kedua variabel (Yang *et al.*, 2021). Untuk menghitung nilai koefisien korelasi *Pearson*, dapat digunakan rumus sebagai berikut (Munandar *et al.*, 2020):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Keterangan:

- r : koefisien korelasi *Pearson*
- x_i, y_i : nilai dari variabel x, y untuk data ke-i
- \bar{x}, \bar{y} : nilai rata-rata dari variabel x, y

Interpretasi koefisien korelasi *Pearson* tidak memiliki standar yang universal dan berlaku untuk semua konteks. Rentang nilai koefisien korelasi *Pearson* (r) berkisar antara -1 hingga 1, di mana interpretasinya bisa berbeda tergantung pada disiplin ilmu. Namun, secara umum, interpretasi berikut sering digunakan untuk menafsirkan kekuatan korelasi:

Tabel 2. 1 Interpretasi Nilai Koefisien Korelasi

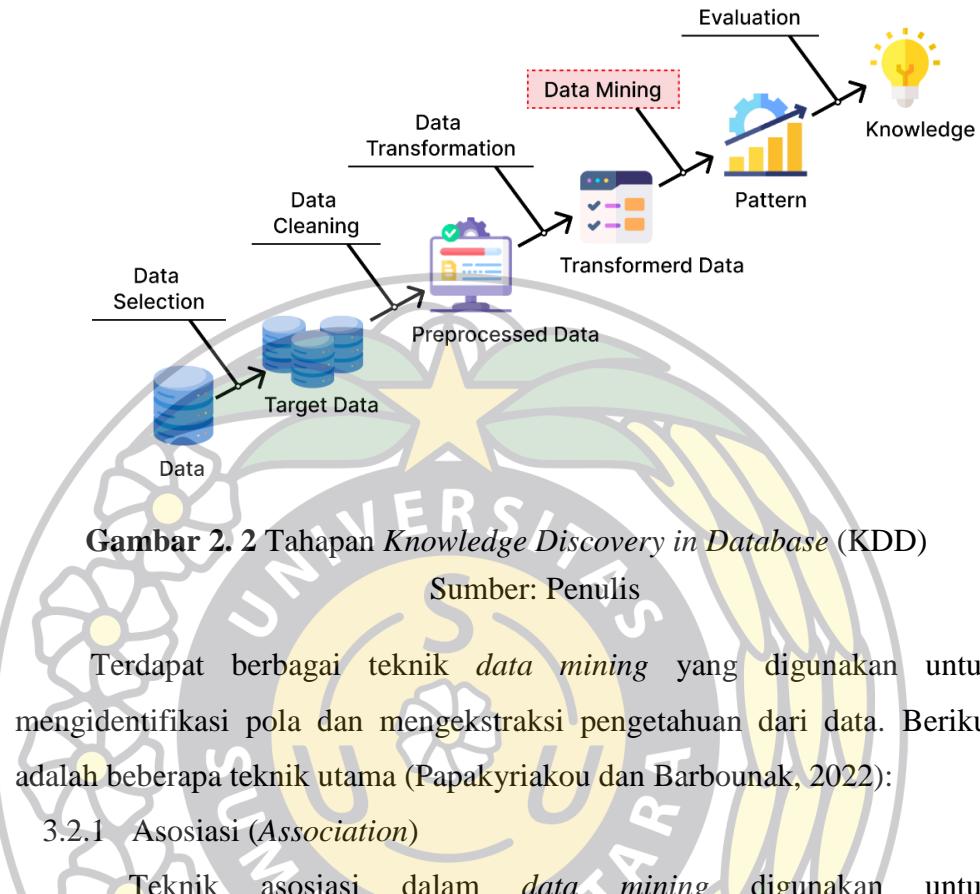
Nilai Koefisien (r)	Tingkat Korelasi
0	Tidak ada korelasi
0.01 – 0.09	Korelasi Tidak Signifikan
0.10 – 0.29	Korelasi Lemah
0.30 – 0.49	Korelasi Sedang
0.50 – 0.69	Korelasi Kuat
0.70– 0.89	Korelasi Sangat Kuat
> 0.90	Korelasi Mendekati Sempurna
1	Korelasi Sempurna

Analisis korelasi dapat memberikan wawasan tentang pola hubungan antara variabel, namun korelasi tidak mengindikasikan sebab-akibat. Hubungan yang kuat antara dua variabel tidak selalu menunjukkan bahwa satu variabel menyebabkan perubahan pada variabel lainnya (Negri, 2023). Analisis korelasi membantu dalam mengidentifikasi keterkaitan yang ada, tetapi studi lebih lanjut diperlukan untuk memahami dinamika sebab-akibat yang mendasarinya.

2.4 Data Mining

Data mining adalah proses untuk mengekstrak pengetahuan atau informasi penting dari kumpulan data yang besar dan kompleks. Tujuan utama dari *data mining* adalah untuk mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak langsung terlihat dalam data, sehingga memberikan wawasan yang lebih mendalam dan bermanfaat. Proses ini melibatkan penggunaan berbagai disiplin ilmu seperti kecerdasan buatan (*artificial intelligence*), *machine learning*, dan teknik statistik untuk menganalisis data secara sistematis dan otomatis (Jassim dan Abdulwahid, 2020). *Data mining* adalah salah satu tahap dalam proses *Knowledge Discovery In Database* (KDD), yang merupakan pencarian informasi baru

dan berharga dari kumpulan data atau *database*. Proses KDD dimulai dari seleksi data hingga evaluasi (Mandour *et al.*, 2024).



Terdapat berbagai teknik *data mining* yang digunakan untuk mengidentifikasi pola dan mengekstraksi pengetahuan dari data. Berikut adalah beberapa teknik utama (Papakyriakou dan Barbounak, 2022):

3.2.1 Asosiasi (Association)

Teknik asosiasi dalam *data mining* digunakan untuk mengidentifikasi hubungan atau pola tersembunyi antara item-item dalam kumpulan data yang besar. Beberapa algoritma yang umum digunakan ialah algoritma *Apriori*, *Eclat*, dan *Frequent Pattern (FP) Growth*.

3.2.2 Klasifikasi (Classification)

Klasifikasi adalah teknik *data mining* yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas berdasarkan karakteristik atau atribut tertentu. Model klasifikasi dilatih dengan *dataset* yang sudah berlabel, kemudian digunakan untuk memprediksi label kategori dari data baru.

3.2.3 Regresi (*Regression*)

Regresi adalah teknik *data mining* yang digunakan untuk menganalisis hubungan antara variabel dependen dan satu atau lebih variabel independen. Teknik ini membantu untuk memahami korelasi dan independensi variabel untuk menentukan faktor yang penting dan faktor yang dapat diabaikan.

3.2.4 Prediksi (*Forecasting*)

Prediksi dalam *data mining* adalah proses memperkirakan nilai atau hasil masa depan berdasarkan data historis dan model prediktif. Teknik ini digunakan untuk memprediksi perilaku dan tren untuk meningkatkan strategi pengambilan keputusan dan kinerja bisnis.

3.2.5 Klasterisasi (*Clustering*)

Klasterisasi adalah teknik *data mining* yang digunakan untuk mengelompokkan data ke dalam kelompok atau klaster berdasarkan kesamaan fitur atau karakteristik tertentu.

2.5 Data *Time Series*

Data *time series* adalah sekumpulan titik data yang dikumpulkan atau dicatat pada interval waktu yang berurutan dan biasanya dalam jarak waktu yang sama (Prater *et al.*, 2024). Data ini menjadi landasan dalam melakukan peramalan deret waktu (*time series forecasting*) untuk memprediksi nilai di masa depan. Teknik pemodelan prediktif ini sangat diperlukan dalam berbagai lintas bidang, seperti ilmu sains, ekonomi, dan keuangan, di mana pemahaman dan prediksi keadaan masa depan sangat penting untuk pengambilan keputusan yang tepat. Terdapat tiga komponen utama dalam data *time series* (Fatima dan Rahimi, 2024), yaitu sebagai berikut:

- a) *Trend* (Tren): Mengacu pada pola yang menunjukkan peningkatan atau penurunan dari waktu ke waktu yang dapat bersifat linier atau non-linier.

- b) *Seasonality* (Musiman): Merujuk pada pola berulang yang terjadi secara teratur dalam rentang waktu tertentu, seperti harian, mingguan, bulanan, atau tahunan.
- c) *Residuals* (Sisa): Mencerminkan *noise* yang tidak dapat dijelaskan oleh model yang mencakup tren dan musiman.

2.6 Linear Regression

Linear Regression adalah salah satu algoritma statistik dan *machine learning* yang paling dasar dan sering digunakan untuk memodelkan hubungan antara variabel dependen dan variabel independen (Acharya *et al.*, 2019). Algoritma ini bertujuan untuk menemukan fungsi linier yang optimal, yaitu menentukan sekumpulan koefisien (bobot) sehingga fungsi tersebut dapat memprediksi nilai variabel dependen seakurat mungkin (Qu, 2024). Variabel dependen atau target adalah variabel yang dipengaruhi oleh variabel lain dan diwakili oleh Y, sedangkan variabel independen adalah variabel yang tidak dipengaruhi oleh variabel lain dan diwakili oleh X. Berdasarkan jumlah variabel independen (X), *Linear Regression* terbagi menjadi dua jenis (Alita *et al.*, 2021) :

- a) *Simple Linear Regression*

Regresi Linier Sederhana (*Simple Linear Regression*) merupakan algoritma regresi paling dasar, yang hanya melibatkan satu variabel independen dan satu variabel target. Algoritma ini dapat digunakan untuk mengeksplorasi hubungan antara dua variabel dengan mengungkapkan bagaimana satu variabel berubah seiring dengan perubahan variabel lain. Keterkaitan antara dua variabel tersebut dapat dinyatakan dengan persamaan berikut:

$$Y = \alpha + \beta X$$

Keterangan :

α : nilai konstanta

β : nilai koefisien

Y : variabel target

X : variabel independen

b) *Multiple Linear Regression*

Regresi Linier Berganda (*Multiple Linear Regression*) merupakan algoritma regresi yang melibatkan satu variabel target dan lebih dari satu variabel independen. Algoritma ini lebih kompleks dibandingkan dengan *Simple Linear Regression* karena melibatkan lebih dari satu variabel independen (X_1, X_2, \dots). Algoritma ini dapat dituliskan dengan persamaan berikut:

$$Y = \alpha + \beta_1X_1 + \beta_2X_2 \dots + \beta_nX_n$$

Keterangan:

- α : nilai konstanta
- $\beta_1, \beta_2, \dots, \beta_n$: nilai koefisien
- Y : variabel target
- X_1, X_2, \dots, X_n : variabel independen

2.7 Random Forest

Random forest merupakan algoritma *ensemble learning* yang dikenalkan oleh Breiman pada tahun 2001. Algoritma ini mengkombinasikan sejumlah pohon keputusan (*decision trees*) dan menggabungkan seluruh hasil prediksi dari pohon-pohon tersebut untuk meningkatkan akurasi prediksi. *Random forest* memperbaiki kelemahan dari *decision tree* yang cenderung mengalami *overfitting* terhadap data latih (Fu *et al.*, 2023). *Random Forest* sering digunakan untuk menangani masalah dalam klasifikasi, regresi, dan lainnya. Pada masalah regresi, hasil prediksi digabungkan dengan cara menghitung rata-rata, sementara pada masalah klasifikasi, hasil prediksi digabungkan dengan cara melakukan *majority vote* (Wang *et al.*, 2024). Adapun langkah-langkah dalam pembentukan model *Random Forest* adalah sebagai berikut:

- Pengambilan Sampel: Algoritma mengambil beberapa sampel dari *dataset*. Setiap sampel diambil secara acak, dengan penggantian, sehingga data yang sama bisa terpilih lebih dari sekali.
- Membangun Pohon Keputusan: Dari setiap sampel, *Random Forest* membangun pohon keputusan, yaitu pohon regresi untuk prediksi nilai kontinu atau pohon klasifikasi untuk prediksi kategori. Untuk setiap pemisahan (*split*) di pohon, hanya sebagian fitur yang dipilih secara acak, sehingga pohon berbeda-beda satu sama lain. Setiap pohon tumbuh dari atas ke bawah menggunakan metode CART (*Classification and Regression Tree*), yang terus melakukan pemisahan sampai mencapai hasil terbaik di setiap cabang, tanpa memangkas pohon.
- Menggabungkan Hasil: Setelah semua pohon selesai dibangun, hasil prediksi dari setiap pohon digabungkan. Untuk regresi, hasil akhir diperoleh dengan rata-rata semua prediksi pohon, sedangkan untuk klasifikasi, hasilnya ditentukan dengan *voting* berdasarkan suara mayoritas dari semua pohon.

Jika terdapat satu set data input $\{H(x, \theta_i), i=1, 2, \dots, k\}$ dan nilai prediksi dari satu pohon keputusan adalah $\{H(x, \theta_i)\}$, maka hasil akhir dari model prediksi *Random Forest* akan menjadi rata-rata dari hasil prediksi seluruh pohon keputusan seperti pada persamaan di bawah ini (Li *et al.*, 2022).

$$\bar{H}(x) = \frac{1}{k} \sum_{i=1}^k H(x, \theta_i)$$

Keterangan:

- | | |
|------------------|--|
| $\bar{H}(x)$ | : nilai prediksi akhir |
| $H(x, \theta_i)$ | : prediksi pohon keputusan ke- <i>i</i> untuk input <i>x</i> . |
| <i>k</i> | : jumlah pohon keputusan |

2.8 Evaluasi Performa Model

Evaluasi performa model merupakan langkah untuk menilai tingkat akurasi model *machine learning* dalam menghasilkan prediksi berdasarkan data yang digunakan. Evaluasi ini penting untuk memahami sejauh mana

keakuratan hasil prediksi yang diperoleh dari suatu model. Ada beberapa metrik atau metode evaluasi yang umum digunakan untuk mengukur performa model, namun penelitian ini akan menggunakan tiga metode evaluasi, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2). Dengan menerapkan metode-metode tersebut, peneliti dapat mengukur dan membandingkan performa model dalam memprediksi produksi tanaman pangan di Pulau Kalimantan dengan melihat perbedaan antara nilai prediksi dengan nilai yang sebenarnya.

2.8.1 *Mean Absolute Error* (MAE)

Mean Absolute Error (MAE) merupakan metrik evaluasi yang mengukur rata-rata nilai absolut dari selisih antara nilai aktual dan nilai prediksi. Metrik ini mengukur seberapa besar kesalahan prediksi dalam satuan yang sama dengan data. Semakin kecil nilai MAE, maka selisih antara nilai prediksi dengan nilai sebenarnya semakin kecil yang menunjukkan bahwa model mempunyai akurasi prediksi yang semakin tinggi. Formula MAE dapat dinyatakan sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Keterangan:

- n : jumlah data
- y_i : nilai aktual
- \hat{y}_i : nilai prediksi

2.8.2 *Root Mean Square Error* (RMSE)

Root Mean Squared Error (RMSE) adalah akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi. RMSE digunakan untuk mengukur seberapa besar kesalahan prediksi model dalam satuan yang sama dengan variabel target. Metrik ini lebih mudah diinterpretasikan karena satunya sama dengan data yang dievaluasi. Semakin rendah nilai RMSE, semakin baik model dalam

memprediksi nilai yang mendekati hasil aktual (Wang, 2024). Formula RMSE dapat dinyatakan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Keterangan:

- n : jumlah data
- y_i : nilai aktual
- \hat{y}_i : nilai prediksi

2.8.3 *R-Squared (R²)*

R-Squared (atau koefisien determinasi) adalah metrik evaluasi yang digunakan untuk menilai kualitas suatu model regresi. R^2 bernilai antara 0 dan 1, di mana nilai yang mendekati 1 berarti model sangat baik dalam memprediksi data, dan nilai yang mendekati 0 menunjukkan bahwa model kurang baik (Pahlevi *et al.*, 2023). Formula R^2 dapat dinyatakan sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Keterangan:

- n : jumlah data
- y_i : nilai aktual
- \hat{y}_i : nilai prediksi
- \bar{y} : rata-rata dari nilai aktual

2.9 Penelitian Relevan

Beberapa penelitian terdahulu yang relevan dengan penelitian yang dilakukan dalam penelitian ini, antara lain:

- 1) Penelitian oleh Satria *et al.*, 2023 yang berjudul “Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning”

melakukan prediksi hasil panen beberapa komoditas pangan, yaitu padi, jagung, kacang tanah, kedelai, ubi kayu, dan ubi jalar di Pulau Sumatera. Penelitian ini bertujuan untuk mengetahui algoritma terbaik dalam melakukan prediksi dengan membandingkan beberapa algoritma *machine learning*, yaitu *Random Forest* (RF), *Decision Tree* (DT), *Gradient Boosting* (GB), *Extra Tree* (ET), *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN). Berdasarkan hasil evaluasi, disimpulkan bahwa algoritma *Extra Tree* merupakan model terbaik dengan akurasi 0,968 untuk dataset padi dan 0,913 untuk dataset tanaman pangan.

- 2) Penelitian oleh Wijaya dan Fauziah, 2024 dengan judul “*A Comparative Analysis Of Linear Regression And Random Forest Methods For Predicting Palm Oil Price Sales*” membandingkan algoritma *Linear Regression* dan *Random Forest* untuk memprediksi harga penjualan *Crude Palm Oil* (CPO) menggunakan data historis transaksi penjualan dari 2018 sampai 2020. Penelitian ini membagi data latih dan data uji menjadi tiga skema: 90:10, 80:20, dan 70:30. Pada setiap skema pengujian, model Regresi Linier menghasilkan nilai MAE, MSE, dan RMSE yang lebih rendah daripada model Random Forest. Skema pengujian 80:20 menghasilkan nilai MAE, MSE, dan RMSE terendah untuk Regresi Linear. Berdasarkan hasil pengujian menunjukkan bahwa Regresi Linier lebih efektif dibandingkan Random Forest.
- 3) Penelitian oleh Gao, 2024 dengan judul “*The Prediction Of Apple Stock Price based on Linear Regression Model and Random Forest Model*” membandingkan algoritma *Linear Regression* dan *Random Forest* dalam memprediksi harga saham perusahaan Apple menggunakan data dari tahun 2018 hingga 2023. Berdasarkan nilai evaluasi MAE, MSE, dan RMSE, ditemukan bahwa algoritma *Linear Regression* menghasilkan nilai evaluasi yang lebih kecil dibandingkan *Random Forest*, yaitu MAE 2.0988, MSE 7.9001, dan RMSE 2.8017. Penelitian ini menyimpulkan bahwa algoritma *Linear Regression* adalah yang terbaik.

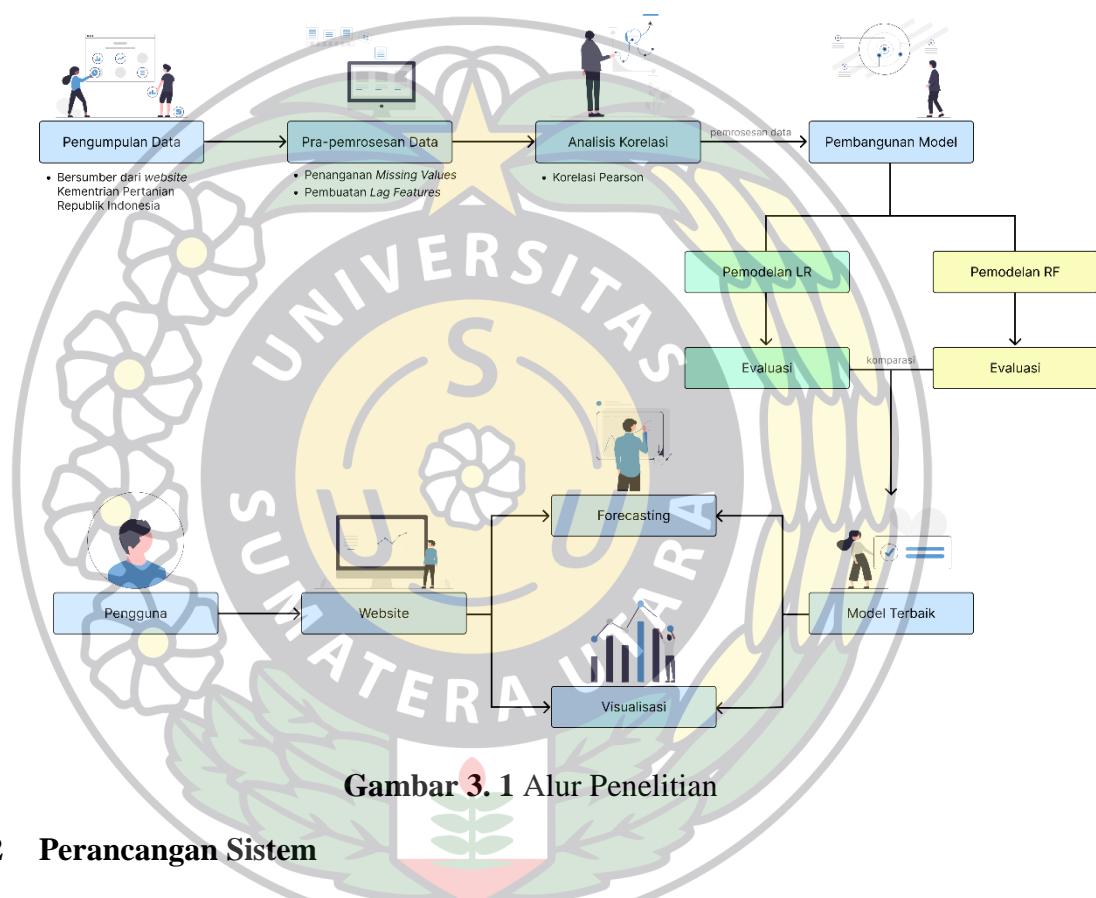
- 4) Penelitian oleh Pashankar *et al.*, 2023 dengan judul “*Machine Learning Techniques for Stock Price Prediction-A Comparative Analysis of Linear Regression, Random Forest, and Support Vector Regression*” memprediksi harga saham dari perusahaan *Tata Consultancy Services* (TCS) dengan membandingkan algoritma *machine learning*, yaitu *Linear Regression*, *Random Forest*, dan *Support Vector Regression* (SVR). Penelitian ini menggunakan data dari Januari 2005 hingga Januari 2024. Dengan membandingkan nilai evaluasi MSE dan RMSE, penelitian ini menyimpulkan bahwa algoritma *Linear Regression* merupakan algoritma dengan kinerja terbaik yang menghasilkan nilai MSE sebesar 74.92 dan RMSE sebesar 8.66.
- 5) Penelitian oleh Hu, 2024 yang berjudul “*Comparison and Analysis of the Effectiveness of Linear Regression, Decision Tree, and Random Forest Models for Health Insurance Premium Forecasting*” membandingkan efektivitas algoritma *Linear Regression*, *Decision Tree*, dan *Random Forest* dalam memprediksi biaya asuransi kesehatan menggunakan dataset dari Kaggle yang berisi 1338 informasi pelanggan dan biaya asuransi kesehatan mereka. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi sebesar 0.8564, diikuti oleh *Linear Regression* sebesar 0.7584, dan *Decision Tree* sebesar 0.7097.

BAB III

ANALISIS DAN PERANCANGAN

3.1 Alur Penelitian

Tahapan alur penelitian dan proses perancangan pembangunan model dan sistem dapat dilihat pada Gambar 3.1 berikut ini.

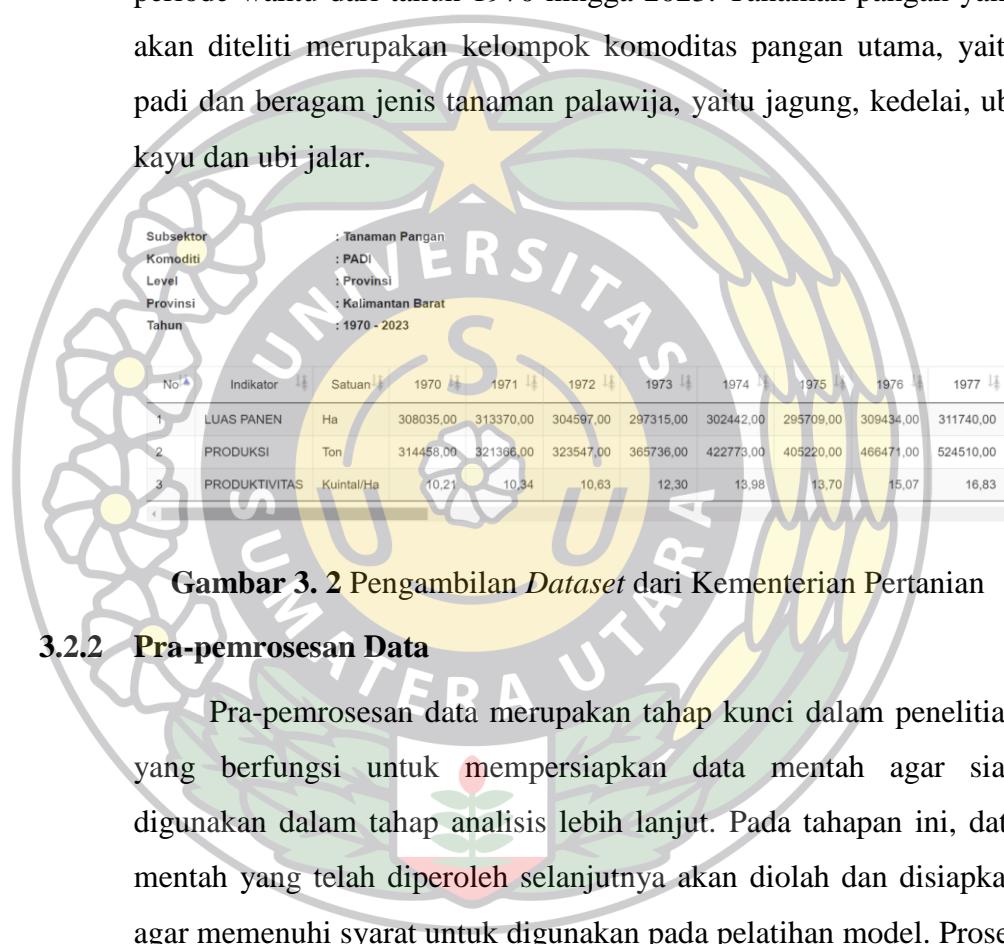


3.2 Perancangan Sistem

Perancangan sistem disusun berdasarkan alur penelitian yang terlihat pada Gambar 3.1. Berikut adalah tahapan dan proses dalam perancangan sistem prediksi produksi tanaman pangan di Pulau Kalimantan menggunakan algoritma *Linear Regression* dan *Random Forest* berdasarkan analisis korelasi luas panen.

