

PERAMALAN HARGA CABAI RAWIT TINGKAT GEROMBOL PROVINSI DI INDONESIA: PENDEKATAN *FEATURE-BASED CLUSTERING TIME SERIES* DAN XGBOOST

HAKIM ZOELVA MAHESA



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025



- 



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Peramalan Harga Cabai Rawit Tingkat Gerombol Provinsi di Indonesia: Pendekatan *Feature-based Clustering Time Series* dan XGBoost” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2025

Hakim Zoelva Mahesa
G1401211039



ABSTRAK

HAKIM ZOELVA MAHESA. Peramalan Harga Cabai Rawit Tingkat Gerombol Provinsi di Indonesia: Pendekatan *Feature-based Clustering Time Series* dan XGBoost. Dibimbing oleh ANANG KURNIA dan I MADE SUMERTAJAYA.

Cabai rawit merupakan salah satu komoditas pertanian strategis di Indonesia yang sering mengalami fluktuasi harga signifikan dan besar pengaruhnya terhadap inflasi pangan dan stabilitas ekonomi nasional. Penelitian ini bertujuan untuk mengefisienkan beban komputasi dan meningkatkan akurasi peramalan harga cabai rawit provinsi di Indonesia dengan mengintegrasikan metode *feature-based clustering time series* dan metode peramalan XGBoost. Data yang digunakan dalam penelitian ini adalah harga cabai rawit merah dari masing-masing provinsi, diperoleh dari PIHPS Bank Indonesia. Analisis dan pemodelan data dilakukan menggunakan R Studio dan Python. Representasi fitur dilakukan dengan menggunakan nilai *Autocorrelation Function* (ACF), kemudian dilanjutkan dengan penggerombolan hierarki menggunakan jarak *euclidean* dan *complete linkage* sebagai metode terbaik. Hasil analisis menunjukkan bahwa data harga cabai rawit dari 31 provinsi dapat digerombolkan ke dalam 16 gerombol yang relatif paralel dan berhimpit. Peramalan kemudian dilakukan pada *prototype* masing-masing gerombol yang direpresentasikan oleh nilai rata-rata harga dari setiap gerombol. Hasil peramalan menunjukkan akurasi yang baik, dengan nilai MAPE di bawah 20%. Pendekatan ini berhasil memodelkan 16 *prototype* dari 31 provinsi.

Kata kunci: *acf, feature-based clustering time series, jarak euclidean, penggerombolan hierarki, xgboost.*

ABSTRACT

HAKIM ZOELVA MAHESA. Peramalan Harga Cabai Rawit Tingkat Gerombol Provinsi di Indonesia: Pendekatan *Feature-based Clustering Time Series* dan XGBoost. Supervised by ANANG KURNIA dan I MADE SUMERTAJAYA.

Bird's eye chili is one of Indonesia's strategic agricultural commodities, frequently experiencing significant price fluctuations that affect food inflation and national economic stability. This study aims to improve the efficiency and accuracy of forecasting by integrating feature-based clustering time series with the XGBoost forecasting method. The data used in this study consist of bird's eye chili prices from each province, obtained from PIHPS Bank Indonesia. Data analysis and modeling were conducted using R Studio and Python. Feature representation was performed using autocorrelation function (ACF) values, followed by hierarchical clustering with Euclidean distance and complete linkage as the optimal linkage method. The analysis showed that bird's eye chili price data from 31 provinces could be grouped into 16 clusters. Forecasting was then carried out on the prototypes, represented by the median price of each cluster. The forecasting results demonstrated good accuracy, with MAPE values below 20%. This approach reduced the number of required models by forecasting only 16 prototype out of 31 province.

Keywords: acf, euclidean distance, feature-based clustering time series, hierarchical clustering, xgboost.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERAMALAN HARGA CABAI RAWIT TINGKAT GEROMBOL PROVINSI DI INDONESIA: PENDEKATAN *FEATURE-BASED CLUSTERING TIME SERIES* DAN XGBOOST

HAKIM ZOELVA MAHESA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Peramalan Harga Cabai Rawit Tingkat Gerombol Provinsi di Indonesia: Pendekatan *Feature-based Clustering Time Series* dan XGBoost

Nama : Hakim Zoelva Mahesa

NIM : G1401211039

Disetujui oleh

Pembimbing 1:

Prof. Dr. Anang Kurnia S.Si., M.Si.



Pembimbing 2:

Prof. Dr. Ir. I Made Sumertajaya M.Si.



Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:

Dr. Bagus Sartono, S.Si., M.Si

NIP. 197804112005011002



Tanggal Ujian:
(17 Juli 2025)

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2025 sampai bulan Juli 2025 ini ialah “Peramalan Harga Cabai Rawit Tingkat Gerombol Provinsi di Indonesia: Pendekatan *Feature-based Clustering Time Series* dan XGBoost”. Pada proses pembuatan karya ilmiah ini, penulis tak terlepas mendapatkan dukungan, arahan, bimbingan, dan do'a dari berbagai pihak. Oleh karena itu, penulis ingin memberikan ucapan terimakasih setinggi-tingginya kepada:

1. Bapak Indra Nurhidayat, Ibu Irnasari, Nayla Aura Malika, dan Keisya Raina Hakyla selaku keluarga penulis yang telah memberikan doa'a, dukungan dan motivasi.
2. Bapak Prof. Dr. Anang Kurnia S.Si., M.Si. dan Bapak Prof. Dr. Ir. I Made Sumertajaya M.Si. selaku dosen pembimbing yang memberikan bimbingan dan arahan kepada penulis dalam menyelesaikan penulisan karya ilmiah.
3. Seluruh dosen dan staf akademik Program Studi Statistika dan Sains Data IPB University yang telah memberikan ilmu dan pelayanan yang baik selama masa perkuliahan.
4. Santi dan Trya Destiyani selaku sahabat penulis yang selalu memberikan dukungan dan motivasi.
5. Yogi Nur Hamid, Muhammad Kukuh Kurniawan, Nabil Naufal, Adisti Suci Rahmah, Jonathan Marjono, dan Dewi Kunthi Siswati Suryo selaku teman penulis yang telah menemani dan memberikan masukan selama penulisan karya ilmiah ini.
6. Alfikri Ihsan, Andy Fatihatul Fuadiya Akbar, Fedora Ilahi, dan Arfiah Kania Sektiaruni yang telah memberikan waktu, tenaga, serta kesan yang sangat berarti kepada penulis selama proses penulisan karya ilmiah ini.
7. Teman-teman dan sahabat Statistika IPB 58 terutama Windi Gunawan, Muhammad Luthfi Al Gifari, dan Akmal Basis Jatining Kusumah yang selalu kebersamai dan memberikan dukungan selama proses perkuliahan.
8. Seluruh pihak yang tidak dapat disebutkan satu persatu yang sudah memberikan dukungan dan bantuan selama proses perkuliahan.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2025

Hakim Zoelva Mahesa



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 <i>Feature-based Clustering Time Series</i>	3
2.2 Metode Penggerombolan Hierarki <i>Agglomerative</i>	3
2.3 Jarak <i>Euclidean</i>	4
2.4 Evaluasi Ukuran Ketidakmiripan	5
2.5 Ukuran Penentuan Jumlah Gerombol	5
2.6 <i>Extreme Gradient Boosting (XGBoost)</i>	6
2.7 Validasi Silang Deret Waktu	8
2.8 Evaluasi Ketepatan Model Peramalan	8
III METODE	10
3.1 Data	10
3.2 Prosedur Analisis	10
IV HASIL DAN PEMBAHASAN	12
4.1 Penanganan Data Tidak Lengkap	12
4.2 Eksplorasi Data	12
4.3 Penggerombolan Deret Waktu	14
4.4 Pemodelan Deret Waktu	17
4.5 Peramalan Harga Cabai Rawit	21
V SIMPULAN DAN SARAN	23
5.1 Simpulan	23
5.2 Saran	23
DAFTAR PUSTAKA	24
LAMPIRAN	26
RIWAYAT HIDUP	3



DAFTAR TABEL

1	Klasifikasi nilai MAPE	9
2	Jumlah provinsi yang memiliki nilai ACF yang signifikan	14
3	Hasil penggerombolan 16 gerombol	17
4	Kombinasi <i>hyperparameter</i> terbaik, evaluasi data latih, dan data uji	18
5	Perbandingan evaluasi data uji prototype dan data uji anggota gerombol	19

DAFTAR GAMBAR

1	Ilustrasi <i>expanding window</i>	8
2	Plot deret waktu harga cabai rawit setiap provinsi Di Indonesia	12
3	Plot kotak garis harga cabai rawit setiap provinsi Di Indonesia	13
4	Ekstraksi fitur ACF untuk Provinsi Sumatera Utara	14
5	Plot <i>average silhouette coefficient</i>	15
6	Dendogram hasil penggerombolan 16 Gerombol	16
7	Plot garis tujuh gerombol dengan anggota lebih dari satu provinsi	16
8	Visualisasi peta hasil penggerombolan provinsi di Indonesia	17
9	Hasil peramalan 52 periode kedepan untuk setiap <i>prototype</i>	21

DAFTAR LAMPIRAN

1	Lampiran 1 Hasil ekstraksi nilai ACF untuk setiap provinsi	27
2	Lampiran 2 Hasil perhitungan matriks jarak <i>euclidean</i>	28
3	Lampiran 3 Dendogram penggerombolan hierarki menggunakan tiga pautan	29
4	Lampiran 4 Plot deret waktu tujuh gerombol optimal	30
5	Lampiran 5 Plot deret waktu 16 gerombol optimal	32