3.2.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari situs resmi Kementerian Pertanian Republik Indonesia. *Dataset* ini merupakan data deret waktu (*time series*) yang terdiri dari data luas panen dan jumlah produksi tanaman pangan untuk wilayah Pulau Kalimantan. *Dataset* ini mencakup periode waktu dari tahun 1970 hingga 2023. Tanaman pangan yang akan diteliti merupakan kelompok komoditas pangan utama, yaitu padi dan beragam jenis tanaman palawija, yaitu jagung, kedelai, ubi kayu dan ubi jalar.



Subsektor	: Tanaman Pangan									
Komoditi	: PADI									
Level	: Provinsi									
Provinsi	: Kalimantan Barat									
Tahun	: 1970 - 2023									
No	Indikator	Satuan	1970	1971	1972	1973	1974	1975	1976	1977
1	LUAS PANEN	Ha	308035,00	313370,00	304597,00	297315,00	302442,00	295709,00	309434,00	311740,00
2	PRODUKSI	Ton	314458,00	321306,00	323547,00	365736,00	422773,00	405220,00	466471,00	524510,00
3	PRODUKTIVITAS	Kuintal/Ha	10,21	10,34	10,63	12,30	13,98	13,70	15,07	16,83

Gambar 3. 2 Pengambilan *Dataset* dari Kementerian Pertanian

3.2.2 Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahap kunci dalam penelitian yang berfungsi untuk mempersiapkan data mentah agar siap digunakan dalam tahap analisis lebih lanjut. Pada tahapan ini, data mentah yang telah diperoleh selanjutnya akan diolah dan disiapkan agar memenuhi syarat untuk digunakan pada pelatihan model. Proses ini bertujuan untuk mengatasi masalah-masalah dalam data seperti inkonsistensi, *noise* atau *missing values*, serta untuk mengekstrak informasi yang lebih bermakna melalui transformasi data tertentu.

Tahap pra-pemrosesan data dalam penelitian ini mencakup dua langkah utama, yaitu penanganan *missing values* dan pembuatan *lag features* (fitur lag). Pada langkah pertama, penanganan *missing values* dilakukan untuk memastikan bahwa data yang digunakan tidak

memiliki nilai yang hilang atau tidak sah (*valid*), yang dalam *dataset* penelitian ini direpresentasikan sebagai nilai 0. Metode yang digunakan untuk menangani *missing values* melibatkan penggantian nilai 0 dengan estimasi yang lebih akurat, seperti menggunakan rata-rata (*mean*) atau metode interpolasi.

Langkah kedua adalah pembuatan *lag features*, yaitu menambahkan fitur baru ke dalam *dataset* yang merepresentasikan nilai-nilai historis dari variabel target atau variabel lainnya. Pembuatan *lag features* ini bertujuan untuk menangkap pola temporal atau keterkaitan antar waktu dalam data, yang sangat penting dalam konteks peramalan deret waktu, karena nilai pada suatu waktu dapat dipengaruhi oleh nilai pada waktu sebelumnya. Kedua langkah utama pada tahapan pra-pemrosesan ini dilakukan untuk memastikan data dapat diproses dengan benar oleh algoritma prediksi yang digunakan, sehingga model dapat mempelajari pola dalam data secara lebih akurat yang dapat meningkatkan akurasi model.

3.2.3 Analisis Korelasi

Dalam penelitian ini, tahapan analisis korelasi bertujuan untuk mengidentifikasi hubungan antara variabel luas panen dengan jumlah produksi tanaman pangan di Kalimantan. Metode yang diterapkan adalah korelasi *Pearson*, yang berfungsi untuk menilai kekuatan serta arah hubungan linear antara dua variabel kuantitatif. Dengan penerapan metode ini, akan dihasilkan koefisien korelasi *Pearson* yang nilainya berada dalam rentang antara -1 hingga 1. Koefisien yang mendekati 1 menunjukkan adanya hubungan positif yang kuat, yang berarti ketika luas panen bertambah, produksi cenderung meningkat. Sebaliknya, jika koefisien mendekati -1, hal ini mengindikasikan hubungan negatif yang kuat, di mana peningkatan luas panen diikuti dengan penurunan produksi. Sedangkan, jika nilai koefisien mendekati 0, maka hubungan linear antara kedua variabel tersebut

dapat dianggap tidak signifikan. Analisis ini penting untuk mengidentifikasi seberapa besar pengaruh luas panen terhadap produksi dan bagaimana hubungan ini dapat digunakan untuk meningkatkan akurasi model prediksi di tahap berikutnya.

3.2.4 Pembangunan Model

Tahapan pembangunan model dalam penelitian ini melibatkan beberapa proses penting untuk mempersiapkan dan melatih model prediktif dengan menggunakan dua algoritma, yaitu *Linear Regression* dan *Random Forest*. Berikut adalah langkah-langkah yang dilakukan pada tahapan ini:

a) Pemisahan Fitur dan Target

Pada langkah pertama, *dataset* dibagi menjadi data fitur dan target. Fitur mencakup variabel-variabel yang akan digunakan untuk memprediksi target, yaitu luas panen dan *lag features* yang telah dibuat. *Lag features* mencakup nilai-nilai historis dari variabel produksi yang bertujuan untuk menangkap pola temporal dalam data. Target yang diprediksi adalah produksi tanaman pangan. Pemisahan ini penting agar model dapat mempelajari hubungan antara variabel-variabel *input* (fitur) dan variabel *output* (target).

b) Pelatihan dan Evaluasi Model *Linear Regression*

Pada tahap ini, model *Linear Regression* dilatih menggunakan data yang telah dipersiapkan. Setelah proses pelatihan selesai, model kemudian dievaluasi menggunakan metrik evaluasi. Beberapa metrik evaluasi yang akan digunakan untuk menilai kinerja model adalah MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), dan R² (*R-Squared*). Metrik ini berfungsi untuk menilai sejauh mana model *Linear Regression* dapat memprediksi produksi tanaman pangan berdasarkan luas panen dan *lag features*.

c) Pelatihan dan Evaluasi Model *Random Forest*

Setelah model *Linear Regression* dilatih dan dievaluasi, langkah selanjutnya adalah melatih model *Random Forest*. Sama seperti *Linear Regression*, model *Random Forest* juga dievaluasi menggunakan metrik evaluasi yang sama, yaitu MAE, RMSE, dan R^2 . Dengan mengevaluasi model ini, dapat dilihat seberapa baik model *Random Forest* menangkap pola dalam data.

Tahapan pembangunan model secara keseluruhan dirancang untuk memastikan bahwa model yang dihasilkan memiliki akurasi tinggi dan mampu memprediksi produksi tanaman pangan dengan baik, menggunakan kombinasi antara variabel luas panen dan *lag features* sebagai input.

3.2.5 Perbandingan Evaluasi Kinerja Model

Setelah model dibangun, tahapan selanjutnya yang akan dilakukan ialah membandingkan evaluasi kinerja model. Kinerja kedua model, yaitu model yang menggunakan algoritma *Linear Regression* dan *Random Forest* akan dievaluasi menggunakan beberapa metrik evaluasi, yaitu MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), dan R^2 (*R-Squared*). Metrik ini digunakan untuk mengukur tingkat kesalahan dari prediksi yang dihasilkan oleh masing-masing model. Evaluasi kinerja dari kedua model kemudian akan dicatat dan dibandingkan. Hasil perbandingan kinerja model akan memberikan wawasan yang jelas mengenai performa setiap model dalam memprediksi produksi tanaman pangan. Jika salah satu model menunjukkan nilai evaluasi yang lebih rendah dibandingkan model lainnya, model tersebut akan dianggap lebih baik dan lebih dapat diandalkan untuk diterapkan dalam konteks peramalan produksi.

3.2.6 Visualisasi Hasil Prediksi pada Aplikasi Berbasis *Website*

Setelah model dengan kinerja terbaik dipilih, langkah selanjutnya adalah membuat visualisasi hasil prediksi menggunakan model dengan kinerja terbaik tersebut pada aplikasi berbasis *website*. Implementasi dilakukan menggunakan bahasa pemrograman *Python* dan *framework Streamlit*, yang memudahkan pembuatan aplikasi *web* dengan antarmuka yang sederhana dan interaktif. *Website* yang dibangun memiliki kemampuan untuk menerima *input* dari pengguna, yang meliputi pilihan wilayah provinsi dan jenis tanaman. Setelah pengguna memberikan input, *website* akan menampilkan hasil prediksi produksi tanaman pangan untuk wilayah provinsi dan jenis tanaman yang telah dipilih.



BAB IV

IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Sistem

Penelitian ini membutuhkan beberapa komponen dan perangkat pendukung untuk melakukan pemanggilan dan pra-pemrosesan data, pembangunan model *Linear Regression* dan *Random Forest*, evaluasi performa model, serta penerapannya pada aplikasi berbasis *website*. Berikut adalah spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam penelitian ini:

4.1.1 Spesifikasi Perangkat Keras

Berikut spesifikasi perangkat keras yang dipakai pada penelitian ini:

- 1) Processor AMD Athlon Gold 3150U 2.40 GHz
- 2) AMD Radeon™ Graphics
- 3) RAM 8GB

4.1.2 Spesifikasi Perangkat Lunak

Berikut spesifikasi perangkat lunak yang dipakai pada penelitian ini:

- 1) Sistem operasi *Windows 10 Home Single Language 64-bit*
- 2) *Google Colaboratory*
- 3) *Python 3.10.12*
- 4) *Library pandas, numpy, matplotlib, scikit-learn, math, seaborn*
- 5) *Framework Streamlit*
- 6) *Visual Studio Code*
- 7) *Browser Google Chrome*

4.2 Implementasi Tahap Pengumpulan Data

Dalam penelitian ini, *dataset* diperoleh dari *website* resmi Kementerian Pertanian Republik Indonesia, yang dapat diakses melalui <https://bdsp2.pertanian.go.id/bdsp/>. *Dataset* ini mencakup data dari 5 provinsi di Pulau Kalimantan, meliputi data luas panen dalam satuan hektar (Ha) dan

jumlah produksi dalam satuan ton dengan rentang waktu dari tahun 1970 hingga 2023. Jenis tanaman yang diteliti mencakup padi serta beragam jenis tanaman palawija, yaitu jagung, kedelai, ubi kayu, dan ubi jalar. Secara keseluruhan, penelitian ini mengumpulkan sebanyak 25 *dataset* dengan format data CSV (*Comma-Separated Values*) yang kemudian disimpan di *Google Drive* untuk memudahkan akses dan kolaborasi selama proses penelitian berlangsung.

Dataset untuk provinsi Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Timur mencakup data dari tahun 1970 hingga 2023 untuk tanaman padi dan jagung. Namun, untuk tanaman kedelai, ubi kayu, dan ubi jalar di keempat provinsi tersebut, data hanya tersedia hingga tahun 2022. Khusus untuk tanaman kedelai di Kalimantan Tengah, data yang tersedia dimulai dari tahun 1978. Berbeda dengan provinsi lainnya, Kalimantan Utara yang merupakan hasil pemekaran dari Provinsi Kalimantan Timur dan diresmikan pada tahun 2012, memiliki cakupan data yang lebih terbatas. Data untuk tanaman padi dan jagung tersedia dari tahun 2015 hingga 2023, untuk tanaman kedelai hingga tahun 2021, sedangkan untuk ubi kayu dan ubi jalar tersedia hingga tahun 2022. Berikut adalah gambaran dari isi *dataset* yang akan digunakan:

Tabel 4. 1 Gambaran isi *Dataset* yang Digunakan

Tahun	Luas Panen	Produksi
1970	308035	314458
1971	313370	321366
1972	304597	323547
1973	297315	365736
1974	302442	422773
1975	295709	405220
...
2022	241478	731225
2023	224068	700290

Untuk melihat keseluruhan *dataset* yang digunakan dalam penelitian ini, Anda dapat mengunjungi tautan berikut:

https://drive.google.com/drive/folders/1JkqfuPwHmp_hb6efI3NfSNVCmoGZLEJm.

4.3 Implementasi Tahap Pra-pemrosesan Data

Tahap pra-pemrosesan *dataset* merupakan langkah penting untuk mempersiapkan data mentah agar dapat digunakan pada pelatihan model. Pada penelitian ini, proses pra-pemrosesan *dataset* terdiri dari beberapa tahap, dimulai dari pemuatan data (*data loading*), penanganan nilai yang hilang atau tidak valid (*missing values*), pembuatan fitur lag (*lag features*), dan diakhiri dengan visualisasi *dataset* untuk menemukan pola, tren, dan hubungan antar variabel. Gambar 4.1 di bawah ini menunjukkan beberapa *library* yang digunakan pada tahap pra-pemrosesan data.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Gambar 4.1 Import *Library* untuk Proses Pra-pemrosesan Data

Adapun fungsi dari *library-library* yang digunakan ialah:

- a) *NumPy*: *Library* untuk operasi matematika dan komputasi ilmiah, terutama manipulasi *array/matriks*.
- b) *Matplotlib*: *Library* untuk membuat grafik dan visualisasi data.
- c) *Pandas*: *Library* untuk manipulasi dan analisis data tabular, seperti CSV.

4.3.1 *Data Loading*

Tahapan awal yang dilakukan ialah memuat data (*data loading*) dengan mengakses tautan *dataset* dari *Google Drive* menggunakan *library pandas*.

```
#load dataset
url = 'https://drive.google.com/uc?id=1g4rFyt6q3dATvqW_4y4V0Tf3LSfN4h1U'
dataset = pd.read_csv(url)

dataset.head()
```

	tahun	luas panen	produksi
0	1970	114946	131008
1	1971	101857	133017
2	1972	94676	125954
3	1973	117950	140448
4	1974	111021	156292

Gambar 4. 2 Data Loading

Gambar 4.2 menunjukkan proses pemanggilan serta pemuatan *dataset* dengan menggunakan fungsi `read_csv` yang terdapat pada *library pandas*. *Dataset* yang telah dibaca kemudian disimpan pada variabel *dataset* dalam bentuk *DataFrame*. Fungsi `head()` digunakan untuk melihat lima entri pertama dalam *dataset*. Untuk memperoleh informasi detail *dataset*, dapat memanggil fungsi `info()` seperti yang ditunjukkan Gambar 4.3 berikut.

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54 entries, 0 to 53
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   tahun        54 non-null    int64  
 1   luas panen  54 non-null    object  
 2   produksi     54 non-null    object  
dtypes: int64(1), object(2)
memory usage: 1.4+ KB
```

Gambar 4. 3 Informasi Struktur Awal Dataset

Gambar 4.3 di atas menampilkan rincian data pada salah satu *dataset* yang digunakan dalam penelitian ini. Berdasarkan informasi tersebut dapat diketahui bahwa *dataset* tersebut memiliki 54 baris data yang terdiri dari 3 kolom, yaitu tahun, luas panen, dan produksi.

Kolom tahun bertipe data *integer*, sedangkan kolom luas panen dan produksi bertipe data *object*. Sebelum melanjutkan ke proses analisis dan pemodelan data, ada beberapa langkah yang perlu dilakukan untuk memastikan *dataset* siap digunakan. Langkah pertama adalah menjadikan kolom tahun sebagai *index* pada *DataFrame*. Hal ini penting dikarenakan tahun merupakan variabel waktu, yang sering kali digunakan sebagai acuan dalam analisis *time series*.

Selain itu, tipe data kolom luas panen dan produksi juga perlu diubah menjadi *float*. Tipe data *float* memungkinkan untuk melakukan operasi numerik dan statistik dengan benar, seperti korelasi atau pemodelan menggunakan algoritma *machine learning* seperti *Linear Regression* dan *Random Forest*. Mengubah kedua kolom tersebut menjadi *float* memastikan bahwa *dataset* siap untuk pemrosesan lebih lanjut, dan mencegah terjadinya error atau hasil yang tidak akurat akibat penggunaan tipe data yang tidak sesuai. Langkah-langkah ini dapat diikuti seperti yang ditunjukkan pada Gambar 4.4 berikut.

```
# Mengubah kolom 'tahun' menjadi index
dataset.set_index('tahun', inplace=True)
dataset.index.name = 'Y'

# Menghapus karakter non-numerik seperti koma dan spasi
dataset['luas panen'] = dataset['luas panen'].str.replace('.', ',').str.replace(',', '.')
dataset['produksi'] = dataset['produksi'].str.replace('.', ',').str.replace(',', '.')

# Mengubah kolom 'luas panen' dan 'produksi' menjadi float
dataset['luas panen'] = dataset['luas panen'].astype(float)
dataset['produksi'] = dataset['produksi'].astype(float)
```

Gambar 4.4 Proses Konversi Tipe Data pada *Dataset*

4.3.2 Penanganan *Missing Values*

Tahapan selanjutnya adalah penanganan *missing values* yang melibatkan identifikasi dan penanganan nilai-nilai yang hilang atau tidak *valid* dalam *dataset*. Nilai yang hilang dalam penelitian ini direpresentasikan sebagai nilai 0. Untuk melihat jumlah nilai yang

hilang pada *dataset* dapat dilakukan langkah seperti Gambar 4.5 berikut.

```
luas_panen = (dataset['luas panen'] == 0).sum()
produksi = (dataset['produksi'] == 0).sum()

print('Nilai 0 di kolom luas panen ada: ', luas_panen)
print('Nilai 0 di kolom produksi ada: ', produksi)

Nilai 0 di kolom luas panen ada:  1
Nilai 0 di kolom produksi ada:  1
```

Gambar 4. 5 Identifikasi Jumlah Nilai yang Hilang

Berdasarkan Gambar 4.5, terlihat bahwa salah satu *dataset* yang diuji memiliki nilai yang hilang, baik pada kolom luas panen maupun produksi. Salah satu pendekatan untuk menangani nilai yang hilang adalah dengan menerapkan metode interpolasi linear. Pada interpolasi linear, nilai yang hilang dihitung dengan menghubungkan dua titik data terdekat yang tidak hilang dengan sebuah garis lurus, lalu menghitung nilai titik yang hilang berdasarkan posisi relatifnya di antara dua titik tersebut. Penerapan metode interpolasi linear dapat dilihat pada Gambar 4.6.

```
# Mengganti nilai 0 dengan NaN (agar bisa diimputasi)
dataset.replace(0, np.nan, inplace=True)

# Mengisi missing values dengan melakukan interpolasi linear
dataset = dataset.interpolate(method='linear')
```

Gambar 4. 6 Penanganan Data Hilang dengan Interpolasi Linear

4.3.3 Pembuatan *Lag Features*

Pada tahap ini akan dilakukan pembuatan fitur *lag* (*lag features*) yang merupakan langkah penting dalam analisis data *time series*. Langkah ini memungkinkan model untuk memanfaatkan informasi historis dalam membuat prediksi. Proses ini dilakukan dengan menggeser nilai-nilai dalam kolom variabel yang digunakan berdasarkan jumlah *lag* yang ditentukan. Penelitian ini akan membuat

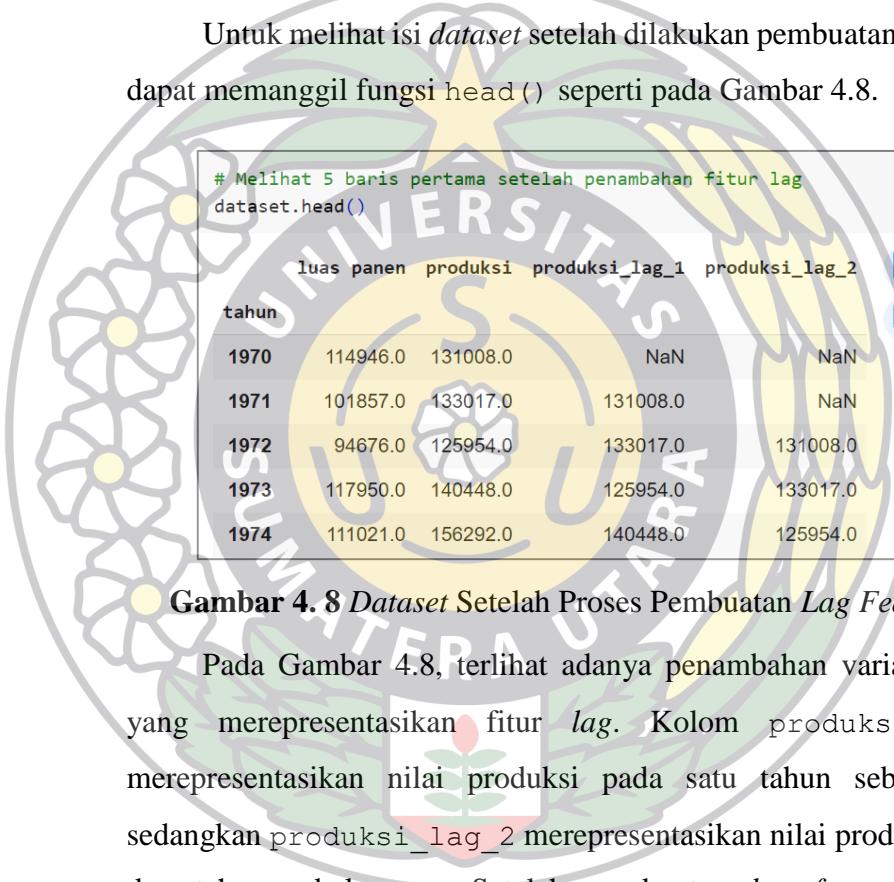
dua *lag* untuk variabel target, yaitu produksi. Penerapan pembuatan *lag features* dapat dilihat pada Gambar 4.7 berikut.

```
# Menentukan jumlah lag
lag_days = 2

# Membuat lag features produksi
for lag in range(1, lag_days + 1):
    dataset[f'produksi_lag_{lag}'] = dataset['produksi'].shift(lag)
```

Gambar 4.7 Pembuatan *Lag Features* Produksi

Untuk melihat isi *dataset* setelah dilakukan pembuatan fitur *lag*, dapat memanggil fungsi `head()` seperti pada Gambar 4.8.



tahun	luas panen	produksi	produksi_lag_1	produksi_lag_2
1970	114946.0	131008.0	NaN	NaN
1971	101857.0	133017.0	131008.0	NaN
1972	94676.0	125954.0	133017.0	131008.0
1973	117950.0	140448.0	125954.0	133017.0
1974	111021.0	156292.0	140448.0	125954.0

Gambar 4.8 Dataset Setelah Proses Pembuatan *Lag Features*

Pada Gambar 4.8, terlihat adanya penambahan variabel baru yang merepresentasikan fitur *lag*. Kolom *produksi_lag_1* merepresentasikan nilai produksi pada satu tahun sebelumnya, sedangkan *produksi_lag_2* merepresentasikan nilai produksi pada dua tahun sebelumnya. Setelah pembuatan *lag features*, dapat ditemukan nilai *NaN* (*Not a Number*) pada baris awal, di mana tidak ada cukup data sebelumnya untuk mengisi nilai *lag*. Nilai-nilai ini dapat ditangani dengan beragam metode, tetapi dalam penelitian ini, nilai tersebut akan dihapus, sebagaimana yang diperlihatkan pada Gambar 4.9.

```
# Menghapus baris dengan nilai NaN setelah penambahan lag
dataset.dropna(inplace=True)
```

Gambar 4. 9 Penghapusan Nilai NaN

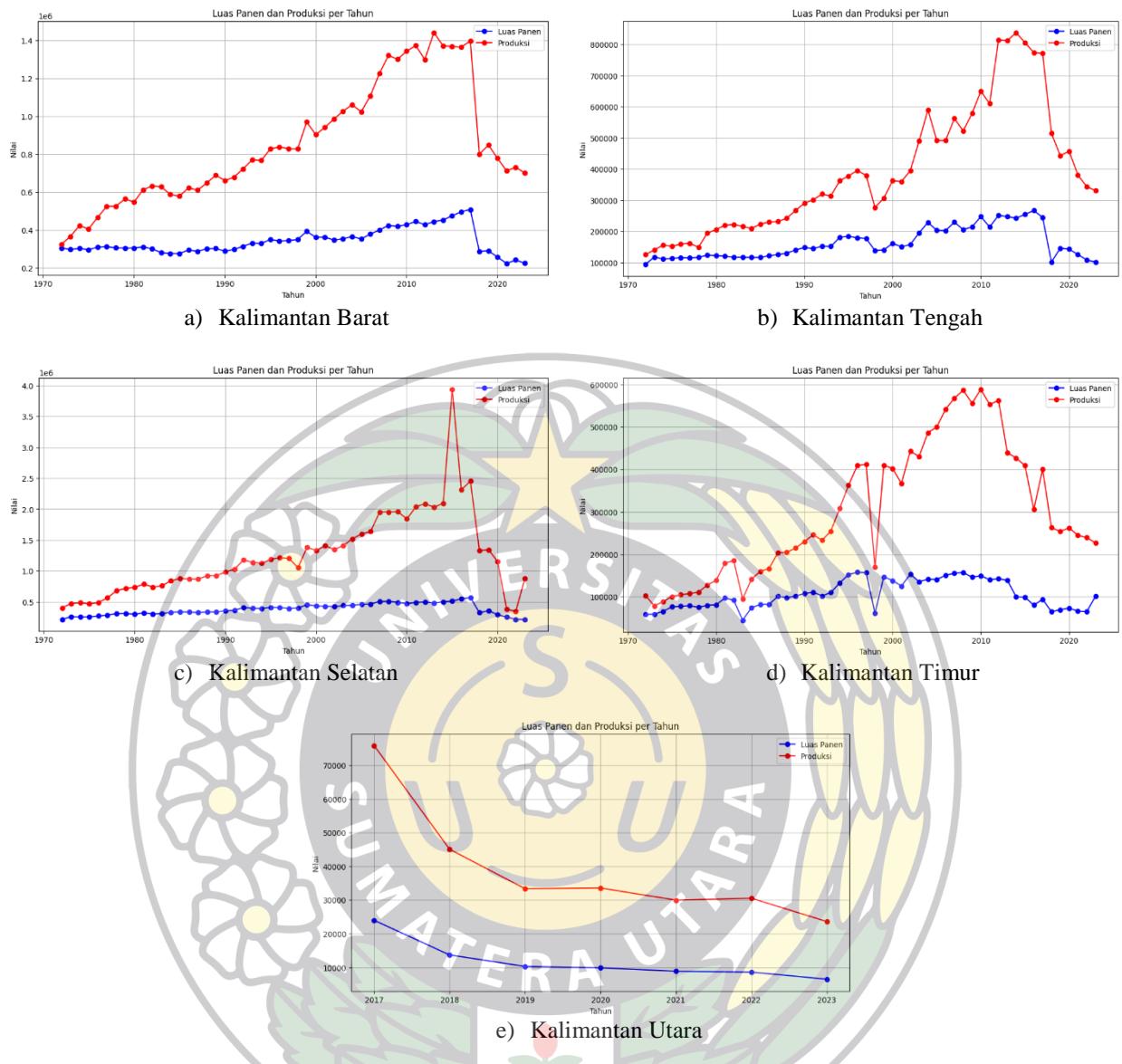
4.4 Visualisasi *Dataset*

Visualisasi *dataset* merupakan langkah penting dalam analisis data *time series* karena dapat membantu dalam memahami tren, pola, dan karakteristik unik dari data. Melalui visualisasi, tren jangka panjang seperti peningkatan atau penurunan dapat terlihat dengan jelas, sehingga dapat membantu untuk memahami bagaimana variabel tertentu berubah seiring waktu. Selain itu, visualisasi mempermudah untuk membandingkan data dari tahun yang berbeda serta memungkinkan untuk melihat korelasi antara nilai-nilai data pada variabel yang dianalisis. Dengan visualisasi, data yang kompleks dan berisi banyak informasi dapat dipresentasikan dengan cara yang lebih sederhana, intuitif, dan informatif, sehingga mempercepat proses pengambilan keputusan dan pemahaman pola dalam data *time series*. Pada penelitian ini, visualisasi *dataset* dibangun menggunakan *library matplotlib* seperti yang ditunjukkan pada Gambar 4.10.

```
# Membuat visualisasi dataset
plt.figure(figsize=(12, 6))
plt.plot(dataset.index, dataset['luas panen'], label='Luas Panen', marker='o', color='b')
plt.plot(dataset.index, dataset['produksi'], label='Produksi', marker='o', color='r')
plt.xlabel('Tahun')
plt.ylabel('Nilai')
plt.title('Luas Panen dan Produksi per Tahun')
plt.legend()
plt.grid(True)
plt.show()
```

Gambar 4. 10 Proses Visualisasi *Dataset*

Di bawah ini ditampilkan hasil visualisasi dari keseluruhan *dataset* yang digunakan dalam penelitian ini.



Gambar 4. 11 Hasil Visualisasi Dataset Tanaman Padi

Gambar 4.11 di atas memperlihatkan grafik yang menggambarkan tren luas panen dan produksi tanaman padi di wilayah Kalimantan. Berikut ini adalah penjelasan dari setiap grafik tersebut:

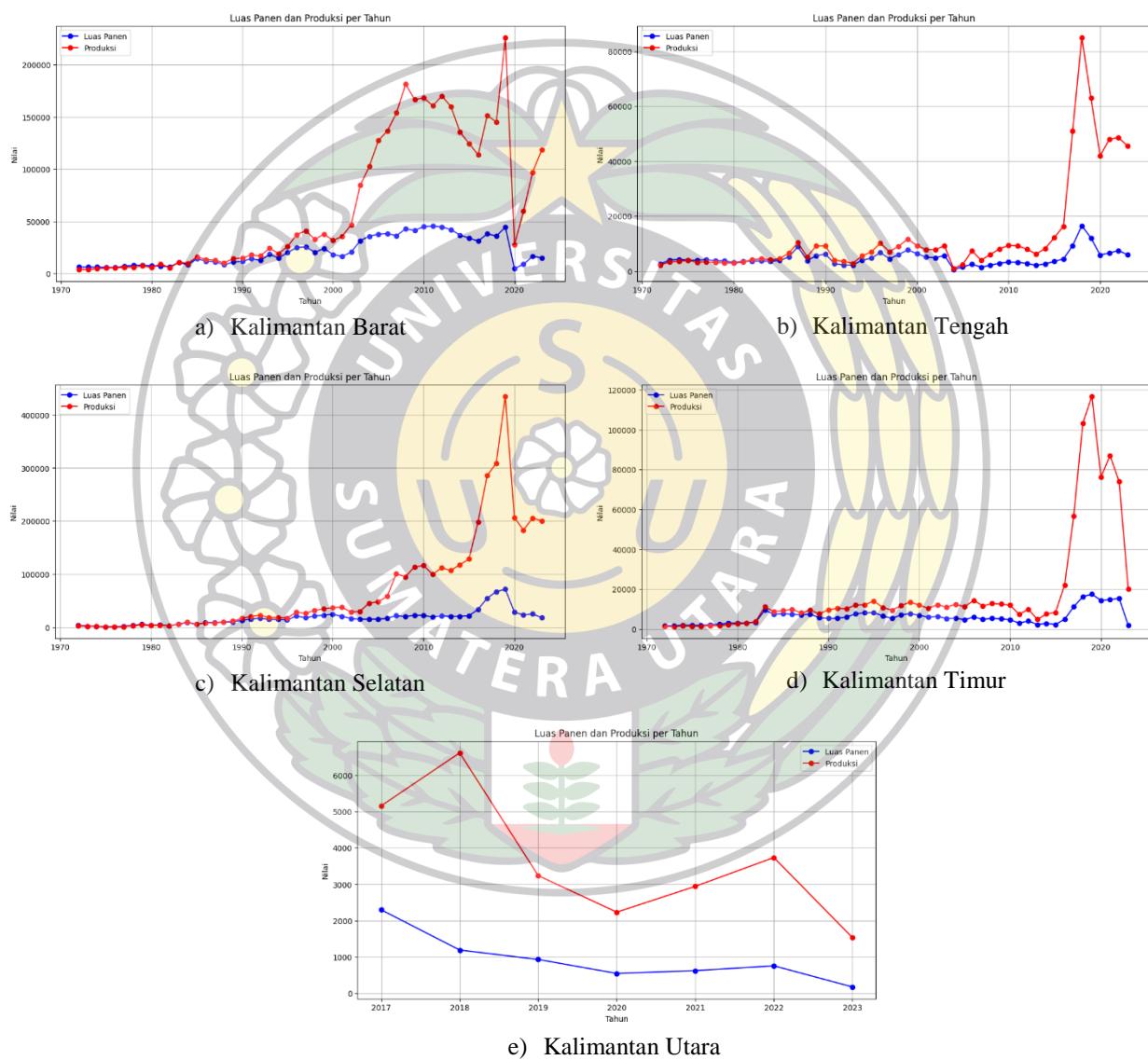
- Kalimantan Barat:** Dari tahun 1970 hingga 2010, baik luas panen maupun produksi meningkat dengan laju yang stabil. Produksi mengalami puncak pada sekitar tahun 2013 dan kemudian menurun secara drastis. Penurunan ini mungkin disebabkan oleh faktor cuaca, hama, atau kebijakan pertanian yang berubah. Luas panen juga

menunjukkan sedikit penurunan setelah puncak produksi, tetapi tetap lebih stabil dibandingkan produksi.

- b) Kalimantan Tengah: Baik luas panen maupun produksi meningkat secara konsisten hingga tahun 2014. Pada tahun 2011-2012 terjadi lonjakan produksi yang sangat besar, yang kemudian menurun dengan cepat pada tahun 2018. Ini bisa menunjukkan adanya anomali atau kejadian tertentu yang mempengaruhi hasil produksi. Setelah puncak tersebut, tren cenderung menurun, baik pada luas panen maupun produksi.
- c) Kalimantan Selatan: Terlihat tren kenaikan yang stabil pada luas panen dan produksi hingga sekitar tahun 2014. Ada peningkatan tajam di produksi pada sekitar tahun 2015, yang kemudian menurun drastis setelahnya. Hal ini mungkin disebabkan oleh peristiwa luar biasa yang memengaruhi produksi. Luas panen relatif stabil dibandingkan produksi, tetapi ada penurunan kecil dalam beberapa tahun terakhir.
- d) Kalimantan Timur: Produksi meningkat dengan stabil dari tahun 1970 hingga sekitar tahun 2010, meskipun terjadi fluktuasi yang cukup besar di sekitar tahun 1997-1999 yang mungkin disebabkan oleh adanya anomali data atau telah terjadi perubahan kebijakan pertanian di wilayah ini. Luas panen menunjukkan pola yang serupa, meskipun tidak sebesar fluktuasi pada produksi. Setelah tahun 2010, baik luas panen maupun produksi terus menunjukkan tren penurunan secara bertahap yang mungkin mengindikasikan adanya penurunan efektivitas hasil panen.
- e) Kalimantan Utara: Data yang tersedia hanya mencakup beberapa tahun terakhir, dengan tren keseluruhan yang memperlihatkan penurunan pada luas panen dan produksi. Pola ini mungkin mencerminkan tantangan yang dihadapi sektor pertanian di wilayah ini.

Secara umum, grafik-grafik ini menunjukkan bahwa fluktuasi produksi lebih besar dibandingkan dengan luas panen di berbagai provinsi. Ini menunjukkan adanya variabel lain yang memengaruhi hasil panen, seperti

kondisi cuaca, teknik pertanian, atau penggunaan pupuk. Beberapa daerah menunjukkan lonjakan produksi yang diikuti penurunan tajam, yang mungkin disebabkan oleh kejadian tertentu seperti peristiwa iklim, bencana alam, atau kebijakan khusus. Tren menurun dalam beberapa tahun terakhir pada beberapa provinsi bisa menunjukkan tantangan berkelanjutan dalam produksi padi atau pergeseran prioritas pertanian di Kalimantan.



Gambar 4. 12 Hasil Visualisasi *Dataset* Tanaman Jagung

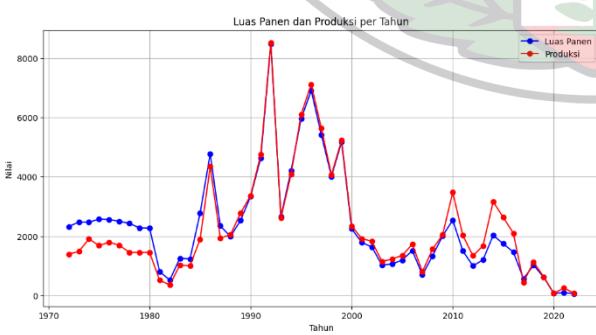
Gambar 4.12 di atas memperlihatkan grafik yang menggambarkan tren luas panen dan produksi tanaman jagung di wilayah Kalimantan. Berikut ini adalah penjelasan dari setiap grafik tersebut:

- a) Kalimantan Barat: Dari tahun 1970 hingga sekitar tahun 2000, baik luas panen maupun produksi menunjukkan pertumbuhan yang lambat dan stabil. Setelah tahun 2000, terdapat peningkatan yang signifikan pada luas panen dan produksi, dengan puncak tertinggi sekitar tahun 2019. Setelah puncak tersebut, terjadi penurunan tajam, terutama pada produksi, yang kemungkinan disebabkan oleh faktor eksternal seperti cuaca, kebijakan, atau masalah dalam pertanian.
- b) Kalimantan Tengah: Baik produksi maupun luas panen menunjukkan tren yang stabil sejak tahun 1970 hingga 2014. Produksi mengalami kenaikan tajam sekitar tahun 2016-2018 yang kemudian diikuti dengan penurunan. Produksi menunjukkan lonjakan yang tidak sepenuhnya sesuai dengan kenaikan luas panen. Penurunan produksi setelah tahun 2018 mengindikasikan terjadinya penurunan produktivitas atau faktor eksternal yang mempengaruhi produksi meskipun luas panen relatif stabil.
- c) Kalimantan Selatan: Ada peningkatan produksi dan luas panen yang stabil hingga sekitar tahun 2019, di mana produksi melonjak secara signifikan yang kemudian diikuti dengan penurunan di tahun setelahnya. Luas panen mengalami peningkatan yang sedikit, sementara produksi berfluktuasi lebih besar. Tren ini menunjukkan tantangan dalam mempertahankan tingkat produksi tinggi, yang mungkin menunjukkan perlunya teknik manajemen tanaman yang lebih baik atau penyesuaian terhadap kondisi yang berubah.
- d) Kalimantan Timur: Produksi dan luas panen mengikuti pola pertumbuhan yang serupa seperti di Kalimantan Tengah dengan puncak tertinggi terjadi sekitar tahun 2019, tetapi kemudian jumlah produksi menurun di tahun-tahun setelahnya. Lonjakan produksi sekitar tahun 2015-2019 yang diikuti dengan penurunan mengindikasikan periode

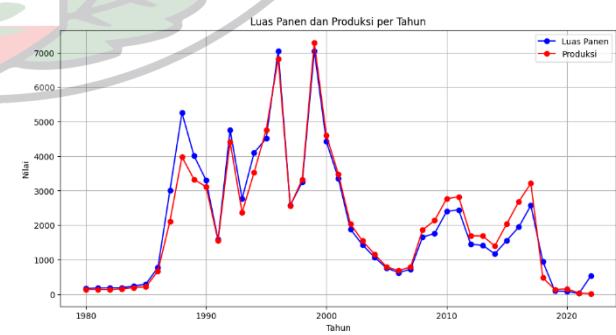
hasil yang luar biasa namun tidak berkelanjutan di tahun-tahun selanjutnya. Kemungkinan, hal ini disebabkan oleh peningkatan sementara dalam upaya pertanian atau kondisi iklim yang menguntungkan.

- e) Kalimantan Utara: Data di Kalimantan Utara hanya mencakup periode 2017 hingga 2023, lebih singkat daripada provinsi lainnya. Terlihat bahwa produksi cenderung fluktuatif selama periode tersebut, dengan penurunan yang signifikan setelah tahun 2019, meskipun luas panen tidak menunjukkan perubahan besar. Ketidakstabilan ini mungkin disebabkan oleh sumber daya yang terbatas atau kondisi tanam yang tidak dapat diprediksi.

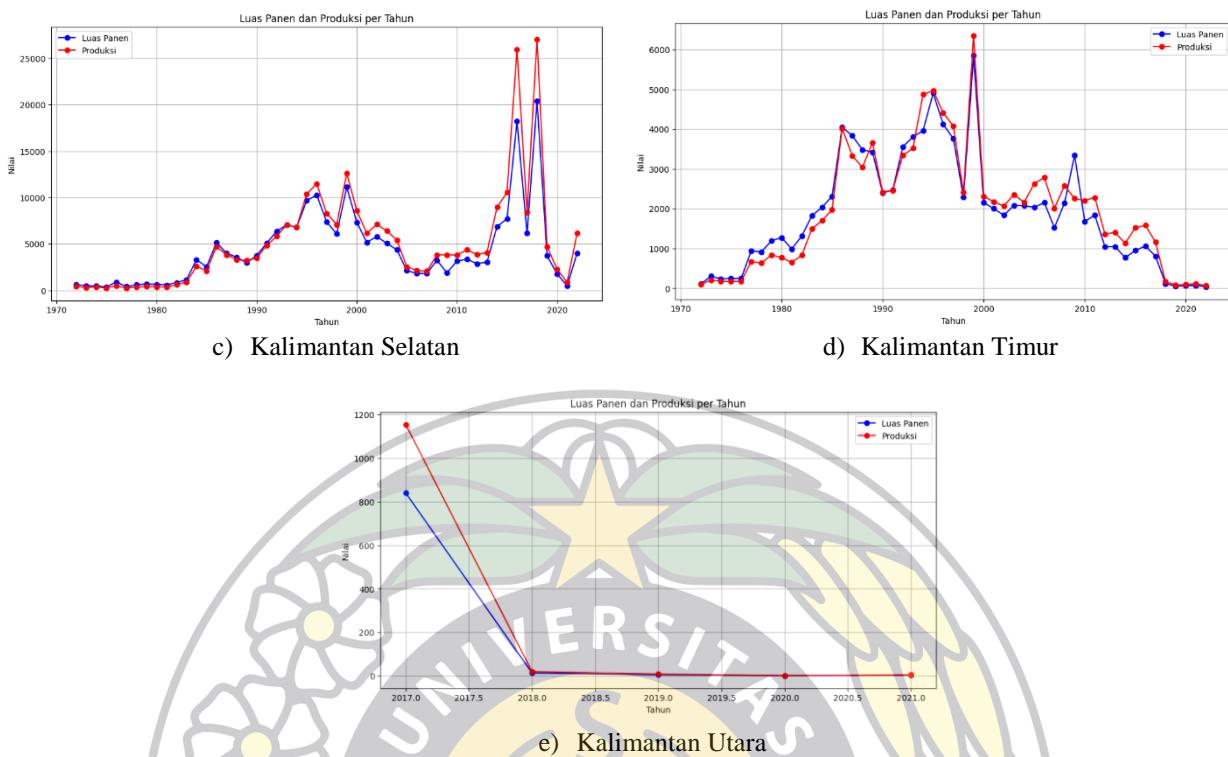
Secara umum, sebagian besar provinsi menunjukkan tren peningkatan luas panen dan produksi jagung yang pesat hingga sekitar tahun 2019 yang kemudian diikuti oleh penurunan yang signifikan pada tahun-tahun selanjutnya. Penurunan ini mungkin dipengaruhi oleh beberapa faktor, seperti perubahan iklim, penurunan kualitas tanah, masalah ekonomi, atau kebijakan pertanian yang berubah. Di sebagian besar wilayah, luas panen menunjukkan tren yang lebih stabil dibandingkan produksi. Ini mengindikasikan bahwa meskipun area lahan untuk budidaya jagung tidak meningkat secara signifikan, variasi jumlah produksi mungkin disebabkan oleh faktor lain, seperti manajemen tanaman, teknologi, atau kondisi iklim.



a) Kalimantan Barat



b) Kalimantan Tengah



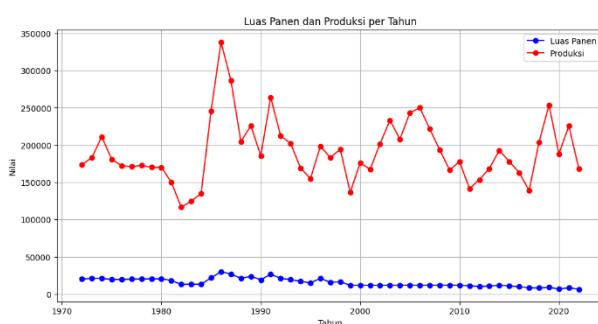
Gambar 4. 13 Hasil Visualisasi Dataset Tanaman Kedelai

Gambar 4.13 di atas memperlihatkan grafik yang menggambarkan tren luas panen dan produksi tanaman kedelai di wilayah Kalimantan. Berikut ini adalah penjelasan dari setiap grafik tersebut:

- Kalimantan Barat: Grafik memperlihatkan fluktuasi yang cukup besar dalam luas panen dan produksi. Kedua variabel ini cenderung berfluktuasi bersamaan. Puncak tertinggi terjadi pada awal 1990-an dan sekitar tahun 2010. Setelah puncak tersebut, terdapat penurunan yang signifikan diikuti oleh fluktuasi hingga 2020. Hubungan antara luas panen dan produksi menunjukkan pola yang serupa, mengindikasikan bahwa produksi sangat bergantung pada luas panen.
- Kalimantan Tengah: Terdapat pola yang mirip dengan Kalimantan Barat, namun dengan nilai yang sedikit lebih rendah. Puncak produksi dan luas panen terjadi pada sekitar tahun 1990 dan awal 2010. Setelah 2010, baik luas panen maupun produksi mengalami penurunan dengan fluktuasi yang lebih kecil hingga 2020.

- c) Kalimantan Selatan: Provinsi ini menunjukkan peningkatan yang lebih stabil pada luas panen dan produksi dibandingkan provinsi lainnya. Puncak besar terjadi pada sekitar tahun 2016 dan 2018, yang kemudian turun tajam. Kalimantan Selatan memiliki variasi data yang lebih ekstrem dibandingkan Kalimantan Barat dan Tengah, terutama dalam beberapa tahun terakhir.
- d) Kalimantan Timur: Tren pada Kalimantan Timur juga menunjukkan fluktuasi, namun dengan pola yang sedikit lebih stabil dari tahun 1980 hingga 2000. Puncak tertinggi terlihat pada awal hingga akhir 1990-an, dan setelah itu ada beberapa fluktuasi hingga mencapai titik rendah pada tahun 2020. Luas panen dan produksi cenderung sejalan, meskipun terdapat beberapa tahun di mana produksi tidak berbanding lurus dengan luas panen.
- e) Kalimantan Utara: Data untuk Kalimantan Utara sangat terbatas dan dimulai pada tahun 2017, sehingga tidak banyak informasi yang dapat diambil. Terlihat penurunan drastis dari tahun 2017 hingga 2018, diikuti dengan nilai yang tetap rendah hingga 2020.

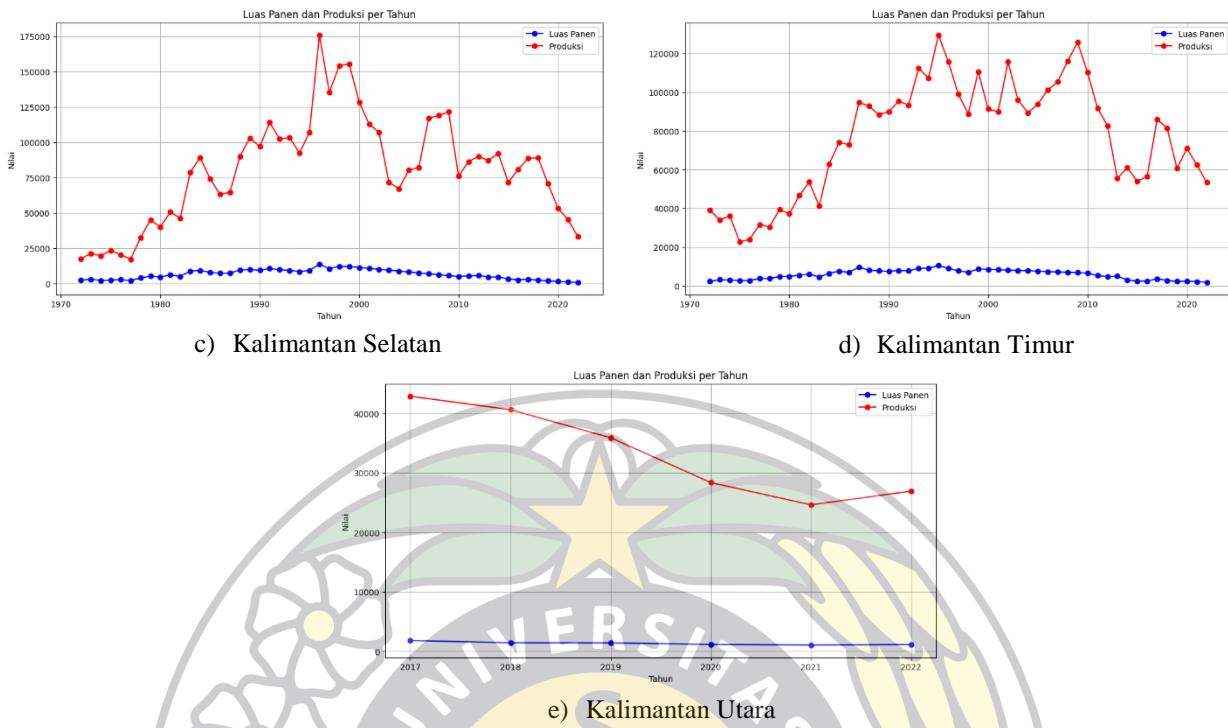
Secara keseluruhan, terdapat korelasi positif antara luas panen dan produksi di seluruh wilayah Kalimantan, yang mana penambahan luas panen biasanya disertai dengan peningkatan jumlah produksi. Fluktuasi yang signifikan terjadi di setiap provinsi, yang mungkin disebabkan oleh faktor cuaca, kebijakan pertanian, atau kondisi lingkungan. Kalimantan Selatan menunjukkan variasi yang paling besar dalam beberapa tahun terakhir, sedangkan Kalimantan Utara memiliki data yang terbatas.



a) Kalimantan Barat



b) Kalimantan Tengah



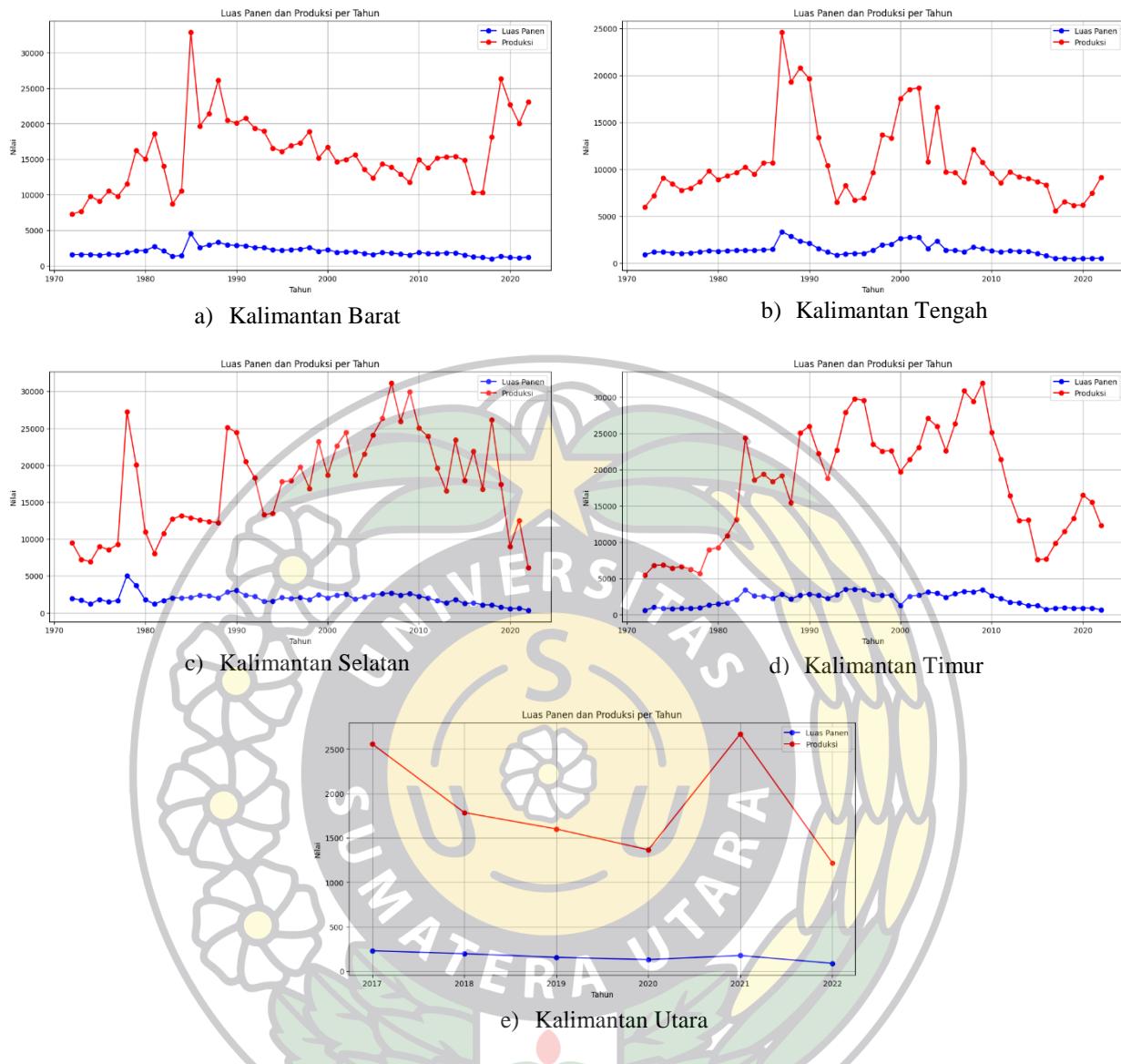
Gambar 4. 14 Hasil Visualisasi Dataset Tanaman Ubi Kayu

Gambar 4.14 di atas memperlihatkan grafik yang menggambarkan tren luas panen dan produksi tanaman ubi kayu di wilayah Kalimantan. Berikut ini adalah penjelasan dari setiap grafik tersebut:

- Kalimantan Barat: Produksi ubi kayu sangat tinggi dibandingkan dengan luas panen, dengan grafik produksi menunjukkan nilai yang sangat fluktuatif. Puncak produksi terjadi pada pertengahan 1980-an, sementara luas panen tetap relatif rendah. Ketidaksesuaian besar antara produksi dan luas panen menunjukkan adanya faktor eksternal, seperti peningkatan efisiensi dari penggunaan teknologi yang meningkatkan hasil panen per hektar.
- Kalimantan Tengah: Produksi ubi kayu di Kalimantan Tengah juga jauh lebih tinggi daripada luas panennya. Puncak besar dalam produksi terjadi pada akhir 1980-an, awal 2000-an dan diakhiri 2010-an. Luas panen menunjukkan nilai yang tetap rendah selama periode tersebut, sehingga peningkatan produksi mungkin disebabkan oleh peningkatan hasil per hektar.

- c) Kalimantan Selatan: Tren yang serupa terlihat di Kalimantan Selatan, dengan produksi ubi kayu yang meningkat dan mencapai puncak pada akhir 1980-an hingga pertengahan 1990-an, kemudian menurun secara bertahap. Luas panen tetap rendah yang mengindikasikan bahwa produksi tidak berbanding lurus dengan luas panen. Penurunan produksi dalam beberapa tahun terakhir mungkin menunjukkan adanya masalah pada faktor eksternal yang memengaruhi hasil per hektar.
- d) Kalimantan Timur: Grafik produksi memperlihatkan fluktuasi signifikan sepanjang periode waktu, dengan puncak produksi terbesar terjadi pada akhir 1980-an hingga akhir 2000-an. Sama seperti provinsi lainnya, luas panen tetap rendah yang menunjukkan bahwa peningkatan produksi bergantung pada hasil per hektar. Setelah puncak pada akhir 2000-an, produksi menunjukkan tren menurun, meskipun masih fluktuatif.
- e) Kalimantan Utara: Data untuk Kalimantan Utara dimulai pada tahun 2017, dengan tren produksi yang menurun setiap tahun. Luas panen tetap sangat rendah, sementara produksi menunjukkan penurunan yang konsisten dari tahun 2017 hingga 2020. Data terbatas membuat sulit untuk menilai tren jangka panjang, tetapi penurunan produksi dapat mengindikasikan adanya kendala dalam budidaya atau hasil per hektar yang menurun.

Semua provinsi menunjukkan produksi ubi kayu yang jauh lebih tinggi daripada luas panen, menunjukkan kemungkinan bahwa hasil per hektar cukup tinggi atau ada peningkatan efisiensi dalam produksi. Luas panen yang relatif rendah dibandingkan produksi menunjukkan bahwa peningkatan produksi sebagian besar berasal dari faktor selain ekspansi lahan, seperti praktik agrikultural yang lebih baik atau penggunaan varietas yang lebih unggul. Ketidakstabilan dalam produksi menunjukkan adanya faktor-faktor eksternal yang mungkin memengaruhi hasil, seperti kondisi cuaca, ketersediaan sumber daya, atau perubahan teknologi.



Gambar 4. 15 Hasil Visualisasi Dataset Tanaman Ubi Jalar

Gambar 4.15 di atas memperlihatkan grafik yang menggambarkan tren luas panen dan produksi tanaman ubi jalar di wilayah Kalimantan. Berikut ini adalah penjelasan dari setiap grafik tersebut:

- Kalimantan Barat: Produksi ubi jalar jauh lebih tinggi dibandingkan luas panen, dengan grafik produksi menunjukkan fluktuasi yang cukup signifikan. Produksi mencapai puncaknya pada pertengahan 1980-an dan akhir 2010-an, meskipun luas panen tetap berada di level rendah. Tidak adanya korelasi langsung antara luas panen dan produksi

menunjukkan bahwa produktivitas per hektar berperan penting dalam meningkatkan produksi.

- b) Kalimantan Tengah: Produksi menunjukkan pola fluktuatif dengan puncak tertinggi pada akhir tahun 1980-an dan terjadi penurunan yang cukup tajam setelahnya. Luas panen tetap pada angka yang rendah sepanjang periode, sama seperti provinsi lainnya. Penurunan produksi di tahun-tahun terakhir bisa menunjukkan adanya penurunan produktivitas, atau kondisi yang memengaruhi hasil per hektar.
- c) Kalimantan Selatan: Tren produksi ubi jalar menunjukkan dua puncak besar, yaitu sekitar akhir tahun 1970-an dan akhir tahun 2000-an. Luas panen tetap stabil dengan fluktuasi kecil sepanjang periode. Perbedaan signifikan antara produksi dan luas panen mengindikasikan bahwa peningkatan dalam hasil per hektar mungkin memiliki peran yang penting.
- d) Kalimantan Timur: Produksi memiliki fluktuasi yang cukup besar, dengan puncak pada akhir tahun 2000-an yang kemudian diikuti oleh penurunan bertahap di tahun-tahun setelahnya. Luas panen tetap rendah yang menunjukkan bahwa perubahan dalam produksi bukan disebabkan oleh ekspansi lahan. Penurunan produksi pada beberapa tahun terakhir bisa menunjukkan penurunan hasil per hektar atau faktor eksternal yang memengaruhi produktivitas.
- e) Kalimantan Utara: Data untuk Kalimantan Utara terbatas, dimulai dari tahun 2017, dengan tren produksi yang menurun secara bertahap hingga tahun 2020. Luas panen tetap rendah dan relatif stabil di sepanjang periode. Penurunan produksi yang terus-menerus dapat menunjukkan adanya masalah dalam faktor-faktor yang memengaruhi hasil per hektar.

Semua provinsi menunjukkan produksi ubi jalar yang jauh lebih tinggi daripada luas panen, menunjukkan bahwa hasil per hektar yang tinggi atau teknologi pertanian mungkin berkontribusi pada produksi yang lebih besar. Stabilitas dalam luas panen menunjukkan bahwa peningkatan produksi tidak

berasal dari ekspansi lahan, melainkan dari produktivitas yang lebih tinggi per hektarnya. Fluktuasi dalam produksi, khususnya di Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan dan Kalimantan Timur, menunjukkan adanya variabilitas dalam faktor-faktor eksternal seperti kondisi cuaca, perubahan teknologi, atau kebijakan pertanian yang mempengaruhi hasil panen tiap tahun.

Dari hasil visualisasi *dataset* yang ditunjukkan pada Gambar 4.11 hingga Gambar 4.15, terlihat bahwa setiap dataset menunjukkan pola dan tren yang berbeda-beda. Selain itu, nilai variabel luas panen dan produksi tidak selalu bergerak searah. Dengan kata lain, kenaikan atau penurunan luas panen tidak selalu diikuti oleh kenaikan atau penurunan produksi. Berdasarkan hal tersebut, penelitian ini akan menganalisis korelasi antara variabel luas panen dan jumlah produksi untuk menentukan sejauh mana luas panen memengaruhi jumlah produksi yang dihasilkan. Tujuan analisis ini adalah untuk mengidentifikasi apakah terdapat hubungan yang signifikan antara kedua variabel tersebut dan seberapa kuat pengaruh luas panen terhadap perubahan jumlah produksi. Tahapan analisis korelasi akan dijelaskan pada bagian selanjutnya.

4.5 Implementasi Tahap Analisis Korelasi

Setelah melakukan tahapan pra-pemrosesan data, tahap selanjutnya adalah melakukan analisis korelasi antara variabel luas panen dan jumlah produksi. Analisis ini bertujuan untuk mengukur dan mendeskripsikan hubungan atau keterkaitan antara kedua variabel tersebut. Dalam penelitian ini, metode yang digunakan adalah korelasi *Pearson* (*Pearson Correlation Coefficient*), yaitu metode yang paling umum digunakan untuk menghitung korelasi antara dua variabel kontinu.

Berikut ini merupakan langkah-langkah untuk menghitung korelasi *Pearson* pada salah satu *dataset*, yaitu data tanaman padi untuk wilayah Kalimantan Barat.

Menentukan variabel:

- $x \rightarrow \text{luas panen}$
- $y \rightarrow \text{produksi}$

Menghitung nilai rata-rata:

- $\bar{x} = 339484,3462$
- $\bar{y} = 847511,2538$

Menghitung nilai deviasi:

Tabel 4. 2 Perhitungan Nilai Deviasi

No	x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	304597	323547	-34887,34615	-523964,2538	1217126922	274538539309	18279722296
2	297315	365736	-42169,34615	-481775,2538	1778253755	232107395219	20316147448
3	302442	422773	-37042,34615	-424738,2538	1372135409	180402584280	15733301424
4	295709	405220	-43775,34615	-442291,2538	1916280931	195621553229	19361452738
5	309434	466471	-30050,34615	-381040,2538	903023304	145191675051	11450391527
6	311740	524510	-27744,34615	-323001,2538	769748743,5	104329809986	8961458595
7	305071	525252	-34413,34615	-322259,2538	1184278394	103851026689	11090019254
8	304477	563430	-35007,34615	-284081,2538	1225514285	80702158787	9944930789
9	304142	547570	-35342,34615	-299941,2538	1249081432	89964755759	10600627619
10	310785	611837	-28699,34615	-235674,2538	823652469,7	55542353926	6763696991
11	299783	632818	-39701,34615	-214693,2538	1576196886	46093193247	8523611188
12	280275	627662	-59209,34615	-219849,2538	3505746672	48333694417	13017130573
13	275483	587008	-64001,34615	-260503,2538	4096172310	67861945264	16672558924
14	275269	578755	-64215,34615	-268756,2538	4123610682	72229923981	17258275872
15	294403	622203	-45081,34615	-225308,2538	2032327771	50763809251	10157199383
16	286870	609850	-52614,34615	-237661,2538	2768269421	56482871580	12504391477
17	300783	647973	-38701,34615	-199538,2538	1497794194	39815514748	7722399033
18	302176	689049	-37308,34615	-158462,2538	1391912693	25110285894	5911964619
19	287708	659858	-51776,34615	-187653,2538	2680790021	35213743679	9715999828
20	296894	678763	-42590,34615	-168748,2538	1813937586	28475973176	7187046544
21	313716	722389	-25768,34615	-125122,2538	664007663,5	15655578408	3224193549

No	x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
22	330442	770136	-9042,346154	-77375,25385	81764023,97	5986929908	699653829
23	328445	766542	-11039,34615	-80969,25385	121867163,5	6556020068	893847621
24	350044	827669	10559,65385	-19842,25385	111506289,4	393715037,7	-209527332,1
25	341104	838563	1619,653846	-8948,253846	2623278,581	80071246,9	-14493073,76
26	343809	829106	4324,653846	-18405,25385	18702630,89	338753369,1	-79596351,84
27	349421	827499	9936,653846	-20012,25385	98737089,66	400490304	-198854839,2
28	392717	969658	53232,65385	122146,7462	2833715436	14919827596	6502195456
29	361163	903191	21678,65385	55679,74615	469964032,6	3100234132	1207061943
30	361944	941630	22459,65385	94118,74615	504436050,9	8858338378	2113874459
31	346572	985489	7087,653846	137977,7462	50234837,04	19037858434	977938503,2
32	353434	1027122	13949,65385	179610,7462	194592842,4	32260020134	2505507736
33	365218	1060652	25733,65385	213140,7462	662220940,3	45428977671	5484890182
34	352305	1023684	12820,65385	176172,7462	164369165	31036836487	2258649796
35	378042	1107661	38557,65385	260149,7462	1486692670	67677890424	10030763860
36	399832	1225259	60347,65385	377747,7462	3641839325	142693359724	22796190226
37	423601	1321443	84116,65385	473931,7462	7075611454	224611300012	39865552638
38	418929	1300798	79444,65385	453286,7462	6311453025	205468874239	36011208641
39	428461	1343890	88976,65385	496378,7462	7916844930	246391859633	44166119873
40	444353	1372989	104868,6538	525477,7462	10997434560	276126861703	55106143865
41	427798	1300101	88313,65385	452589,7462	7799301456	204837478324	39969854176
42	444070	1441876	104585,6538	594364,7462	10938158990	353269451471	62162025600
43	452242	1372695	112757,6538	525183,7462	12714288501	275817967224	59218487054
44	474300	1368609,5	134815,6538	521098,2462	18175260522	271543382145	70252200773
45	496358	1364524	156873,6538	517012,7462	24609343271	267302179686	81105678574
46	507698	1397952	168213,6538	550440,7462	28295833340	302985015026	92591649136
47	286476	799715,21	-53008,34615	-47796,04385	2809884762	2284461807	2533589237
48	290049	847875,13	-49435,34615	363,8761538	2443853449	132405,8553	-17988343,62
49	256575	778170,36	-82909,34615	-69340,89385	6873959680	4808159559	5749008171
50	223166	711897	-116318,3462	-135614,2538	13529957652	18391225846	15774425722
51	241478	731225	-98006,34615	-116286,2538	9605243886	13522492834	11396790847

No	x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
52	224068	700290	-115416,3462	-147221,2538	13320932960	21674097584	16991739195

Menghitung perkalian dari deviasi x dan y:

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 932241106844$$

Menghitung kuadrat deviasi dari x:

$$\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} = 482131,1956$$

Menghitung kuadrat deviasi dari y:

$$\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} = 2239663,512$$

Menghitung perkalian dari kuadrat deviasi x dan y:

$$\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} = 1079811646884$$

Menghitung koefisien korelasi:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = 0,8633367769 \approx 0,86$$

Berdasarkan perhitungan di atas, diperoleh nilai koefisien korelasi sebesar 0,86 antara luas panen dengan jumlah produksi tanaman padi di Kalimantan Barat. Untuk melihat perhitungan analisis korelasi untuk seluruh dataset yang digunakan dalam penelitian ini, Anda dapat mengunjungi tautan di bawah ini:

<https://drive.google.com/drive/folders/1JMKR6QIfX5oGefsrvbs87FXUVDglY8I-.>

Implementasi dari tahapan analisis ini dalam kode program ditunjukkan seperti pada Gambar 4.16 berikut.

```
# Menghitung korelasi dengan metode Korelasi Pearson
correlation = dataset['luas panen'].corr(dataset['produksi'], method='pearson')
print(f"Korelasi antara luas panen dan produksi: {correlation}")

Korelasi antara luas panen dan produksi: 0.8633367768669283
```

Gambar 4. 16 Proses Analisis Korelasi

Setelah menjalankan kode program seperti yang ditunjukkan pada Gambar 4.12 di atas, akan dihasilkan nilai koefisien korelasi yang menunjukkan kekuatan dan arah korelasi antara variabel luas panen dan jumlah produksi. Berikut merupakan rangkuman nilai koefisien korelasi dari seluruh *dataset* yang mencakup 5 provinsi dan 5 jenis tanaman yang diteliti.

Tabel 4. 3 Nilai Koefisien Korelasi Provinsi Kalimantan Barat

Provinsi	Jenis Tanaman Pangan	Nilai Korelasi
Kalimantan Barat	Padi	0,86
	Jagung	0,93
	Kedelai	0,96
	Ubi Kayu	0,43
	Ubi Jalar	0,61

Berikut adalah analisis nilai koefisien korelasi untuk Provinsi Kalimantan Barat yang terdapat pada Tabel 4.3 berdasarkan interpretasi dari Tabel 2.1:

- a) Padi memiliki nilai koefisien korelasi 0,86, yang termasuk dalam kategori korelasi sangat kuat. Ini menunjukkan bahwa luas panen sangat memengaruhi jumlah produksi padi di Kalimantan Barat.
- b) Jagung dengan koefisien korelasi 0,93 menunjukkan korelasi mendekati sempurna, yang berarti hubungan antara luas panen dan jumlah produksi jagung sangat signifikan dan kuat.
- c) Kedelai memiliki nilai korelasi 0,96 yang juga tergolong korelasi mendekati sempurna. Hal ini mengindikasikan bahwa peningkatan luas panen akan sangat memengaruhi peningkatan jumlah produksi kedelai.
- d) Ubi Kayu memiliki koefisien 0,43, yang tergolong korelasi sedang, menunjukkan bahwa pengaruh luas panen terhadap jumlah produksi ubi kayu tidak sekuat tanaman lainnya.
- e) Ubi Jalar dengan nilai 0,61 menunjukkan korelasi kuat, mengindikasikan hubungan yang signifikan antara luas panen dan jumlah produksi ubi jalar.

Tabel 4. 4 Nilai Koefisien Korelasi Provinsi Kalimantan Tengah

Provinsi	Jenis Tanaman Pangan	Nilai Korelasi
Kalimantan Tengah	Padi	0,91
	Jagung	0,81
	Kedelai	0,98
	Ubi Kayu	0,53
	Ubi Jalar	0,94

Berikut adalah analisis nilai koefisien korelasi untuk Provinsi Kalimantan Tengah yang terdapat pada Tabel 4.4 berdasarkan interpretasi dari Tabel 2.1:

- a) Padi memiliki nilai koefisien korelasi 0,91, yang termasuk dalam kategori korelasi mendekati sempurna, menandakan adanya keterkaitan yang sangat kuat antara luas panen dengan jumlah produksi padi di Kalimantan Tengah.
- b) Jagung dengan nilai korelasi 0,81 tergolong korelasi sangat kuat, menunjukkan luas panen sangat berpengaruh terhadap jumlah produksi jagung.
- c) Kedelai memiliki koefisien 0,98, yang termasuk dalam korelasi mendekati sempurna, menunjukkan pengaruh luas panen yang sangat besar pada jumlah produksi kedelai.
- d) Ubi Kayu memiliki korelasi 0,53 yang tergolong korelasi kuat, menandakan luas panen cukup signifikan dalam menentukan jumlah produksi ubi kayu yang dihasilkan.
- e) Ubi Jalar dengan koefisien korelasi 0,94 juga tergolong korelasi mendekati sempurna, menandakan adanya keterkaitan yang sangat kuat antara luas panen dengan jumlah produksi ubi jalar.

Tabel 4. 5 Nilai Koefisien Korelasi Provinsi Kalimantan Selatan

Provinsi	Jenis Tanaman Pangan	Nilai Korelasi
Kalimantan Selatan	Padi	0,87
	Jagung	0,9
	Kedelai	0,99
	Ubi Kayu	0,77
	Ubi Jalar	0,56

Berikut adalah analisis nilai koefisien korelasi untuk Provinsi Kalimantan Selatan yang terdapat pada Tabel 4.5 berdasarkan interpretasi dari Tabel 2.1:

- Padi memiliki korelasi 0,87, termasuk dalam kategori korelasi sangat kuat, yang menunjukkan luas panen sangat memengaruhi jumlah produksi padi.
- Jagung dengan nilai 0,90 tergolong korelasi sangat kuat, menunjukkan pengaruh luas panen yang sangat kuat terhadap jumlah produksi jagung.
- Kedelai dengan korelasi 0,99 termasuk dalam korelasi mendekati sempurna, menunjukkan hubungan yang hampir sempurna antara luas panen dan jumlah produksi kedelai.
- Ubi Kayu memiliki korelasi 0,77 yang tergolong korelasi sangat kuat, mengindikasikan bahwa luas panen sangat memengaruhi jumlah produksi ubi kayu.
- Ubi Jalar dengan korelasi 0,56 tergolong korelasi kuat, menunjukkan bahwa pengaruh luas panen terhadap jumlah produksi ubi jalar cukup signifikan.

Tabel 4. 6 Nilai Koefisien Korelasi Provinsi Kalimantan Timur

Provinsi	Jenis Tanaman Pangan	Nilai Korelasi
Kalimantan Timur	Padi	0,85
	Jagung	0,88
	Kedelai	0,97
	Ubi Kayu	0,77
	Ubi Jalar	0,92

Berikut adalah analisis nilai koefisien korelasi untuk Provinsi Kalimantan Timur yang terdapat pada Tabel 4.6 berdasarkan interpretasi dari Tabel 2.1:

- a) Padi dengan nilai korelasi 0,85 tergolong korelasi sangat kuat, menandakan bahwa luas panen memiliki pengaruh yang besar terhadap jumlah produksi padi.
- b) Jagung memiliki nilai korelasi 0,88, yang juga tergolong korelasi sangat kuat, menunjukkan pengaruh luas panen yang kuat pada jumlah produksi jagung.
- c) Kedelai dengan koefisien korelasi 0,97 tergolong korelasi mendekati sempurna, menunjukkan korelasi yang sangat signifikan antara luas panen dengan jumlah produksi kedelai.
- d) Ubi Kayu dengan korelasi 0,77 juga menunjukkan korelasi sangat kuat, mengindikasikan bahwa luas panen sangat memengaruhi jumlah produksi ubi kayu.
- e) Ubi Jalar dengan koefisien korelasi 0,92 tergolong korelasi mendekati sempurna, mengindikasikan adanya keterkaitan yang sangat kuat antara luas panen dengan jumlah produksi ubi jalar.

Tabel 4. 7 Nilai Koefisien Korelasi Provinsi Kalimantan Utara

Provinsi	Jenis Tanaman Pangan	Nilai Korelasi
Kalimantan Utara	Padi	1
	Jagung	0,73
	Kedelai	1
	Ubi Kayu	0,94
	Ubi Jalar	0,79

Berikut adalah analisis nilai koefisien korelasi untuk Provinsi Kalimantan Utara yang terdapat pada Tabel 4.7 berdasarkan interpretasi dari Tabel 2.1:

- a) Padi memiliki nilai korelasi 1, menunjukkan korelasi sempurna, yang berarti luas panen dan jumlah produksi padi memiliki hubungan yang sangat erat.
- b) Jagung dengan koefisien korelasi 0,73 tergolong korelasi sangat kuat, menunjukkan bahwa luas panen memengaruhi jumlah produksi jagung secara signifikan.
- c) Kedelai juga memiliki korelasi sempurna dengan nilai 1, yang menunjukkan korelasi sempurna, mengindikasikan bahwa jumlah produksi kedelai sangat bergantung pada luas panen.
- d) Ubi Kayu dengan korelasi 0,94 tergolong korelasi mendekati sempurna, menunjukkan adanya hubungan yang sangat erat antara luas panen dan jumlah produksi ubi kayu.
- e) Ubi Jalar memiliki nilai korelasi 0,79, yang tergolong korelasi sangat kuat, menunjukkan bahwa jumlah produksi ubi jalar berpengaruh signifikan terhadap luas area panen.

Dari analisis ini, terlihat bahwa jumlah produksi sebagian besar tanaman pangan memiliki korelasi yang kuat hingga mendekati sempurna dengan luas panen di setiap provinsi. Hal ini menunjukkan bahwa luas panen merupakan faktor penting dalam memprediksi jumlah produksi tanaman pangan di Kalimantan. Padi dan kedelai umumnya memiliki nilai korelasi tertinggi, yang berarti jumlah produksi kedua tanaman ini sangat dipengaruhi

oleh luas area panen yang tersedia. Di sisi lain, ubi kayu dan ubi jalar memiliki korelasi yang bervariasi, tetapi umumnya masih menunjukkan hubungan yang cukup signifikan dengan luas panen.

4.6 Implementasi Tahap Pembangunan Model

Setelah melakukan tahapan analisis korelasi, tahapan selanjutnya ialah membangun model menggunakan algoritma *Linear Regression* dan *Random Forest*. Tahapan ini akan meliputi langkah-langkah, yaitu pemisahan fitur dan target dan melakukan pelatihan serta evaluasi model *Linear Regression* dan *Random Forest*. Gambar 4.17 menunjukkan beberapa *library* yang digunakan pada tahap pembangunan model.

```
from math import sqrt
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

Gambar 4. 17 Import *Library* untuk Proses Pembangunan Model

Berikut penjelasan dari *library-library* yang digunakan:

- sqrt* dari *math*: Fungsi untuk menghitung akar kuadrat.
- LinearRegression* dari *sklearn.linear_model*: Model regresi linear untuk prediksi berdasarkan hubungan linier antara variabel input dan output.
- RandomForestRegressor* dari *sklearn.ensemble*: Model regresi berbasis *ensemble* yang menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang akurat.
- mean_absolute_error* dari *sklearn.metrics*: Fungsi untuk menghitung metrik evaluasi *Mean Absolute Error* (MAE).
- mean_squared_error* dari *sklearn.metrics*: Fungsi untuk menghitung metrik evaluasi *Mean Squared Error* (MSE).
- r2_score* dari *sklearn.metrics*: Fungsi untuk menghitung metrik evaluasi *R-Squared* (R^2)

4.6.1 Pemisahan Fitur dan Target

Pada tahap ini, akan dilakukan pemisahan *dataset* menjadi variabel independen (X) sebagai data fitur dan variabel dependen (y) sebagai data target. Variabel independen (X) meliputi data luas panen dan fitur *lag* dari data produksi, yaitu *produksi_lag_1* dan *produksi_lag_2*. Sedangkan variabel dependen (y) merupakan variabel yang ingin diprediksi, yaitu data produksi. Tahapan pemisahan data fitur dan target ini penting untuk memastikan bahwa model dilatih menggunakan informasi yang relevan untuk memprediksi hasil yang diinginkan. Gambar 4.18 menunjukkan implementasi dari proses ini.

```
# # Memisahkan fitur dan target
X = dataset[['luas panen', 'produksi_lag_1', 'produksi_lag_2']]
y = dataset['produksi']
```

Gambar 4.18 Pemisahan Fitur dan Target

4.6.2 Pelatihan dan Evaluasi Model *Linear Regression*

Setelah melakukan tahapan pemisahan fitur dan target, tahap selanjutnya adalah mengimplementasikan algoritma *Linear Regression*. Model diinisialisasi menggunakan *library scikit-learn (sklearn)*, kemudian dilanjutkan dengan pelatihan menggunakan data fitur dan target, yaitu X dan y. Pada tahap ini, model belajar dari data latih untuk menemukan hubungan antara variabel fitur dan variabel target. Implementasi dari proses ini dapat dilihat pada Gambar 4.19 berikut.

```
# Inisialisasi model Linear Regression
model_lr = LinearRegression()

# Melatih model
model_lr.fit(X, y)

▼ LinearRegression ⓘ ⓘ
LinearRegression()
```

Gambar 4.19 Inisialisasi dan Pelatihan Model *Linear Regression*

Setelah pelatihan, selanjutnya dilakukan evaluasi kinerja dari model tersebut. Proses evaluasi ini bertujuan untuk mengukur seberapa baik model mampu memprediksi data yang sebenarnya. Evaluasi dilakukan menggunakan beberapa metrik evaluasi, yaitu MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), dan R² (*R-Squared*) untuk menilai akurasi prediksi. Perhitungan metrik evaluasi juga akan menggunakan *library sklearn* seperti pada Gambar 4.20 di bawah ini. Hasil metrik evaluasi dari seluruh *dataset* yang diteliti akan dibahas pada bagian selanjutnya.

```
# Menghitung metrik evaluasi Model Linear Regression
mae_lr = mean_absolute_error(y, predict_lr)
rmse_lr = sqrt(mean_squared_error(y, predict_lr))
r2_lr = r2_score(y, predict_lr)

print(f'Training MAE: {mae_lr:.2f}')
print(f'Training RMSE: {rmse_lr:.2f}')
print(f'Training R2: {r2_lr:.2f}')

Training MAE: 22171.10
Training RMSE: 31976.95
Training R2: 0.96
```

Gambar 4. 20 Evaluasi Model *Linear Regression*

4.6.3 Pelatihan dan Evaluasi Model *Random Forest*

Setelah mengimplementasikan algoritma *Linear Regression*, langkah berikutnya adalah membangun model menggunakan algoritma *Random Forest*, yang diinisialisasi menggunakan *library scikit-learn (sklearn)*. Ada beberapa parameter yang digunakan dalam inisialisasi model *Random Forest*, yaitu:

- 1) *n_estimators=100*: Parameter ini menunjukkan jumlah *decision tree* yang akan digunakan dalam proses *ensemble*. Dalam hal ini, model akan membangun 100 *decision tree*. Semakin banyak *decision tree* yang digunakan, biasanya model menjadi lebih akurat.
- 2) *random_state=42*: Parameter ini digunakan untuk memastikan hasil yang didapatkan konsisten setiap kali model dilatih dengan

data yang sama. Nilai 42 adalah nilai yang sering digunakan sebagai contoh *default*, tetapi angka lainnya juga dapat digunakan. Dengan mengatur *random_state*, kita memastikan bahwa pembagian data dan hasil model tidak berubah setiap kali kode dijalankan.

Setelah model diinisialisasikan, langkah selanjutnya adalah melakukan pelatihan model. Proses inisialisasi dan pelatihan model *Random Forest* ditampilkan pada Gambar 4.21.



```
# Inisialisasi model Random Forest
model_rf = RandomForestRegressor(n_estimators=100, random_state=42)

# Melatih model
model_rf.fit(X, y)
```

• RandomForestRegressor ⓘ ⓘ

```
RandomForestRegressor(random_state=42)
```

Gambar 4. 21 Inisialisasi dan Pelatihan Model *Random Forest*

Setelah proses pelatihan selesai, model dievaluasi untuk menilai sejauh mana prediksi yang dihasilkan oleh model tersebut akurat. Metrik evaluasi yang diterapkan untuk model *Random Forest* sama seperti yang diterapkan pada model *Linear Regression*, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan R^2 (*R-Squared*). Evaluasi ini penting untuk menentukan apakah model *Random Forest* memberikan hasil prediksi yang lebih akurat dibandingkan model *Linear Regression*. Selain itu, algoritma *Random Forest* cenderung lebih kuat terhadap *overfitting* karena menggunakan banyak *decision tree* dalam proses *ensemble*-nya. Proses perhitungan metrik evaluasi dilakukan menggunakan *library sklearn*, seperti yang ditunjukkan pada Gambar 4.22.

```
# Menghitung metrik evaluasi Model Random Forest
mae_rf = mean_absolute_error(y, predict_rf)
rmse_rf = sqrt(mean_squared_error(y, predict_rf))
r2_rf = r2_score(y, predict_rf)

print(f'Training MAE: {mae_rf:.2f}')
print(f'Training RMSE: {rmse_rf:.2f}')
print(f'Training R2: {r2_rf:.2f}')

Training MAE: 9324.02
Training RMSE: 14697.71
Training R2: 0.99
```

Gambar 4. 22 Evaluasi Model *Random Forest*

Hasil evaluasi dari model *Random Forest* terhadap keseluruhan dataset akan dibahas secara rinci pada bagian berikutnya, guna membandingkan performa kedua model dan memahami model mana yang lebih efektif dalam memprediksi produksi tanaman pangan di Kalimantan.

4.7 Hasil Perbandingan Evaluasi Kinerja Model

Pada tahap ini, kinerja antara model *Linear Regression* dan *Random Forest* akan dibandingkan berdasarkan nilai metrik evaluasi. Metrik evaluasi yang digunakan adalah *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2). Metrik-metrik ini memberikan gambaran mengenai seberapa akurat model dalam memprediksi data dengan tingkat kesalahan yang sekecil mungkin. Perbandingan nilai-nilai evaluasi ini bertujuan untuk mengidentifikasi model yang memberikan kinerja terbaik, dengan fokus pada model yang menghasilkan error paling rendah dan nilai R^2 yang paling tinggi. Hasil perbandingan ini akan dilakukan berdasarkan jenis tanaman yang dianalisis, sehingga dapat memberikan wawasan yang lebih spesifik mengenai kinerja masing-masing model untuk tanaman tertentu. Dengan membandingkan hasil evaluasi berdasarkan jenis tanaman, diharapkan dapat diidentifikasi model mana yang paling efektif untuk memprediksi produksi berbagai jenis tanaman pangan.

Tabel 4. 8 Evaluasi Kinerja Model untuk Tanaman Padi

Provinsi	Hasil Evaluasi					
	Model <i>Linear Regression</i>			Model <i>Random Forest</i>		
	MAE	RMSE	R ²	MAE	RMSE	R ²
Kalimantan Barat	47885.42	71257.16	0.95	16117.49	23741.10	0.99
Kalimantan Tengah	23335.98	34035.68	0.97	9872.23	13357.30	1
Kalimantan Selatan	128956.25	293044.78	0.80	56349.83	128037.70	0.96
Kalimantan Timur	22171.10	31976.95	0.96	9324.02	14697.71	0.99
Kalimantan Utara	305.78	437.37	1	2328.05	4123.34	0.94

Catatan : MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), R² (*R-Squared*)

Berdasarkan Tabel 4.8, secara umum model *Random Forest* memperlihatkan kinerja yang lebih baik daripada model *Linear Regression* dalam memprediksi produksi padi di Kalimantan. Hal ini terlihat dari nilai MAE dan RMSE yang lebih kecil serta nilai R² yang lebih tinggi di sebagian provinsi. Di Kalimantan Tengah, Kalimantan Barat, dan Kalimantan Timur, model *Random Forest* memiliki nilai R² mendekati atau bahkan sempurna, menunjukkan akurasi yang sangat tinggi. Meski model *Linear Regression* memberikan hasil terbaik di Kalimantan Utara dengan nilai MAE dan RMSE sangat rendah serta R² sempurna, model *Random Forest* tetap menunjukkan performa yang lebih stabil dan akurat di sebagian besar wilayah lainnya.

Tabel 4. 9 Evaluasi Kinerja Model untuk Tanaman Jagung

Provinsi	Hasil Evaluasi					
	Model <i>Linear Regression</i>			Model <i>Random Forest</i>		
	MAE	RMSE	R ²	MAE	RMSE	R ²
Kalimantan Barat	12164.10	18465.82	0.92	3858.60	7654.50	0.99
Kalimantan Tengah	4345.20	6048.17	0.88	1312.67	2930.34	0.97
Kalimantan Selatan	18752.52	24219.15	0.93	4340.21	9581.15	0.99
Kalimantan Timur	6204.56	8329.60	0.90	1664.01	3496.11	0.98
Kalimantan Utara	804.97	1032.52	0.59	386.41	530.52	0.89

Catatan : MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), R² (*R-Squared*)

Berdasarkan Tabel 4.9, model *Random Forest* kembali menunjukkan kinerja yang lebih baik dibandingkan model *Linear Regression* dalam memprediksi produksi jagung di wilayah Kalimantan. Model *Random Forest* menunjukkan nilai MAE dan RMSE yang lebih rendah serta nilai R² yang lebih tinggi di seluruh wilayah Kalimantan. Di Kalimantan Barat, Kalimantan Selatan, dan Kalimantan Tengah, model *Random Forest* memiliki nilai R² yang mendekati sempurna. Secara keseluruhan, model *Random Forest* menghasilkan prediksi yang lebih akurat dibandingkan model *Linear Regression* di seluruh wilayah Kalimantan untuk tanaman jagung.

Tabel 4. 10 Evaluasi Kinerja Model untuk Tanaman Kedelai

Provinsi	Hasil Evaluasi					
	Model <i>Linear Regression</i>			Model <i>Random Forest</i>		
	MAE	RMSE	R ²	MAE	RMSE	R ²
Kalimantan Barat	361.16	461.27	0.93	123.21	188.83	0.99
Kalimantan Tengah	252.67	354.55	0.96	119.68	188.66	0.99
Kalimantan Selatan	683.90	909.97	0.97	379.00	662.13	0.98
Kalimantan Timur	253.27	333.10	0.95	103.76	154.14	0.99
Kalimantan Utara	0.09	0.14	1	88.56	154.23	0.89

Catatan : MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), R² (*R-Squared*)

Berdasarkan Tabel 4.10 yang menunjukkan evaluasi kinerja model untuk tanaman kedelai, model *Random Forest* kembali menunjukkan kinerja yang lebih baik dibandingkan model *Linear Regression* di hampir semua provinsi. Pada provinsi Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Timur, model *Random Forest* menghasilkan nilai MAE dan RMSE yang lebih rendah, dengan nilai R² mendekati sempurna, yaitu 0,99. Secara keseluruhan, model *Random Forest* menghasilkan prediksi yang lebih akurat di sebagian besar wilayah, meskipun model *Linear Regression* menunjukkan performa yang sangat baik untuk wilayah Kalimantan Utara.

Tabel 4. 11 Evaluasi Kinerja Model untuk Tanaman Ubi Kayu

Provinsi	Hasil Evaluasi					
	Model <i>Linear Regression</i>			Model <i>Random Forest</i>		
	MAE	RMSE	R ²	MAE	RMSE	R ²
Kalimantan Barat	26309.77	32070.16	0.41	8930.10	11994.63	0.92
Kalimantan Tengah	9621.22	13650.87	0.58	3946.49	6048.63	0.92
Kalimantan Selatan	11441.80	15101.46	0.84	4571.02	5691.42	0.98
Kalimantan Timur	8231.79	10519.72	0.87	3208.78	4275.57	0.98
Kalimantan Utara	189.67	232.48	1	1567.28	1787.50	0.93

Catatan : MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), R² (*R-Squared*)

Berdasarkan Tabel 4.11 yang menampilkan evaluasi kinerja model untuk tanaman ubi kayu, model *Random Forest* menunjukkan kinerja yang lebih baik daripada model *Linear Regression*. Model *Random Forest* memiliki nilai MAE dan RMSE lebih rendah serta nilai R² lebih tinggi di Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Timur, dengan R² mendekati sempurna di Kalimantan Selatan dan Kalimantan Timur. Di Kalimantan Utara, meskipun *Linear Regression* memiliki MAE dan RMSE sangat rendah serta R² sempurna, model *Random Forest* tetap menunjukkan performa yang lebih andal untuk prediksi ubi kayu di hampir semua provinsi.

Tabel 4. 12 Evaluasi Kinerja Model untuk Tanaman Ubi Jalar

Provinsi	Hasil Evaluasi					
	Model <i>Linear Regression</i>			Model <i>Random Forest</i>		
	MAE	RMSE	R ²	MAE	RMSE	R ²
Kalimantan Barat	2181.78	3354.66	0.56	738.81	1325.83	0.93
Kalimantan Tengah	912.45	1283.81	0.91	387.78	660.37	0.98
Kalimantan Selatan	2639.56	3530.17	0.71	1209.26	1662.29	0.94
Kalimantan Timur	1819.07	2276.24	0.92	666.01	916.16	0.99
Kalimantan Utara	231.89	251.73	0.80	219.30	226.12	0.84

Catatan : MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), R² (*R-Squared*)

Berdasarkan Tabel 4.12 yang menunjukkan evaluasi kinerja model untuk tanaman ubi jalar, model *Random Forest* secara konsisten menunjukkan kinerja yang lebih baik dibandingkan model *Linear Regression*. Di Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Timur, model *Random Forest* memiliki nilai MAE dan RMSE yang lebih rendah, dengan nilai R^2 yang lebih tinggi dibandingkan model *Linear Regression*. Provinsi Kalimantan Tengah dan Kalimantan Timur menunjukkan performa terbaik untuk model *Random Forest*, dengan nilai R^2 mendekati sempurna, yaitu 0,98 dan 0,99. Secara keseluruhan, model *Random Forest* terbukti lebih andal dalam memprediksi produksi ubi jalar di sebagian besar wilayah.

Berdasarkan keseluruhan tabel evaluasi dari Tabel 4.8 hingga 4.12 dapat disimpulkan bahwa model *Random Forest* secara konsisten menunjukkan kinerja yang lebih baik dibandingkan model *Linear Regression* dalam memprediksi produksi berbagai jenis tanaman pangan yang meliputi padi, jagung, kedelai, ubi kayu, dan ubi jalar untuk seluruh wilayah Kalimantan. Model *Random Forest* menghasilkan nilai MAE dan RMSE yang lebih rendah serta nilai R^2 yang lebih tinggi di sebagian besar provinsi, hal ini menunjukkan akurasi dan konsistensi yang lebih baik dalam memprediksi produksi tanaman pangan dibandingkan model *Linear Regression*.

4.8 Implementasi Prediksi Menggunakan Algoritma *Random Forest*

Berdasarkan hasil evaluasi kinerja kedua model, diketahui bahwa model *Random Forest* memiliki nilai kesalahan yang lebih kecil dengan nilai akurasi yang lebih tinggi dibandingkan dengan model *Linear Regression*. Oleh karena itu, model *Random Forest* akan digunakan untuk memprediksi produksi tanaman pangan di Pulau Kalimantan selama 5 tahun ke depan. Sebelum melakukan prediksi produksi tanaman pangan, langkah pertama yang akan dilakukan ialah memprediksi nilai luas panen. Langkah ini melibatkan pembangunan model prediksi luas panen menggunakan algoritma

Random Forest untuk memprediksi luas panen di semua provinsi Kalimantan. Setelah luas panen diprediksi, hasil prediksi tersebut akan digunakan sebagai *input* untuk memprediksi produksi tanaman pangan. Selanjutnya model prediksi produksi tanaman pangan yang telah dibangun sebelumnya akan digunakan untuk memprediksi produksi tanaman pangan di Kalimantan untuk 5 tahun mendatang. Prediksi ini akan dilakukan untuk setiap jenis tanaman dan wilayah secara terpisah. Berikut adalah tahapan implementasi dari proses ini:

4.8.1 Pembagunan Model Prediksi Luas Panen

Langkah awal pada tahap ini adalah membuat fitur *lag* untuk data luas panen. Fitur *lag* ini merupakan variabel yang merepresentasikan nilai luas panen pada tahun sebelumnya. Pembuatan fitur *lag* sangat penting dalam analisis deret waktu, karena dapat membantu model memahami pola dan tren yang ada. Proses pembuatan fitur *lag* untuk data luas panen dapat dilihat pada Gambar 4.23 berikut.

```
# Menentukan jumlah lag luas panen
lag_days = 2

# Membuat lag features luas panen
for lag in range(1, lag_days + 1):
    dataset[f'luas_panen_lag_{lag}'] = dataset['luas panen'].shift(lag)
```

Gambar 4. 23 Pembuatan *Lag Features* Luas Panen

Setelah fitur *lag* dibuat, langkah selanjutnya adalah memisahkan *dataset* menjadi data fitur *x_panen* dan data target *y_panen*, seperti pada Gambar 4.24. Dalam hal ini, data fitur terdiri dari nilai *lag* luas panen, sedangkan data target merupakan nilai luas panen.

```
# Memisahkan fitur dan target prediksi luas panen
X_panen = dataset[['luas_panen_lag_1', 'luas_panen_lag_2']]
y_panen = dataset['luas panen']
```

Gambar 4. 24 Pemisahan Fitur dan Target Model Prediksi Luas Panen

Setelah dilakukan pembuatan fitur *lag* dan pemisahan data fitur dan data target, langkah selanjutnya adalah membangun model prediksi luas panen menggunakan algoritma *Random Forest*. Model ini diinisialisasi menggunakan *library scikit-learn (sklearn)*. Setelah model dibangun, langkah selanjutnya adalah melatih model dengan menggunakan data fitur dan target yang telah disiapkan. Proses pembangunan dan pelatihan model dapat dilihat pada Gambar 4.25 di bawah ini.

```
# Inisialisasi model Random Forest
model_rf_lp = RandomForestRegressor(n_estimators=100, random_state=42)

# Melatih model
model_rf_lp.fit(X_panen, y_panen)

RandomForestRegressor()
RandomForestRegressor(random_state=42)
```

Gambar 4.25 Insialisasi dan Pelatihan Model Prediksi Luas Panen

4.8.2 Implementasi Prediksi Luas Panen

Setelah model prediksi luas panen dibangun, langkah berikutnya adalah melakukan prediksi luas panen untuk 5 tahun ke depan. Prediksi ini akan diterapkan pada seluruh *dataset* yang digunakan dalam penelitian ini. Hasil prediksi tersebut nantinya akan menjadi input untuk memprediksi produksi tanaman pangan. Implementasi kode untuk prediksi luas panen ditampilkan pada Gambar 4.26 berikut.

```

# Mengambil data dari baris terakhir dataset
last_row_panen = dataset.iloc[-1].copy()

# Menentukan jumlah tahun yang ingin diprediksi
n_years_panen = 5

# Inisialisasi list untuk menyimpan hasil prediksi luas panen
predictions_panen = []

# Perulangan untuk memprediksi luas panen
for i in range(n_years_panen):
    # Data input untuk melakukan prediksi
    data_pred_panen = pd.DataFrame({
        'luas_panen_lag_1': [last_row_panen['luas panen']],
        'luas_panen_lag_2': [last_row_panen['luas_panen_lag_1']]
    })

    # Prediksi luas panen
    prediksi_luas_panen = model_rf_lp.predict(data_pred_panen)

    # Menyimpan hasil prediksi luas panen
    predictions_panen.append(round(prediksi_luas_panen[0], 2))

    # Update last_row_panen untuk mempersiapkan prediksi berikutnya
    last_row_panen['luas_panen_lag_1'] = last_row_panen['luas panen']
    last_row_panen['luas panen'] = prediksi_luas_panen[0]

# Menampilkan hasil prediksi
print(f"Prediksi luas panen untuk {n_years_panen} tahun ke depan: {predictions_panen}")

Prediksi luas panen untuk 5 tahun ke depan: [113225.28, 114414.75, 116266.26, 119436.88, 120581.56]

```

Gambar 4. 26 Source Code Prediksi Luas Panen

Gambar 4.26 menampilkan kode program untuk memprediksi luas panen selama 5 tahun ke depan menggunakan model prediksi yang telah dibangun. Proses dimulai dengan mengambil data dari baris terakhir *dataset* sebagai *input* awal. Prediksi ditetapkan untuk 5 tahun, dan sebuah *list* kosong diinisialisasi untuk menyimpan hasil prediksi. Pada setiap iterasi, input prediksi dibuat menggunakan dua fitur *lag*, yaitu luas panen pada satu dan dua tahun sebelumnya. Model prediksi luas panen, yaitu *model_rf_lp* kemudian digunakan untuk memprediksi luas panen tahun berikutnya, dan hasilnya disimpan dalam *list* *predictions_panen*. Setelah setiap prediksi, input diperbarui menggunakan hasil prediksi terbaru agar dapat digunakan untuk prediksi tahun selanjutnya. Setelah semua iterasi selesai, hasil prediksi 5 tahun ke depan yang tersimpan dalam *predictions_panen* ditampilkan. Prediksi ini nantinya akan menjadi *input* bagi model prediksi produksi tanaman pangan.

4.8.3 Implementasi Prediksi Produksi Tanaman Pangan

Setelah memperoleh prediksi luas panen, langkah berikutnya adalah melakukan prediksi produksi tanaman pangan. Kode yang digunakan untuk memprediksi produksi tanaman pangan dapat dilihat pada Gambar 4.27 di bawah ini. Proses prediksi produksi ini melibatkan penggunaan data prediksi luas panen sebagai salah satu fitur utama, bersama dengan nilai produksi sebelumnya, yaitu fitur *lag* untuk memperkirakan produksi di masa depan.

```
n_years = 5

# Inisialisasi list untuk menyimpan hasil prediksi produksi
predictions_produksi = []

# Mengambil data dari baris terakhir dataset
last_row_produksi = dataset.iloc[-1].copy()

# Perulangan untuk memprediksi produksi
for i in range(n_years):
    # Data input untuk melakukan prediksi
    data_pred_produksi = pd.DataFrame({
        'luas_panen': [predictions_panen[i]],
        'produksi_lag_1': [last_row_produksi['produksi']],
        'produksi_lag_2': [last_row_produksi['produksi_lag_1']]
    })

    # Prediksi produksi
    prediksi_produksi = model_rf.predict(data_pred_produksi)

    # Menyimpan hasil prediksi produksi
    predictions_produksi.append(round(prediksi_produksi[0], 2))

    # Update last_row_produksi untuk mempersiapkan prediksi berikutnya
    last_row_produksi['produksi_lag_1'] = last_row_produksi['produksi']
    last_row_produksi['produksi'] = prediksi_produksi[0]

# Menampilkan hasil prediksi produksi untuk 5 tahun ke depan
print("Prediksi produksi untuk 5 tahun ke depan: {}").format(predictions_produksi)

Prediksi produksi untuk 5 tahun ke depan: [326237.05, 325830.55, 325830.55, 328448.5, 324817.44]
```

Gambar 4.27 Source Code Prediksi Produksi

Gambar 4.27 menunjukkan kode program untuk memprediksi produksi tanaman pangan selama 5 tahun ke depan menggunakan model *Random Forest*. Pertama, sebuah *list* kosong bernama *predictions_produksi* diinisialisasi untuk menyimpan hasil prediksi tiap tahun. Data dari baris terakhir dataset digunakan sebagai referensi untuk mengambil nilai *lag* produksi terakhir. Prediksi dilakukan dengan iterasi lima kali, di mana setiap iterasi mewakili satu

tahun prediksi. Dalam setiap iterasi, *DataFrame* `data_pred_produksi` dibuat yang berisi data luas panen terbaru serta dua nilai lag produksi, yaitu `produksi_lag_1` dan `produksi_lag_2`. Model *Random Forest* kemudian digunakan untuk memprediksi produksi berdasarkan variabel ini. Hasil prediksi dibulatkan dan disimpan dalam `predictions_produksi`. Setelah prediksi dibuat, nilai *lag* produksi diperbarui dengan prediksi terbaru untuk memastikan prediksi berikutnya mempertimbangkan perubahan produksi dari tahun sebelumnya. Proses ini diulang hingga prediksi lima tahun ke depan diperoleh dan ditampilkan.

4.8.4 Hasil Prediksi Luas Panen dan Produksi Tanaman Pangan

Untuk melihat pola yang dihasilkan dari hasil prediksi yang telah didapatkan, penelitian ini melakukan visualisasi menggunakan *library matplotlib*, seperti yang ditunjukkan pada Gambar 4.28 di bawah ini.

```
n_years = 5
historical_luas_panen = dataset['luas panen'] # Data historis luas panen
historical_produksi = dataset['produksi'] # Data historis produksi

# Mengambil tahun terakhir dari data historis
last_year = historical_produksi.index[-1]

# Membuat array tahun untuk periode prediksi
predicted_years = np.arange(last_year + 1, last_year + n_years + 1)

# Visualisasi data historis
plt.figure(figsize=(10, 6))
plt.plot(historical_luas_panen.index, historical_luas_panen, label='Historical Luas Panen', color='blue', marker='o')
plt.plot(historical_produksi.index, historical_produksi, label='Historical Produksi', color='red', marker='o')

# Visualisasi hasil prediksi luas panen
plt.plot(predicted_years, predictions_panen, label='Predicted Luas Panen', linestyle='--', color='green', marker='x')

# Visualisasi hasil prediksi produksi
plt.plot(predicted_years, predictions_produksi, label='Predicted Produksi', linestyle='--', color='darkorange', marker='x')

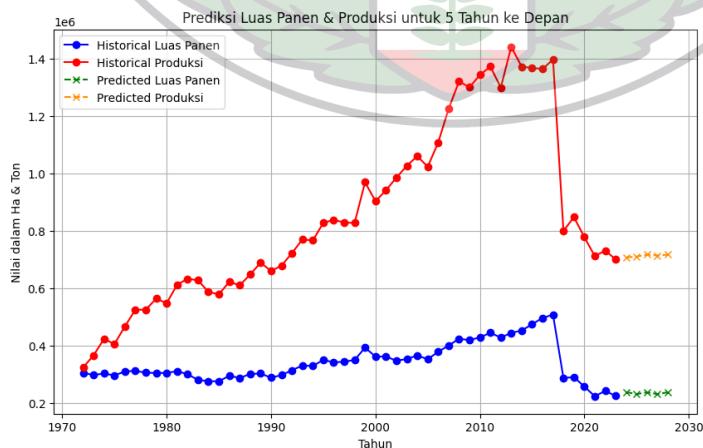
plt.xlabel('Tahun')
plt.ylabel('Nilai dalam Ha & Ton')
plt.xlabel('Tahun')
plt.ylabel('Nilai dalam Ha & Ton')
plt.title('Prediksi Luas Panen & Produksi untuk 5 Tahun ke Depan')
plt.legend()
plt.grid(True)
plt.show()
```

Gambar 4. 28 Source Code Visualisasi Hasil Prediksi

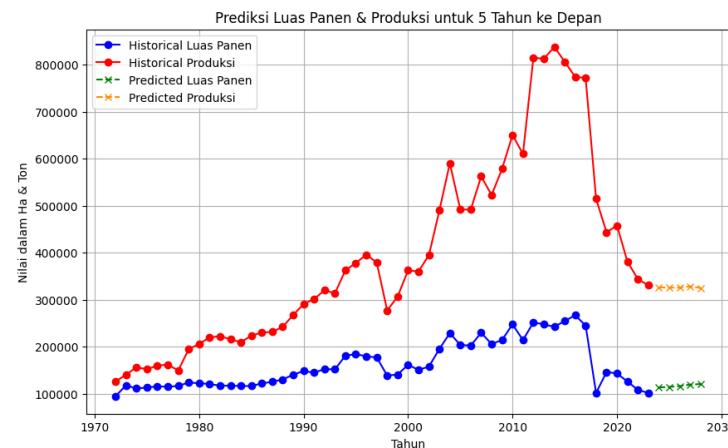
Berikut adalah hasil prediksi luas panen dan produksi tanaman pangan beserta visualisasinya yang dikelompokkan berdasarkan jenis tanaman.

Tabel 4. 13 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Padi

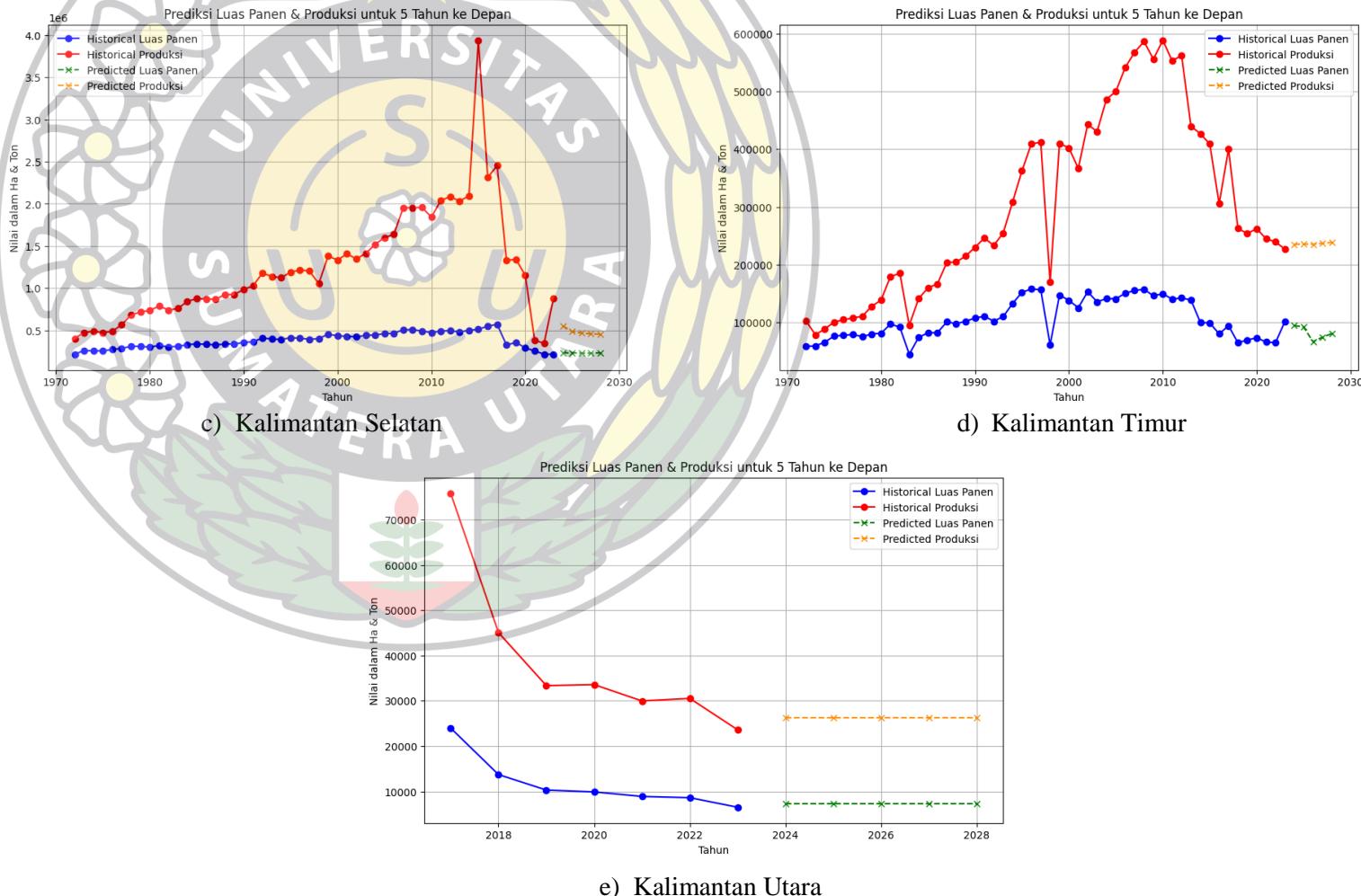
Provinsi	Prediksi Luas Panen dan Produksi Tanaman Padi di Kalimantan									
	Luas Panen (Ha)					Produksi (Ton)				
	2024	2025	2026	2027	2028	2024	2025	2026	2027	2028
Kalimantan Barat	237747.46	230435.26	236876.96	230435.26	236876.96	707687.85	710154.88	717120.26	714070.85	717120.26
Kalimantan Tengah	113225.28	114414.75	116266.26	119436.88	120581.56	326237.05	325830.55	325830.55	328448.5	324817.44
Kalimantan Selatan	230423.57	231879.84	226735.52	226735.52	231879.84	551554.59	482485.2	465050.13	456965.36	452716.55
Kalimantan Timur	94942.75	92018.86	66792.86	74567.9	80840.38	235158.68	235588.35	235331.71	237121.33	238651.56
Kalimantan Utara	7323.4	7323.4	7323.4	7323.4	7323.4	26192.86	26192.86	26192.86	26192.86	26192.86



a) Kalimantan Barat



b) Kalimantan Tengah



Gambar 4. 29 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Padi

Berdasarkan Tabel 4.13 dan Gambar 4.29, berikut adalah analisis hasil prediksi luas panen dan produksi tanaman padi di wilayah Kalimantan:

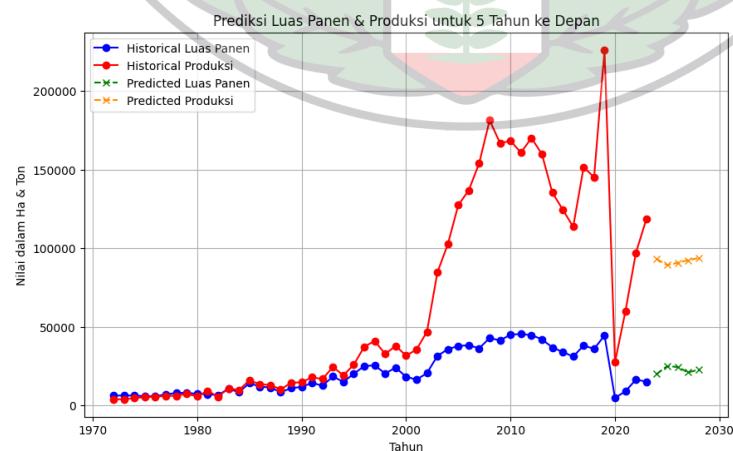
- a) Kalimantan Barat: Luas panen diprediksi akan relatif stabil dari tahun 2024 hingga 2028, dengan sedikit perubahan di sekitar 230.000 hektar. Produksi menunjukkan sedikit peningkatan dari sekitar 707.687 ton menjadi 717.120 ton. Grafik menunjukkan tren yang cukup stabil dengan sedikit kenaikan pada beberapa tahun terakhir prediksi. Hal ini mengindikasikan bahwa Kalimantan Barat memiliki peluang untuk secara bertahap meningkatkan produksi padi.
- b) Kalimantan Tengah: Luas panen diprediksi meningkat secara bertahap dari sekitar 113.225 hektar pada tahun 2024 menjadi 120.581 hektar pada tahun 2028, sementara produksi diperkirakan tetap stabil di kisaran 320.000 ton. Grafik menunjukkan tren yang stabil tanpa adanya peningkatan atau penurunan yang signifikan. Stabilitas ini menunjukkan bahwa Kalimantan Tengah mempertahankan kapasitas produksi yang cukup, meskipun peningkatan signifikan dalam luas panen atau produksi tidak terlihat.
- c) Kalimantan Selatan: Luas panen padi diprediksi bervariasi antara 226.000 hingga 231.000 hektar, dengan sedikit penurunan pada tahun 2026. Sementara itu, produksi padi diprediksi menurun secara bertahap setelah tahun 2024 dari kisaran 551.554 ton menjadi 452.716 ton pada tahun 2028. Grafik menunjukkan perubahan yang tidak terlalu signifikan pada luas panen, disertai dengan tren penurunan konsisten dalam produksi. Tren ini menunjukkan adanya tantangan dalam upaya untuk meningkatkan produksi padi di Kalimantan Selatan.

- d) Kalimantan Timur: Terdapat variasi yang signifikan dalam luas panen, dimulai dari 94.942 hektar pada tahun 2024, turun drastis pada tahun 2026, namun meningkat lagi pada tahun 2027 dan 2028. Produksi padi diprediksi akan meningkat dari sekitar 235.158 ton pada tahun 2024 menjadi 238.651 ton pada tahun 2028. Visualisasi menunjukkan pola stabil dengan peningkatan ringan yang menunjukkan kapasitas produksi yang terbatas namun cenderung meningkat.
- e) Kalimantan Utara: Luas panen dan produksi sangat rendah dan stabil pada angka sekitar 7.323 hektar dan 26.192 ton selama periode prediksi. Grafik menunjukkan tren yang hampir datar tanpa perubahan signifikan pada luas panen maupun produksi, menunjukkan bahwa Kalimantan Utara bukan wilayah utama untuk budidaya padi. Produksi yang stabil pada angka rendah ini menandakan bahwa padi bukanlah komoditas unggulan di provinsi ini, yang mungkin disebabkan keterbatasan lahan atau kondisi yang kurang mendukung.

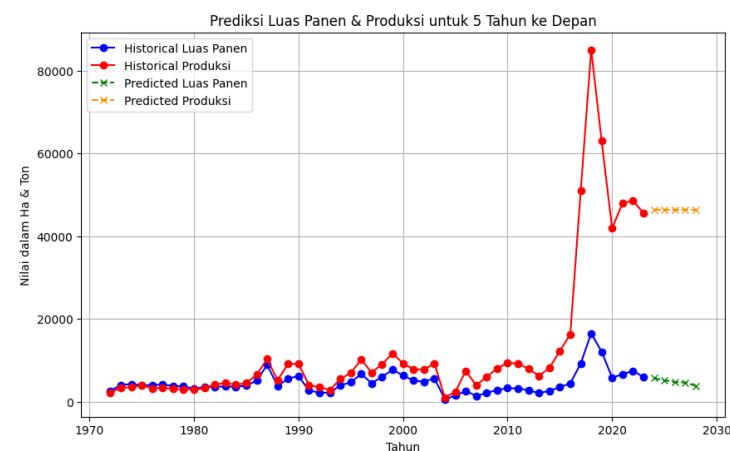
Dari analisis ini, dapat disimpulkan bahwa Kalimantan Barat, Kalimantan Tengah dan Kalimantan Timur menunjukkan stabilitas dalam produksi, dengan peningkatan kecil yang menunjukkan upaya untuk mempertahankan kapasitas produksi. Sementara itu, Kalimantan Utara memiliki luas panen dan produksi yang sangat rendah dan stabil, sehingga berpotensi rendah untuk meningkatkan jumlah produksi padi. Secara keseluruhan, prediksi ini menunjukkan bahwa produksi padi di Kalimantan menunjukkan stabilitas tanpa pertumbuhan yang signifikan.

Tabel 4. 14 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Jagung

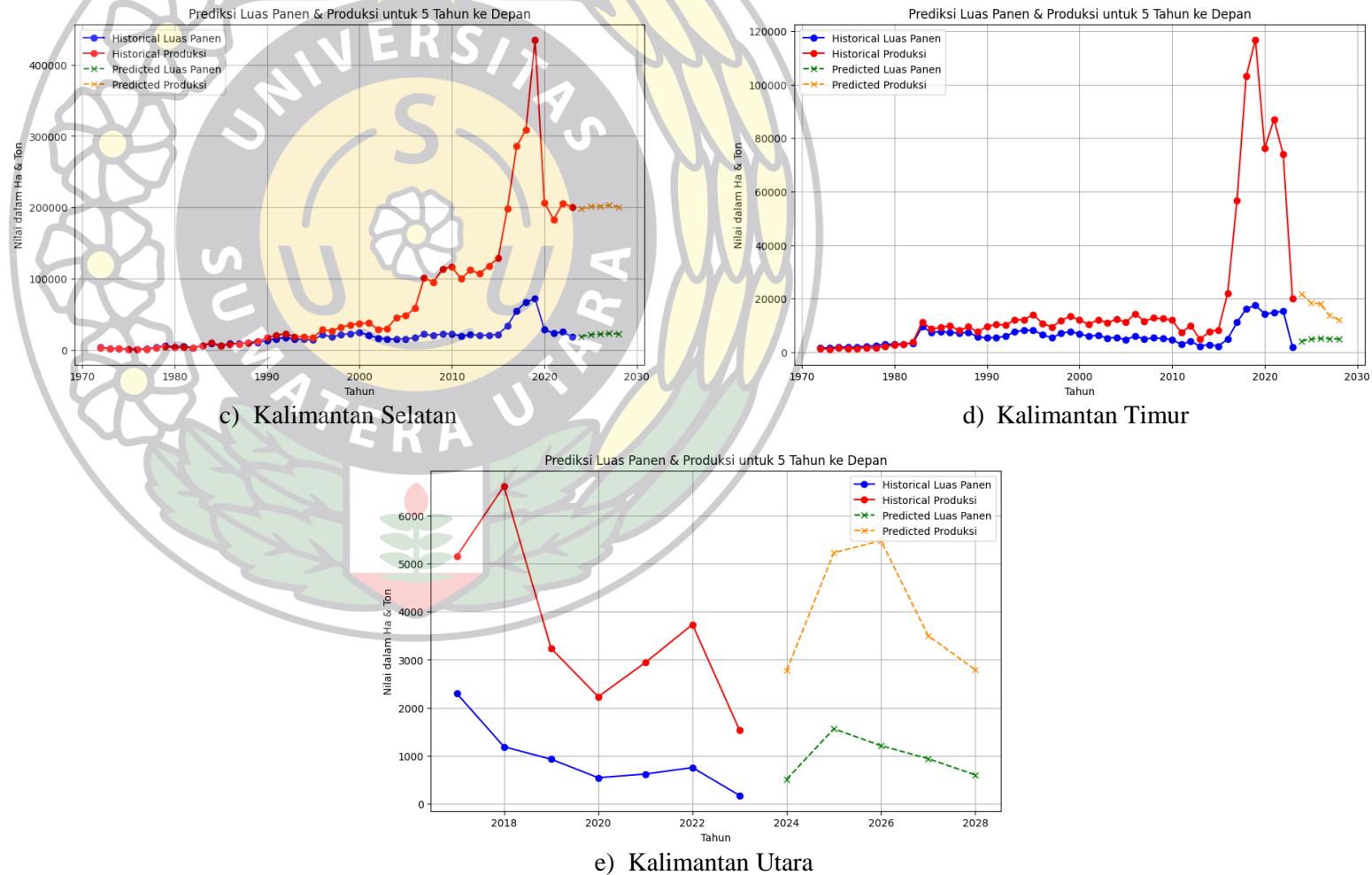
Provinsi	Prediksi Luas Panen dan Produksi Tanaman Jagung di Kalimantan									
	Luas Panen (Ha)					Produksi (Ton)				
	2024	2025	2026	2027	2028	2024	2025	2026	2027	2028
Kalimantan Barat	19924.29	25007.94	24364.18	20973.94	22673.44	93155.46	89247.96	90721.34	92401.33	93706.98
Kalimantan Tengah	5674.3	5152.7	4736.43	4603.94	3829.09	46396.09	46396.09	46396.09	46396.09	46396.09
Kalimantan Selatan	18832.18	21706.14	22157.5	23197.26	22842.28	197914.26	201295.7	201733.54	203012.6	200736.42
Kalimantan Timur	4195.36	4946.03	5155.8	5077.13	4969.54	21720.98	18474.01	18069.89	13827.94	12104.12
Kalimantan Utara	509.89	1560.06	1209.05	938.26	600.41	2774.57	5231.0	5482.76	3502.16	2792.56



a) Kalimantan Barat



b) Kalimantan Tengah



Gambar 4. 30 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Jagung

Berdasarkan Tabel 4.14 dan Gambar 4.30, berikut adalah analisis hasil prediksi luas panen dan produksi tanaman jagung di wilayah Kalimantan:

- a) Kalimantan Barat: Hasil prediksi menunjukkan adanya peningkatan yang stabil dalam luas panen dan produksi, yang mengindikasikan tren pertumbuhan positif dalam produksi jagung. Grafik memperlihatkan tren peningkatan produksi terutama setelah tahun 2020, yang kemungkinan dipengaruhi oleh peningkatan luas panen. Produksi terlihat jauh lebih tinggi dibandingkan luas panen, yang mungkin disebabkan oleh peningkatan hasil per satuan area atau faktor eksternal yang mendukung produksi.
- b) Kalimantan Tengah: Hasil prediksi menunjukkan bahwa tingkat produksi cenderung stabil sepanjang tahun, meskipun terdapat sedikit penurunan dalam luas panen. Dalam grafik juga terlihat bahwa produksi relatif stabil dengan luas panen yang mengalami sedikit penurunan setelah tahun 2025. Stabilitas ini mungkin menunjukkan efisiensi dalam pengelolaan luas panen yang ada atau bahwa produksi di provinsi ini telah mencapai potensi maksimalnya.
- c) Kalimantan Selatan: Hasil prediksi menunjukkan bahwa terdapat peningkatan yang stabil dalam luas panen dan produksi sekitar tahun 2025-2026. Grafik mengonfirmasi adanya peningkatan dalam produksi yang sejalan dengan peningkatan luas panen, kemudian sedikit menurun pada tahun 2028. Hal ini bisa menunjukkan adanya peningkatan sementara yang mungkin terjadi akibat kondisi yang menguntungkan atau inisiatif pemerintah, namun ada kemungkinan masalah keberlanjutan setelah itu.
- d) Kalimantan Timur: Hasil prediksi menunjukkan adanya peningkatan bertahap dalam luas panen sekitar tahun 2024

hingga 2026, namun produksi jagung diprediksi akan turun dari 21.720 ton pada tahun 2024 menjadi 12.104 ton pada tahun 2028. Tren yang sedikit meningkat pada luas panen menunjukkan adanya potensi pertumbuhan luas area panen yang stabil di wilayah ini, meskipun kemungkinan tingkat produksi akan mengalami tren penurunan secara bertahap.

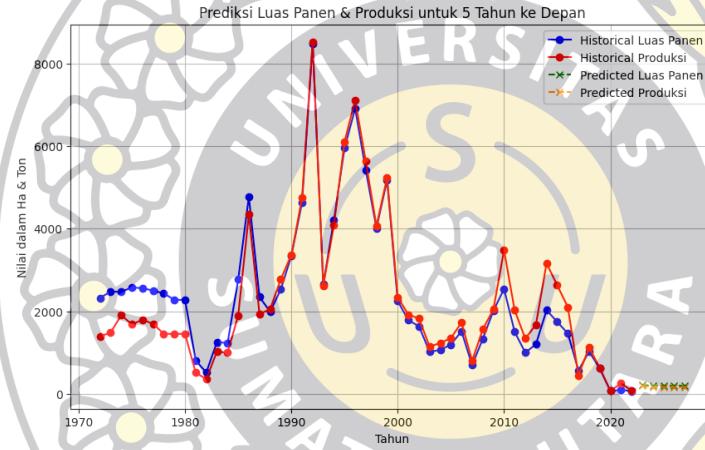
- e) Kalimantan Utara: Hasil prediksi untuk luas panen dan produksi relatif rendah dan menunjukkan fluktuasi kecil tanpa pertumbuhan signifikan dari tahun ke tahun. Grafik mencerminkan nilai-nilai yang rendah dan tidak ada perubahan besar, yang menunjukkan bahwa produksi jagung di wilayah ini mungkin terbatas oleh faktor alam atau kekurangan dukungan pengembangan.

Dari analisis ini, dapat disimpulkan bahwa Kalimantan Barat dan Kalimantan Selatan menunjukkan pertumbuhan produksi yang paling kuat dan berpotensi menjadi wilayah unggulan untuk produksi jagung jika tren ini berlanjut. Sementara itu, fluktuasi di Kalimantan Utara menunjukkan kapasitas produksi yang lebih rendah yang kemungkinan disebabkan oleh kendala seperti lahan yang terbatas atau kondisi yang kurang mendukung untuk budidaya jagung. Secara keseluruhan, prediksi ini menunjukkan adanya potensi pengembangan strategis dalam produksi jagung di Kalimantan, terutama di wilayah-wilayah dengan produktivitas yang masih rendah.

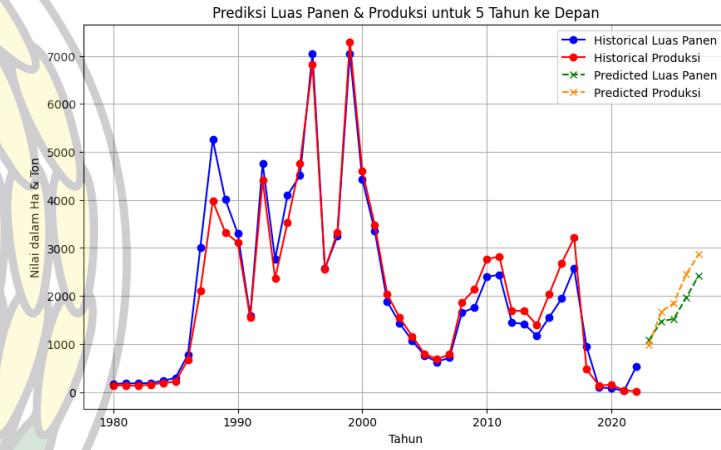
Tabel 4. 15 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Kedelai

Provinsi	Prediksi Luas Panen dan Produksi Tanaman Kedelai di Kalimantan									
	Luas Panen (Ha)					Produksi (Ton)				
	2023	2024	2025	2026	2027	2023	2024	2025	2026	2027
Kalimantan Barat	211.0	199.18	199.18	199.18	199.18	195.03	165.36	165.36	165.36	165.36
Kalimantan Tengah	1079.48	1471.45	1522.53	1967.15	2424.46	988.8	1668.24	1852.8	2457.94	2864.39
Kalimantan Selatan	2246.41	2402.05	3267.33	3569.36	3488.07	2970.9	2854.58	3514.24	3650.48	3804.5
Kalimantan Timur	69.22	61.88	69.29	59.31	70.16	95.08	103.34	93.11	92.86	92.87

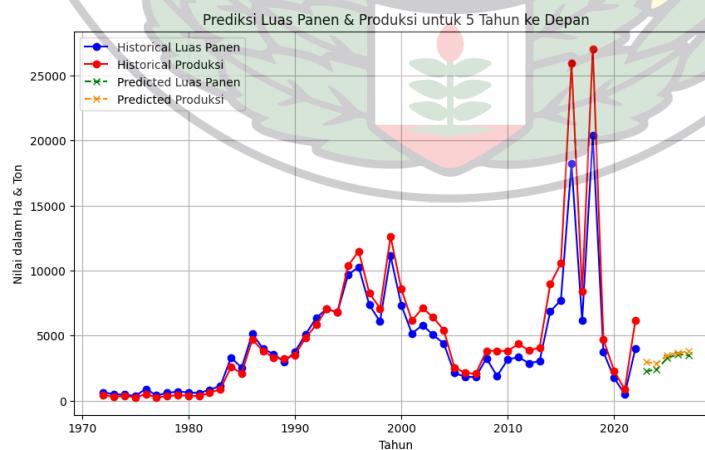
Provinsi	Prediksi Luas Panen dan Produksi Tanaman Kedelai di Kalimantan									
	Luas Panen (Ha)					Produksi (Ton)				
	2022	2023	2024	2025	2026	2022	2023	2024	2025	2026
Kalimantan Utara	2.84	2.84	2.84	2.84	2.84	4.59	4.68	4.59	4.68	4.59



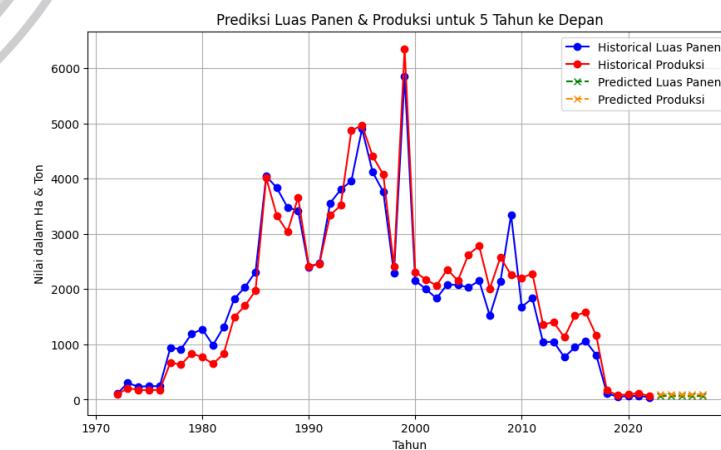
a) Kalimantan Barat



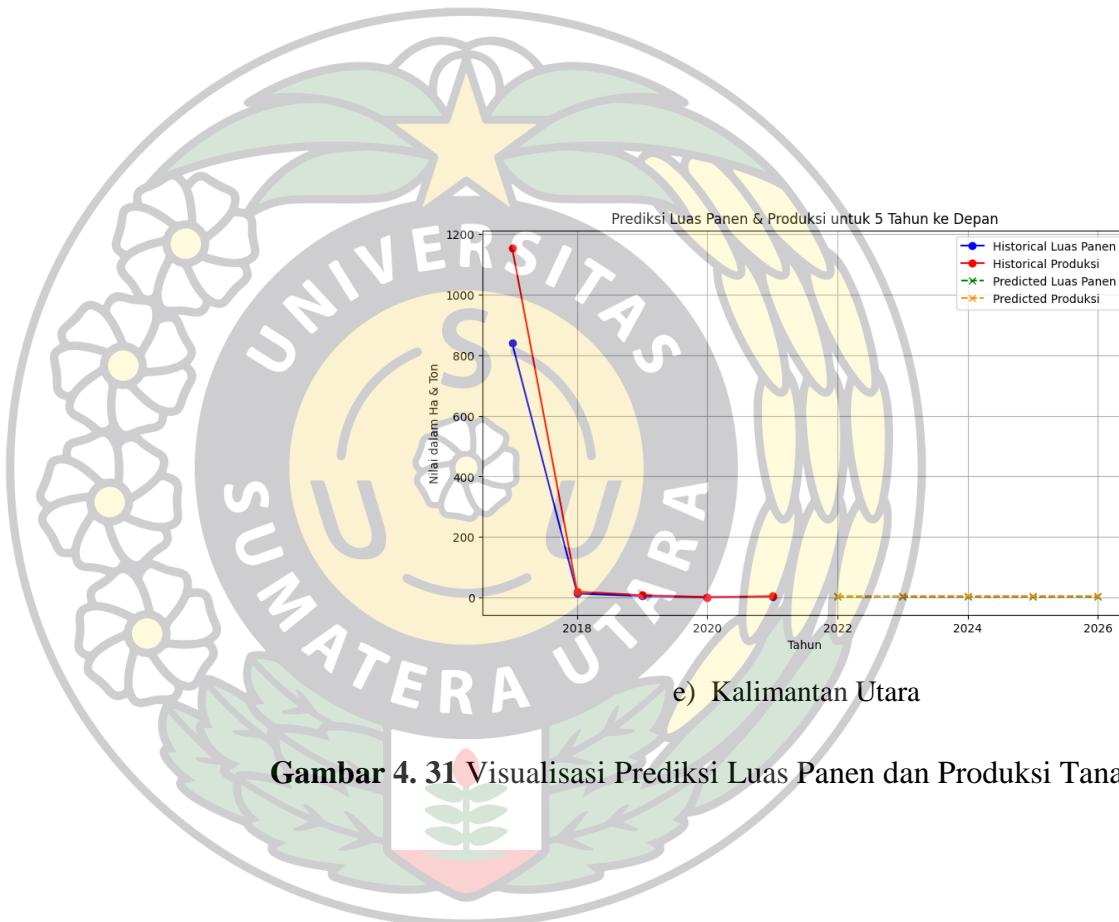
b) Kalimantan Tengah



c) Kalimantan Selatan



d) Kalimantan Timur

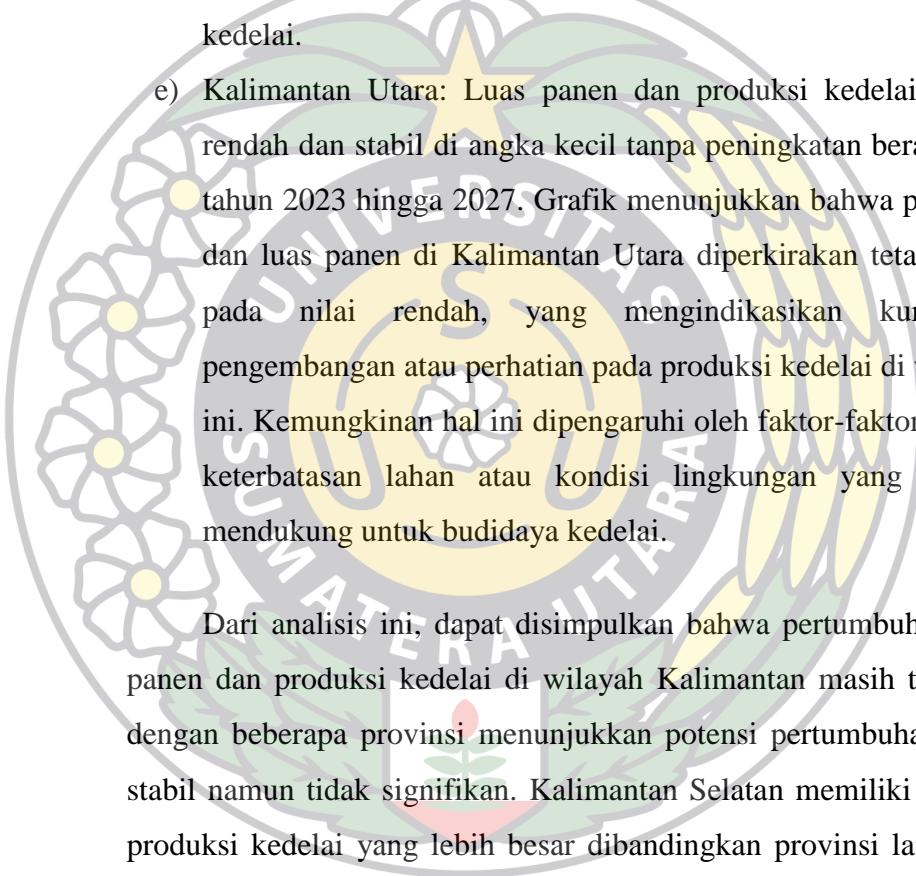


e) Kalimantan Utara

Gambar 4. 31 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Kedelai

Berdasarkan Tabel 4.15 dan Gambar 4.31, berikut adalah analisis hasil prediksi luas panen dan produksi tanaman kedelai di wilayah Kalimantan:

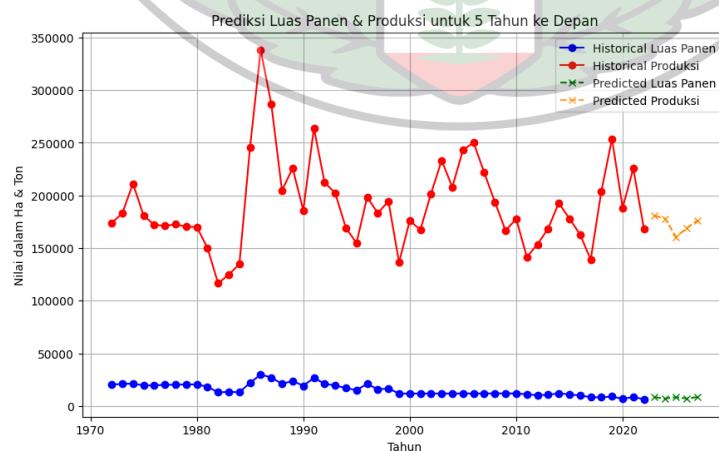
- a) Kalimantan Barat: Luas panen dan produksi kedelai di Kalimantan Barat diprediksi tetap stabil pada kisaran rendah, dengan sedikit peningkatan di tahun 2023. Grafik menunjukkan tren yang konsisten tanpa perubahan signifikan yang menandakan bahwa wilayah ini mungkin kurang optimal untuk budidaya kedelai atau kurang mendapat fokus dalam pengembangan produksi kedelai. Secara umum, tren produksi kedelai tidak menunjukkan pertumbuhan yang signifikan, yang mungkin menunjukkan terbatasnya faktor pendukung produksi kedelai di Kalimantan Barat.
- b) Kalimantan Tengah: Hasil prediksi luas panen maupun produksi kedelai menunjukkan peningkatan yang stabil dari tahun 2023 hingga 2027. Grafik menunjukkan pola fluktuasi yang cukup besar pada periode sebelum 2020, diikuti oleh kenaikan yang lebih stabil pada prediksi untuk beberapa tahun ke depan. Peningkatan ini mungkin menunjukkan adanya inisiatif atau upaya untuk meningkatkan produksi kedelai, meskipun potensi pertumbuhannya masih terbatas.
- c) Kalimantan Selatan: Prediksi menunjukkan peningkatan yang stabil dalam luas panen dan produksi kedelai, dengan angka yang cenderung lebih tinggi dibandingkan provinsi lainnya di Kalimantan. Grafik menunjukkan lonjakan signifikan pada produksi kedelai setelah tahun 2020, yang kemudian stabil pada angka yang cukup tinggi. Tren ini mungkin mencerminkan penggunaan teknologi di sektor pertanian kedelai di Kalimantan Selatan yang menjadikannya salah satu wilayah dengan potensi produksi yang lebih besar.

- 
- d) Kalimantan Timur: Hasil prediksi memperlihatkan bahwa luas panen dan produksi kedelai di Kalimantan Timur relatif rendah dan mengalami fluktuasi ringan tanpa adanya peningkatan signifikan. Grafik memperlihatkan pola yang stabil namun rendah, tanpa adanya lonjakan berarti. Wilayah ini tampaknya memiliki tingkat produksi yang terbatas untuk kedelai, yang kemungkinan disebabkan oleh kondisi yang kurang mendukung atau prioritas yang lebih rendah dalam pengembangan produksi kedelai.
 - e) Kalimantan Utara: Luas panen dan produksi kedelai sangat rendah dan stabil di angka kecil tanpa peningkatan berarti dari tahun 2023 hingga 2027. Grafik menunjukkan bahwa produksi dan luas panen di Kalimantan Utara diperkirakan tetap stabil pada nilai rendah, yang mengindikasikan kurangnya pengembangan atau perhatian pada produksi kedelai di wilayah ini. Kemungkinan hal ini dipengaruhi oleh faktor-faktor seperti keterbatasan lahan atau kondisi lingkungan yang kurang mendukung untuk budidaya kedelai.

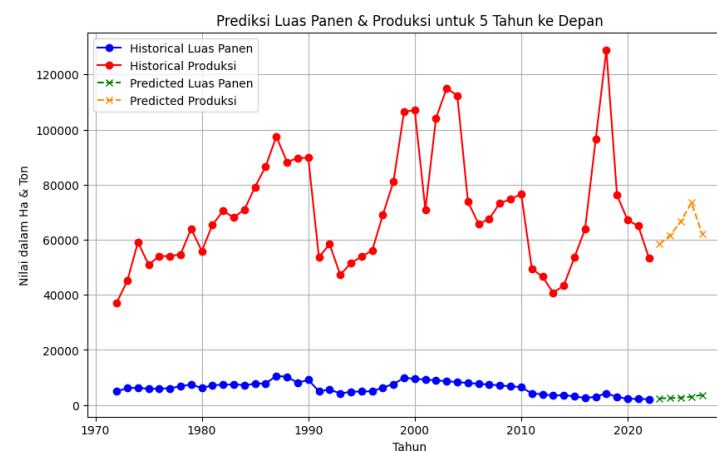
Dari analisis ini, dapat disimpulkan bahwa pertumbuhan luas panen dan produksi kedelai di wilayah Kalimantan masih terbatas, dengan beberapa provinsi menunjukkan potensi pertumbuhan yang stabil namun tidak signifikan. Kalimantan Selatan memiliki potensi produksi kedelai yang lebih besar dibandingkan provinsi lain yang menunjukkan bahwa daerah ini dapat menjadi fokus untuk pengembangan kedelai di Kalimantan. Sementara itu, Kalimantan Utara dan Kalimantan Timur menunjukkan luas panen dan produksi yang sangat rendah, yang mengindikasikan bahwa produksi kedelai mungkin bukan komoditas unggulan di daerah ini.

Tabel 4. 16 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Ubi Kayu

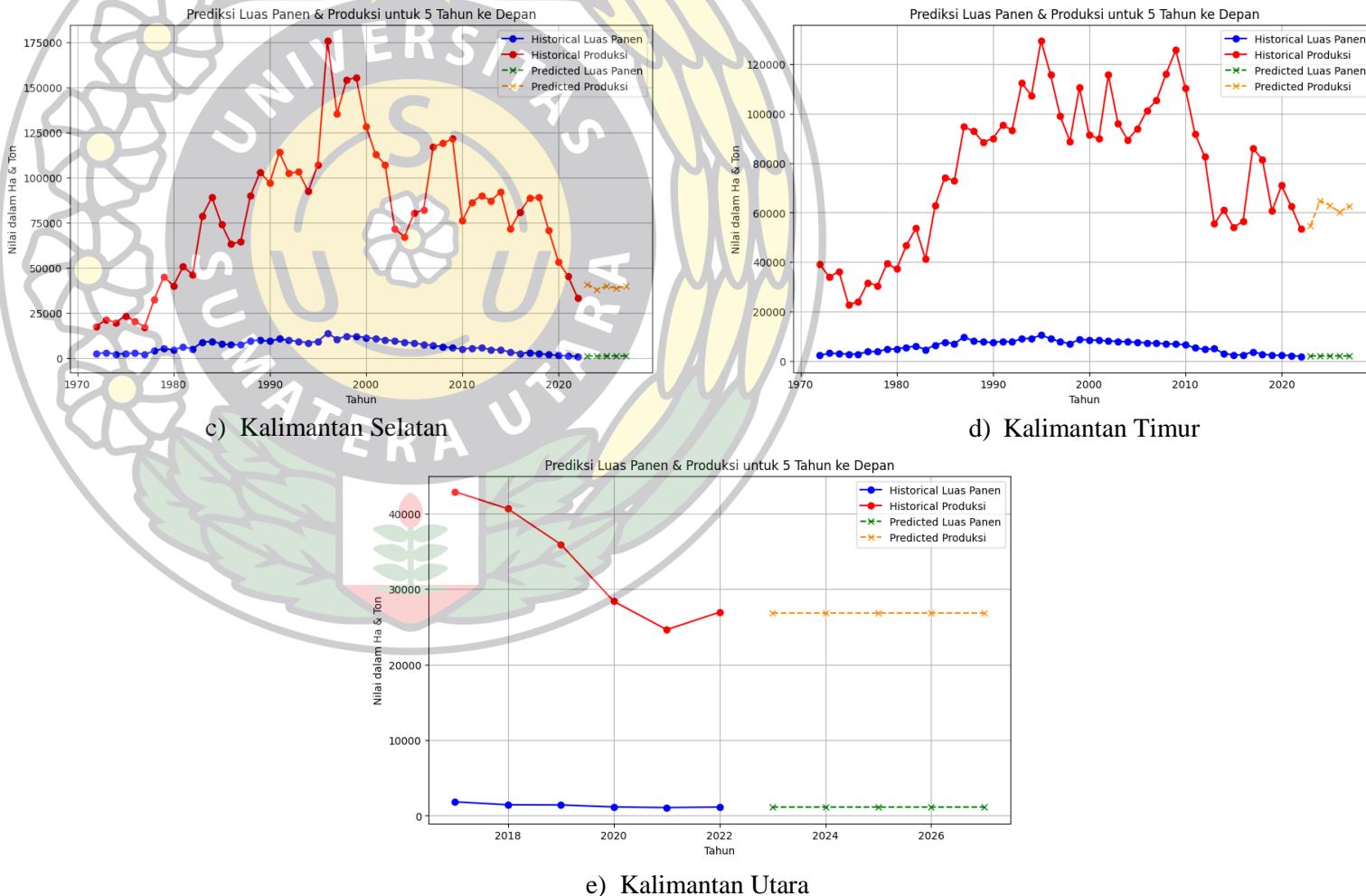
Provinsi	Prediksi Luas Panen dan Produksi Tanaman Ubi Kayu di Kalimantan									
	Luas Panen (Ha)					Produksi (Ton)				
	2023	2024	2025	2026	2027	2023	2024	2025	2026	2027
Kalimantan Barat	8420.82	7113.85	8420.82	7113.85	8420.82	180822.02	177947.9	160680.52	168764.2	176270.47
Kalimantan Tengah	2426.72	2481.95	2726.6	3013.67	3668.67	58421.89	61645.44	66703.55	73458.16	62075.56
Kalimantan Selatan	1293.55	1293.55	1293.55	1293.55	1293.55	41113.95	38085.26	40045.9	38895.27	40045.9
Kalimantan Timur	2074.88	2074.88	2074.88	2074.88	2074.88	54744.53	64745.68	62988.32	60212.02	62581.25
Kalimantan Utara	1116.05	1116.05	1116.05	1116.05	1116.05	26822.95	26822.95	26822.95	26822.95	26822.95



a) Kalimantan Barat



b) Kalimantan Tengah



Gambar 4. 32 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Ubi Kayu

Berdasarkan Tabel 4.16 dan Gambar 4.32, berikut adalah analisis hasil prediksi luas panen dan produksi tanaman ubi kayu di wilayah Kalimantan:

- a) Kalimantan Barat: Luas panen ubi kayu di Kalimantan Barat diprediksi tetap stabil pada angka sekitar 8.420 hektar, sementara produksi berkisar di angka 16.000 hingga 18.000 ton hingga tahun 2027. Grafik menunjukkan pola fluktuasi yang konsisten pada tahun-tahun sebelumnya, namun tren prediksi ke depan relatif stabil tanpa peningkatan signifikan. Pola ini menunjukkan bahwa wilayah ini mungkin tidak difokuskan untuk pengembangan budidaya ubi kayu, atau faktor lingkungan yang kurang mendukung peningkatan produktivitas ubi kayu.
- b) Kalimantan Tengah: Prediksi luas panen menunjukkan peningkatan secara bertahap pada sekitar 2.700 hektar, sedangkan produksi berfluktuasi sekitar 62.000 hingga 67.000 ton. Grafik memperlihatkan pola fluktuatif tanpa peningkatan yang berarti. Hal ini mungkin menunjukkan bahwa meskipun terdapat potensi, perkembangan sektor budidaya ubi kayu di Kalimantan Tengah belum didorong dengan intensif atau adanya kendala dalam meningkatkan produksi lebih lanjut.
- c) Kalimantan Selatan: Luas panen ubi kayu di Kalimantan Selatan diperkirakan stabil pada kisaran 1.300 hektar, sementara produksi berada di kisaran 30.000 hingga 40.000 ton dalam beberapa tahun mendatang. Grafik menunjukkan penurunan produksi yang signifikan di sekitar tahun 2020, tetapi stabil pada prediksi tahun-tahun berikutnya. Stabilitas prediksi ke depan menunjukkan bahwa ekspansi atau peningkatan produksi yang lebih besar untuk tahun mendatang mungkin terbatas.
- d) Kalimantan Timur: Hasil prediksi menunjukkan bahwa luas panen ubi kayu di Kalimantan Timur berada pada angka 2.074 hektar, sementara produksi diprediksi stabil sekitar 62.000 ton.

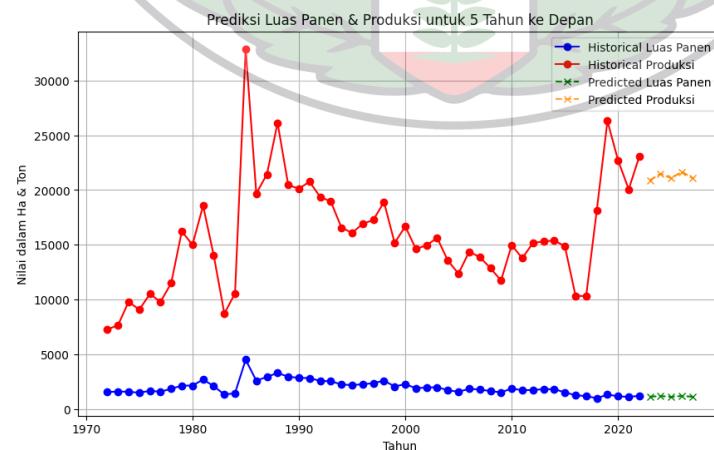
Visualisasi menunjukkan pola stabilitas dengan fluktuasi kecil tanpa adanya lonjakan besar dalam produksi. Wilayah ini tampaknya memiliki tingkat produksi yang stabil namun terbatas, mungkin disebabkan oleh kondisi lingkungan atau pengembangan ubi kayu bukan merupakan prioritas utama.

- e) Kalimantan Utara: Luas panen dan produksi ubi kayu diprediksi tetap rendah, masing-masing sekitar 1.116 hektar dan 26.822 ton hingga tahun 2027. Grafik menunjukkan stabilitas pada tingkat produksi yang rendah tanpa ada peningkatan yang signifikan, hal ini menandakan kurangnya inisiatif pengembangan di sektor ubi kayu di wilayah ini. Keterbatasan produksi ini mungkin disebabkan oleh kondisi lingkungan yang kurang optimal atau fokus pada komoditas lain yang lebih cocok untuk wilayah tersebut.

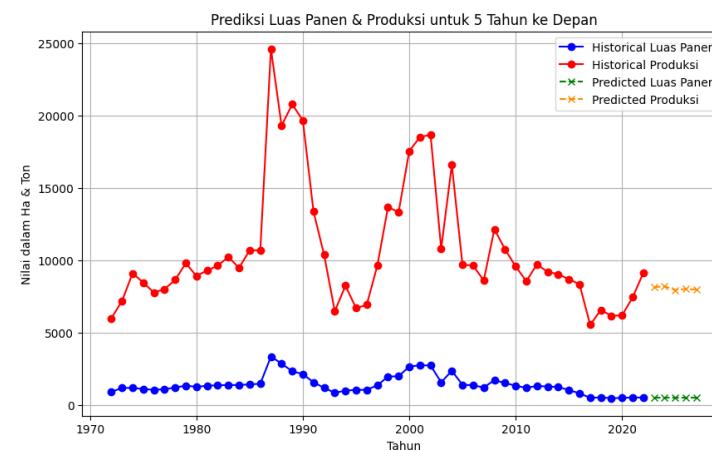
Dari analisis ini, dapat disimpulkan bahwa pertumbuhan luas panen dan produksi ubi kayu di Kalimantan masih terbatas, dengan pola prediksi menunjukkan stabilitas tanpa adanya peningkatan yang signifikan di beberapa provinsi. Kalimantan Selatan memiliki potensi produksi ubi kayu yang lebih tinggi dibandingkan provinsi lain, yang mungkin bisa menjadi wilayah prioritas untuk pengembangan lebih lanjut. Sementara itu, Kalimantan Utara memiliki luas panen dan produksi yang sangat rendah, menunjukkan bahwa produksi ubi kayu bukan komoditas utama di daerah ini. Secara keseluruhan, prediksi ini menunjukkan bahwa produksi ubi kayu di Kalimantan masih memiliki ruang untuk pengembangan, tetapi saat ini kapasitas produksi masih terbatas pada beberapa wilayah tertentu.

Tabel 4. 17 Hasil Prediksi Luas Panen dan Produksi untuk Tanaman Ubi Jalar

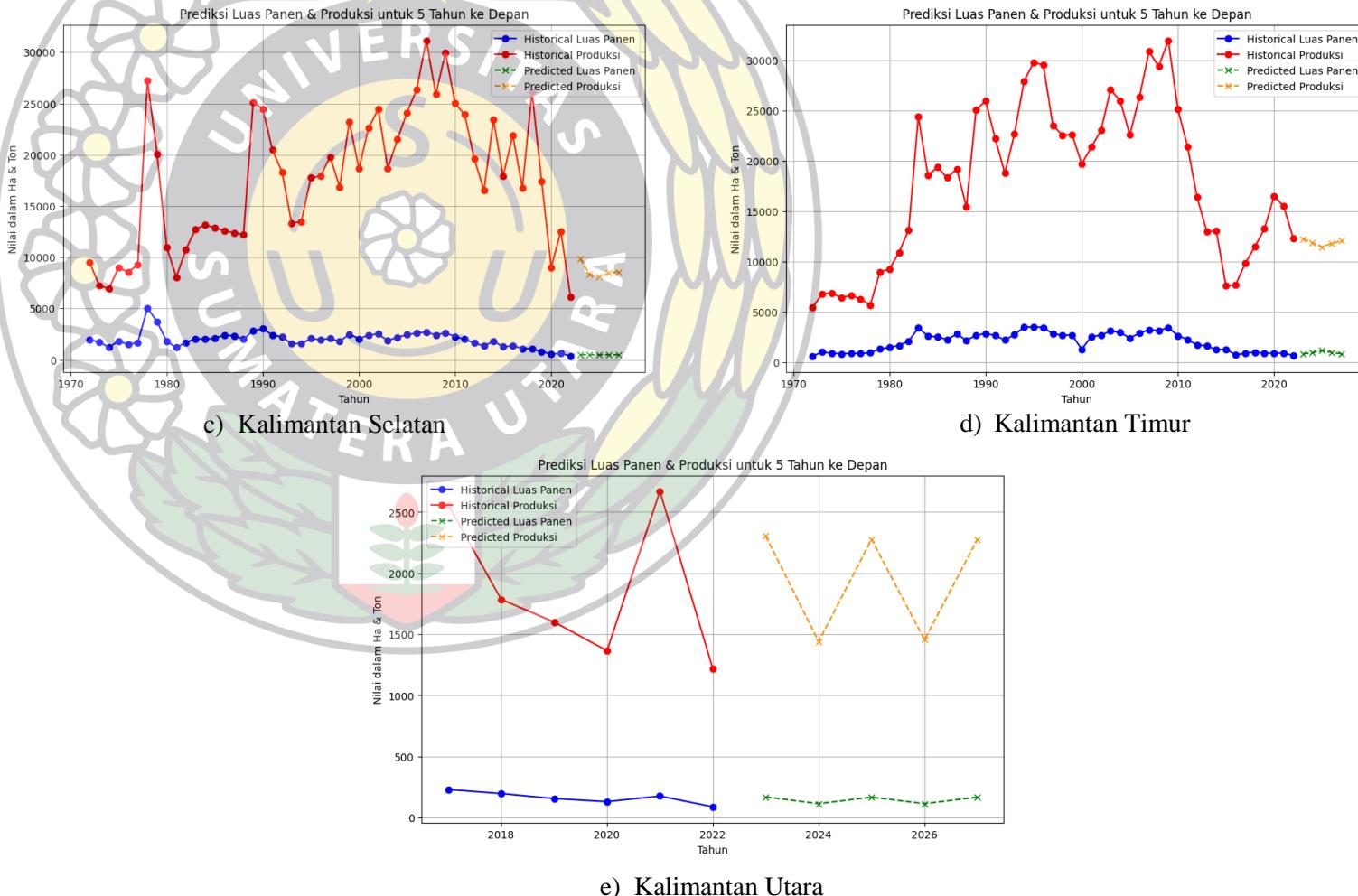
Provinsi	Prediksi Luas Panen dan Produksi Tanaman Ubi Jalar di Kalimantan									
	Luas Panen (Ha)					Produksi (Ton)				
	2023	2024	2025	2026	2027	2023	2024	2025	2026	2027
Kalimantan Barat	1130.24	1182.69	1112.42	1197.61	1121.38	20896.64	21493.83	21076.94	21662.76	21076.94
Kalimantan Tengah	525.77	529.19	525.74	529.19	525.74	8153.42	8209.92	7961.79	8024.32	7993.75
Kalimantan Selatan	455.35	455.35	455.35	455.35	455.35	9849.17	8360.17	8107.85	8455.34	8572.33
Kalimantan Timur	826.0	917.22	1138.51	952.88	827.8	12247.13	11843.3	11431.11	11787.0	12032.51
Kalimantan Utara	168.79	114.49	166.93	114.49	166.93	2303.89	1440.23	2276.55	1454.76	2276.55



a) Kalimantan Barat



b) Kalimantan Tengah



Gambar 4. 33 Visualisasi Prediksi Luas Panen dan Produksi Tanaman Ubi Jalar

Berdasarkan Tabel 4.17 dan Gambar 4.33, berikut adalah analisis hasil prediksi luas panen dan produksi tanaman ubi jalar di wilayah Kalimantan:

- a) Kalimantan Barat: Luas panen ubi jalar di Kalimantan Barat diprediksi berada pada kisaran stabil sekitar 1.130 hektar, sementara produksi menunjukkan peningkatan sedikit dari sekitar 20.800 ton pada tahun 2023 menjadi sekitar 21.076 ton pada tahun 2027. Grafik menunjukkan pola produksi yang relatif stabil, dengan sedikit kenaikan pada prediksi beberapa tahun mendatang. Tren ini mungkin mengindikasikan bahwa Kalimantan Barat memiliki kapasitas produksi ubi jalar yang stabil namun tidak akan mengalami pertumbuhan besar dalam waktu dekat.
- b) Kalimantan Tengah: Luas panen ubi jalar diprediksi stabil pada kisaran 525 hektar, sementara produksi ubi jalar menunjukkan kestabilan dengan sedikit variasi di sekitar 8.000 ton. Grafik menunjukkan pola fluktuasi sebelum periode prediksi, namun diikuti oleh prediksi yang lebih stabil pada tahun-tahun ke depan. Meskipun produksi diperkirakan meningkat sedikit, pola ini menunjukkan bahwa Kalimantan Tengah mungkin tidak difokuskan untuk ekspansi besar dalam produksi ubi jalar.
- c) Kalimantan Selatan: Luas panen di Kalimantan Selatan diperkirakan tetap pada angka sekitar 455 hektar, dengan produksi diprediksi akan mengalami penurunan secara bertahap dari tahun 2024 hingga 2025, kemudian terjadi peningkatan di tahun-tahun berikutnya. Grafik menunjukkan pola yang relatif stabil pada angka rendah, dengan sedikit kenaikan di beberapa tahun terakhir prediksi. Pola ini mengindikasikan bahwa Kalimantan Selatan mungkin tidak memprioritaskan pengembangan ubi jalar, dengan potensi pertumbuhan yang terbatas pada sektor ini.

- d) Kalimantan Timur: Luas panen dan produksi ubi jalar di Kalimantan Timur diperkirakan berada pada angka rendah dan relatif stabil, dengan luas panen sekitar 826 hektar dan produksi berkisar di angka 11.000 hingga 12.000 ton. Grafik menunjukkan pola yang relatif stabil, dengan fluktuasi kecil tanpa adanya lonjakan yang tajam. Wilayah ini tampaknya memiliki tingkat produksi ubi jalar yang terbatas dan stabil, yang mungkin dipengaruhi oleh faktor lingkungan atau prioritas komoditas lain.
- e) Kalimantan Utara: Luas panen dan produksi ubi jalar diprediksi mengalami peningkatan dan penurunan yang berkisar antara 114 hingga 169 hektar dengan rentang produksi antara 1.400 hingga 2.300 ton. Grafik menunjukkan pola naik-turun secara bergantian pada angka rendah untuk beberapa tahun mendatang. Produksi ubi jalar di Kalimantan Utara tampaknya sangat terbatas, menunjukkan bahwa wilayah ini bukan area utama untuk budidaya ubi jalar.

Dari analisis ini, dapat disimpulkan bahwa prediksi luas panen dan produksi ubi jalar di seluruh provinsi Kalimantan menunjukkan stabilitas tanpa adanya peningkatan signifikan. Provinsi-provinsi di Kalimantan menunjukkan kapasitas produksi ubi jalar yang relatif rendah, dengan Kalimantan Barat menjadi provinsi dengan kapasitas tertinggi, namun tetap pada level stabil. Kalimantan Utara memiliki tingkat produksi yang sangat rendah dan tidak mengalami kenaikan yang signifikan, menandakan bahwa ubi jalar mungkin bukan komoditas utama di wilayah ini. Secara keseluruhan, prediksi ini menunjukkan bahwa produksi ubi jalar di Kalimantan masih memiliki ruang untuk pengembangan, namun kapasitanya masih terbatas dan tidak mengalami pertumbuhan yang signifikan.

Selanjutnya, penelitian ini akan berupaya mengaitkan hasil prediksi produksi tanaman pangan dengan implementasi proyek *food estate* yang berlangsung di Kalimantan. Adapun wilayah yang menjadi prioritas di Kalimantan ialah provinsi Kalimantan Tengah. Di wilayah ini, proyek tersebut difokuskan pada pengembangan lahan untuk produksi berbagai komoditas pangan seperti padi, jagung, dan singkong/ubi kayu. Berdasarkan data produksi tanaman padi di Kalimantan Tengah dari tahun 2020 hingga 2023, terlihat adanya tren penurunan yang konsisten dalam jumlah produksi. Penurunan yang signifikan ini terjadi selama pelaksanaan proyek *food estate*, dan prediksi untuk periode 2024 hingga 2028 juga menunjukkan kecenderungan yang stabil, yang mengindikasikan bahwa produksi padi diperkirakan tidak akan mengalami perubahan yang signifikan dalam beberapa tahun mendatang. Selain padi, produksi jagung juga menunjukkan tren yang relatif stabil sejak tahun 2020, dengan hasil prediksi untuk periode 2024 hingga 2028 yang menunjukkan bahwa produksi jagung akan tetap konstan. Sementara itu, produksi ubi kayu dari tahun 2020 hingga 2022 mengalami tren penurunan bertahap, dan hasil prediksi untuk periode 2023 hingga 2027 menunjukkan adanya fluktuasi yang menandakan ketidakstabilan dalam produksi.

Fenomena ini mencerminkan tantangan yang cukup besar dalam upaya untuk meningkatkan produksi padi, jagung, dan ubi kayu di Kalimantan Tengah. Beberapa faktor yang mungkin mempengaruhi tren ini antara lain perubahan iklim, keterbatasan lahan subur, dan faktor lain yang berdampak pada produktivitas pertanian. Oleh karena itu, untuk mengatasi tantangan tersebut, diperlukan perhatian lebih terhadap kebijakan pertanian yang dapat mendukung keberlanjutan produksi pangan, seperti peningkatan sistem irigasi, penggunaan varietas bibit unggul, serta penerapan teknologi pertanian yang dapat meningkatkan efisiensi dan hasil produksi tanaman pangan di masa depan.

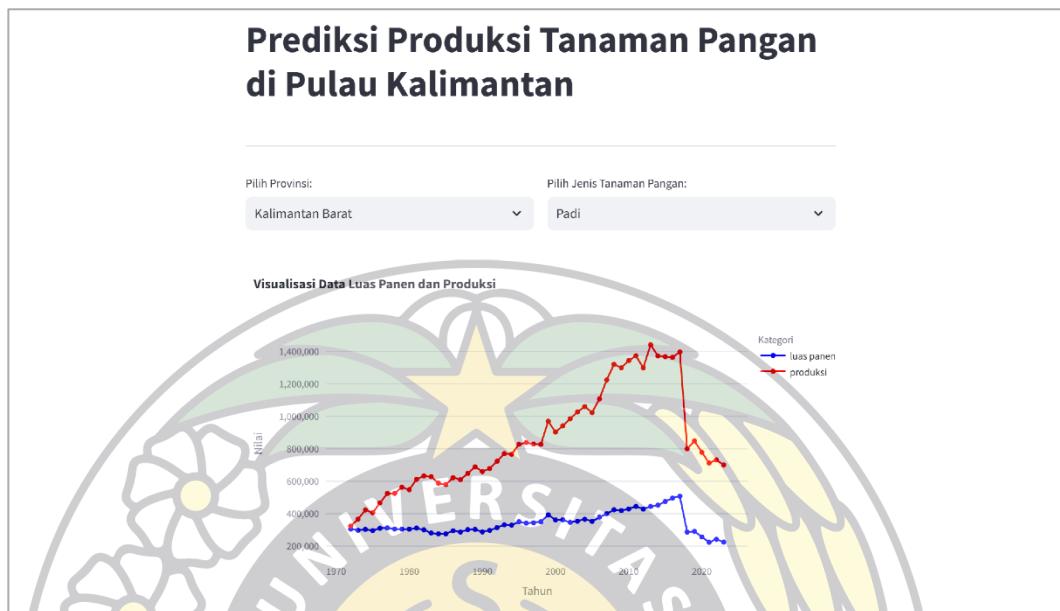
Di sisi lain prediksi produksi tanaman kedelai menunjukkan adanya tren peningkatan yang stabil. Tren ini mengindikasikan bahwa kedelai memiliki potensi untuk berkembang lebih lanjut seiring dengan program *food estate* yang bertujuan meningkatkan kapasitas produksi pangan. Kedelai sebagai sumber protein nabati yang penting dapat menjadi alternatif yang baik untuk meningkatkan keragaman produksi pangan, sehingga tidak hanya bergantung pada padi, jagung, ataupun ubi kayu.

Namun, kondisi wilayah Kalimantan yang didominasi oleh lahan gambut dan tanah podsilik dengan tingkat keasaman tinggi serta kandungan hara rendah menjadi tantangan tersendiri dalam pengembangan tanaman kedelai. Tanaman ini membutuhkan tanah dengan pH netral hingga sedikit asam serta tingkat kesuburan yang memadai untuk pertumbuhan optimal. Untuk mengatasi tantangan tersebut, diperlukan kolaborasi penelitian dengan sektor pertanian guna mengembangkan solusi yang komprehensif. Kolaborasi ini diharapkan mampu menghasilkan inovasi dalam pengelolaan lahan, penerapan teknologi yang tepat guna, serta strategi agronomi yang adaptif terhadap kondisi lokal Kalimantan. Dengan demikian, keberlanjutan produksi pangan di wilayah ini dapat terus didukung dan ditingkatkan.

4.9 Implementasi Visualisasi Hasil Prediksi pada Aplikasi Website

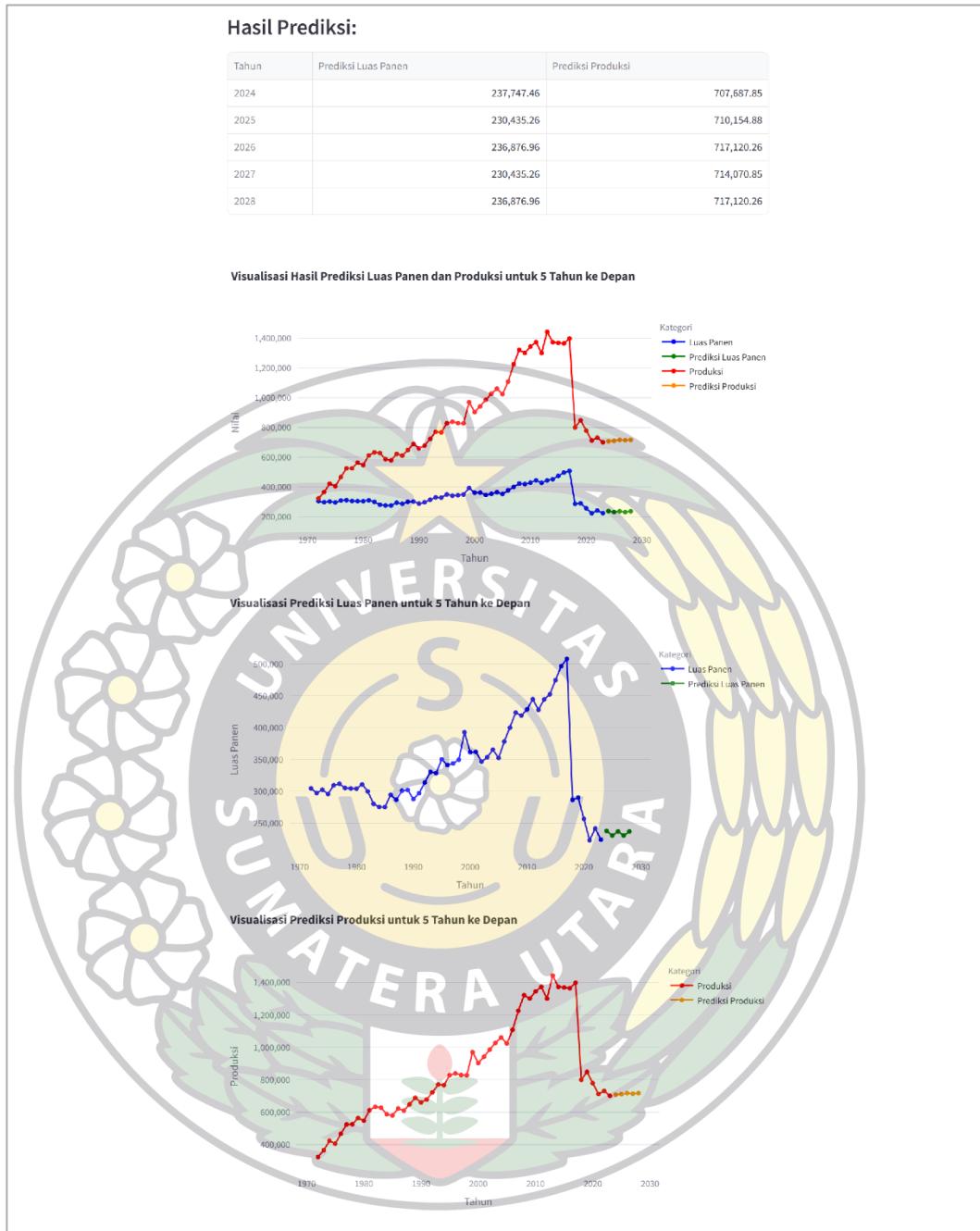
Pada tahap ini, model prediksi menggunakan algoritma *Random Forest* akan diterapkan dan diimplementasikan ke dalam sebuah aplikasi berbasis *website*. Model yang telah dikembangkan akan disimpan dan diunduh dari lingkungan *Google Colaboratory* ke dalam berkas dengan format *Joblib* atau *.pkl*. Selanjutnya, model tersebut akan digunakan dalam kode program aplikasi. Aplikasi ini dibangun menggunakan bahasa pemrograman *Python* dan *framework Streamlit*. Melalui antarmuka aplikasi, pengguna dapat memilih wilayah provinsi dan jenis tanaman yang diinginkan, seperti yang

terlihat pada Gambar 4.34. Setelah itu, aplikasi akan menampilkan hasil prediksi berdasarkan wilayah dan jenis tanaman yang telah dipilih.



Gambar 4. 34 Tampilan Antarmuka Aplikasi Prediksi Produksi

Aplikasi ini dirancang untuk menampilkan hasil prediksi secara informatif, dengan menampilkan grafik yang memudahkan pengguna dalam memahami pola produksi tanaman pangan dari berbagai wilayah. Grafik ini akan memperlihatkan tren dan fluktuasi produksi, sehingga memungkinkan pengguna untuk menganalisis serta membuat keputusan yang lebih baik berdasarkan informasi tersebut. Selain grafik, hasil prediksi juga akan disajikan dalam bentuk tabel data. Tabel ini akan memungkinkan pengguna untuk mengeksplorasi data secara lebih mendalam, melihat angka-angka spesifik, dan membandingkan hasil antara berbagai wilayah dan jenis tanaman. Dengan demikian, aplikasi ini tidak hanya memberikan informasi yang bermanfaat, tetapi juga memungkinkan pengguna untuk menganalisis dan memahami data prediksi secara lebih komprehensif. Tampilan antarmuka aplikasi untuk hasil prediksi dapat dilihat pada Gambar 4.35 di bawah ini.



Gambar 4.35 Tampilan Antarmuka Hasil Prediksi

BAB V

PENUTUP

5. 1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat disimpulkan beberapa hal berikut :

- 1) Berdasarkan analisis korelasi antara luas panen dan produksi pangan, hasil menunjukkan bahwa terdapat hubungan signifikan antara luas panen dengan jumlah produksi. Dengan menggunakan metode korelasi *Pearson*, ditemukan bahwa semakin luas area panen, secara umum produksi pangan akan meningkat, meskipun terdapat variasi yang disebabkan oleh faktor eksternal lainnya.
- 2) Hasil perbandingan kinerja algoritma *Linear Regression* dan *Random Forest* menunjukkan bahwa model *Random Forest* secara konsisten memiliki kinerja yang lebih baik dibandingkan model *Linear Regression* dalam memprediksi produksi berbagai jenis tanaman pangan, termasuk padi, jagung, kedelai, ubi kayu, dan ubi jalar di seluruh wilayah Kalimantan. Model *Random Forest* menghasilkan nilai MAE dan RMSE yang lebih rendah serta nilai R^2 yang lebih tinggi di sebagian besar provinsi. Hal ini menunjukkan akurasi dan konsistensi yang lebih baik dalam memprediksi produksi tanaman pangan dibandingkan model *Linear Regression*.
- 3) Penggunaan model *Random Forest* dalam prediksi produksi pangan dapat membantu menghasilkan estimasi yang akurat, sehingga diharapkan dapat mendukung pemangku kepentingan dalam membuat perencanaan strategis guna mencegah kerawanan pangan akibat ketidakstabilan ketersediaan pangan di wilayah Kalimantan.
- 4) Produksi kedelai di Kalimantan Tengah diprediksi akan meningkat secara stabil, menunjukkan bahwa tanaman kedelai memiliki potensi besar untuk berkembang lebih lanjut seiring dengan pelaksanaan program *food estate*.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, berikut beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya:

- 1) Menggunakan data yang lebih luas, baik dari waktu yang lebih panjang maupun penggunaan data harian atau bulanan, sehingga akan memberikan pemahaman yang lebih mendalam tentang pola produksi pangan di Pulau Kalimantan. Ini juga dapat membantu dalam mengidentifikasi tren jangka panjang yang mungkin tidak terlihat dengan data yang terbatas.
- 2) Penelitian selanjutnya dapat mempertimbangkan untuk mengembangkan model prediksi yang lebih kompleks dengan memasukkan variabel tambahan yang mempengaruhi produksi pangan. Faktor-faktor seperti perubahan iklim, teknologi pertanian, jenis bibit, metode pengolahan lahan, dan praktik pemupukan dapat menjadi pertimbangan penting.
- 3) Melakukan perbandingan dengan algoritma *machine learning* lainnya seperti *Support Vector Machine* (SVM), *Decision Tree*, *Gradient Boosting*, dan lainnya yang dapat memberikan wawasan lebih dalam mengenai keefektifan masing-masing metode dalam memprediksi produksi pangan. Dengan mencoba berbagai algoritma, peneliti dapat menemukan model yang paling sesuai dengan karakteristik data yang digunakan.
- 4) Melakukan kolaborasi penelitian dengan sektor pertanian untuk mengembangkan solusi komprehensif yang mendukung keberlanjutan produksi pangan di wilayah Kalimantan.

DAFTAR PUSTAKA

- Acharya, M. S., Armaan, A., & Antony, A. S. (2019, February). A Comparison Of Regression Models For Prediction Of Graduate Admissions. In *2019 international conference on computational intelligence in data science (ICCIDDS)* (pp. 1-5). IEEE.
- Alita, D., Putra, A. D., & Darwis, D. (2021). Analysis Of Classic Assumption Test And Multiple Linear Regression Coefficient Test For Employee Structural Office Recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 295-306.
- Alonso, E. B., Cockx, L., & Swinnen, J. (2018). Culture And Food Security. *Global food security*, 17, 113-127.
- Ariani, M., & Suryana, A. (2023). Kinerja Ketahanan Pangan Indonesia: Pembelajaran dari Penilaian dengan Kriteria Global dan Nasional. *Analisis Kebijakan Pertanian*, 21(1), 1-20.
- Chaireni, R., Agustanto, D., Wahyu, R. A., & Nainggolan, P. (2020). Ketahanan Pangan Berkelanjutan. *Jurnal Kependudukan Dan Pembangunan Lingkungan*, 1(2), 70-79.
- Christyanto, M., & Mayulu, H. (2021). Pentingnya Pembangunan Pertanian Dan Pemberdayaan Petani Wilayah Perbatasan Dalam Upaya Mendukung Ketahanan Pangan Nasional: Studi kasus di wilayah perbatasan Kalimantan. *Journal of Tropical AgriFood*, 3(1), 1-14.
- Deputi Bidang Kerawanan Pangan dan Gizi Badan Pangan Nasional. (2023). *Indeks Ketahanan Pangan 2023*.
- Direktorat Statistik Tanaman Pangan, Hortikultura, dan Perkebunan. (2022). *Analisis Produktivitas Jagung dan Kedelai di Indonesia 2022 (Hasil Survei Ubinan)*. Jakarta: Badan Pusat Statistik.
- Fatima, S. S. W., & Rahimi, A. (2024). A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems. *Machines*, 12(6), 380.
- Fu, X., Chen, Y., Yan, J., Chen, Y., & Xu, F. (2023). BGRF: A Broad Granular Random Forest Algorithm. *Journal of Intelligent & Fuzzy Systems*, 44(5), 8103-

- 8117.
- Gao, Y. (2024). The Prediction Of Apple Stock Price Based On Linear Regression Model And Random Forest Model. *Theoretical and Natural Science*. 30, 103-109.
- Hadi, A., Rusli, B., & Alexandri, M. B. (2019). Dampak Undang-Undang Nomor 12 Tentang Pangan Terhadap Ketahanan Pangan Indonesia. *Responsive*, 2(4), 173-181.
- Hu, Yaowen. (2024). Comparison and Analysis of the Effectiveness of Linear Regression, Decision Tree, and Random Forest Models for Health Insurance Premium Forecasting. *Advances in Economics, Management and Political Sciences*. 79. 347-353.
- Hussein, A., Agbinya, J., & Satti, I. (2020). A Survey on Data mining Techniques for Water Flow Forecasting. *Australian Journal of Basic and Applied Sciences*, 14(3), 13-27.
- Jassim, M. A., & Abdulwahid, S. N. (2021, March). Data Mining Preparation: Process, Techniques And Major Issues In Data Analysis. In *IOP conference series: materials science and engineering* (Vol. 1090, No. 1, p. 012053). IOP Publishing.
- Kolambe, M., & Arora, S. (2024). Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques. *Journal of Electrical Systems*, 20, 575-586.
- Li, H., Lin, J., Lei, X., & Wei, T. (2022). Compressive Strength Prediction Of Basalt Fiber Reinforced Concrete Via Random Forest Algorithm. *Materials Today Communications*, 30, 103117.
- Mandour, S., Gamal, A., Sleem, A., & Belal, M. (2024). Data Mining Problems Optimization by using Metaheuristic Algorithms: A Survey. *Multicriteria Algorithms with Applications*, 4, 28-52.
- Munandar, T. A., Sumiati, S., & Rosalina, V. (2020, April). Pattern Of Symptom Correlation On Type Of Heart Disease Using Approach Of Pearson Correlation Coefficient. In *IOP Conference Series: Materials Science and Engineering* (Vol. 830, No. 2, p. 022086). IOP Publishing.

- Negri, F. (2023). Correlation Is Not Causation, Yet... Matching And Weighting For Better Counterfactuals. In *Causality in Policy Studies: a Pluralist Toolbox* (pp. 71-98). Cham: Springer International Publishing.
- Pahlevi, M. R., Kusrini, K., & Hidayat, T. (2023). Comparison of LSTM and GRU Models for Forex Prediction. *Sinkron: jurnal dan penelitian teknik informatika*, 7(4), 2254-2263.
- Papakyriakou, D., & Barbounak, I. S. (2022). Data Mining Methods: A Review. *International Journal of Computer Applications*, 183 (48), 5-19.
- Pashankar, S. S., Shendage, J. D., & Pawar, J. (2024). Machine Learning Techniques For Stock Price Prediction-A Comparative Analysis Of Linear Regression, Random Forest, And Support Vector Regression. *Journal of Advanced Zoology*, 45(S-4), 118-127.
- Pemerintah Provinsi Kalimantan Timur. (2019). *Statistik Kalimantan Tahun 2019*.
- Peng, W. & Berry, M. (2019). The Concept of Food Security. *Encyclopedia of Food Security and Sustainability*, 2, 1-7.
- Prater, R., Hanne, T., & Dornberger, R. (2024). Generalized Performance of LSTM in Time-Series Forecasting. *Applied Artificial Intelligence*, 38(1), 2377510.
- Qu, K. (2024). Research On Linear Regression Algorithm. In *MATEC Web of Conferences* (Vol. 395, p. 01046). EDP Sciences.
- Rasman, A., Theresia, E. S., & Aginda, M. F. (2023). Analisis implementasi program food estate sebagai solusi ketahanan pangan Indonesia. *Holistic: Journal of Tropical Agriculture Sciences*, 1(1).
- Rumra, dkk. (2023). Neraca Bahan Makanan Nasional 2021-2023. Jakarta: Direktorat Ketersediaan Pangan Deputi Bidang Ketersediaan dan Stabilitas Pangan.
- Satria, A., Badri, R. M., & Safitri, I. (2023). Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning. *Digital Transformation Technology*, 3(2), 389-398.
- Sofian, S. R. A. (2022). Analisis Korelasi Curah Hujan dan Produktivitas Tanaman Hasil Pertanian Kabupaten Jember. *Jurnal Pendidikan Mipa*, 12(2), 287-293.
- Wang, L., Yang, Y., Xu, L., & Ji, T. (2024). Application Of Random Forest Algorithm In The Detection Of Foreign Objects In Wine. *Applied Mathematics and*

Nonlinear Sciences.

- Wang, W. (2024). Forecasting The Population of China From 2020 To 2025 Based on Random Forest and Linear Regression. *Highlights in Science, Engineering and Technology*, 85, 511-518.
- Wijaya, S., & Fauziah, F. (2024). A Comparative Analysis Of Linear Regression And Random Forest Methods For Predicting Palm Oil Price Sales. *Jurnal Ilmiah Teknologi dan Rekayasa*, 28(3), 265-274.
- Yang, Q., Kang, Q., Huang, Q., Cui, Z., Bai, Y., & Wei, H. (2021, June). Linear Correlation Analysis Of Ammunition Storage Environment Based On Pearson Correlation Analysis. In *Journal of physics: Conference series* (Vol. 1948, No. 1, p. 012064). IOP Publishing.

