



# **PENGGEROMBOLAN DAN PERAMALAN HARGA CABAI MERAH PROVINSI DI INDONESIA DENGAN PENDEKATAN *ENSEMBLE FEATURE-BASED CLUSTERING TIME SERIES* DAN SARIMAX**

@Hak cipta milik IPB University

**HARIYOL**



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2026**



@Hak cipta milik IPB University

IPB University



IPB University  
— Bogor Indonesia —

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.

Perpustakaan IPB University

## PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penggerombolan dan Peramalan Harga Cabai Merah Provinsi di Indonesia dengan Pendekatan *Ensemble Feature-based Clustering Time Series* dan SARIMAX” 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 2026

Hariyol  
G1401221009



## ABSTRAK

HARIYOL. Penggerombolan dan Peramalan Harga Cabai Merah Provinsi di Indonesia dengan Pendekatan *Ensemble Feature-based Clustering Time Series* dan SARIMAX. Dibimbing oleh YENNI ANGRAINI dan AGUS MOHAMAD SOLEH.

Cabai merah merupakan komoditas hortikultura strategis di Indonesia yang memiliki pergerakan harga fluktuatif dan volatilitas tinggi. Penelitian ini bertujuan menggerombolkan provinsi-provinsi di Indonesia berdasarkan karakteristik pergerakan harga cabai merah menggunakan pendekatan *feature-based clustering time series*, memodelkan data deret waktu harga cabai merah pada setiap gerombol yang terbentuk menggunakan pendekatan SARIMAX, serta meramalkan harga cabai merah pada masing-masing gerombol berdasarkan model terbaik yang diperoleh. Data yang digunakan berupa harga cabai merah mingguan tingkat provinsi yang bersumber dari Pusat Informasi Harga Pangan Strategis Nasional (PIHPS) Bank Indonesia pada periode Januari 2019 hingga Desember 2025. Prosedur penggerombolan dilakukan melalui ekstraksi fitur menggunakan *Autocorrelation Function* (ACF), penggerombolan hierarki dengan lima metode pautan, integrasi *ensemble* menggunakan *Cluster-based Similarity Partitioning Algorithm* (CSPA), dan algoritma *Walktrap*. Selanjutnya, pemodelan SARIMAX dilakukan dengan melibatkan hari besar keagamaan sebagai peubah eksogen, sedangkan evaluasi performa model menggunakan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian menunjukkan terbentuknya empat gerombol konsensus optimal. Peubah eksogen hari besar keagamaan berpengaruh signifikan pada model *prototype* Gerombol 1 dan Gerombol 4, tetapi tidak signifikan pada Gerombol 2 dan Gerombol 3. Nilai MAPE data uji berkisar antara 9,53% hingga 16,61% dengan Gerombol 4 berada pada kategori sangat baik ( $MAPE < 10\%$ ) dan Gerombol 1, 2, serta 3 berada pada kategori baik (10–20%). Selain itu, MAPE pada tingkat *prototype* secara konsisten lebih kecil dibandingkan MAPE pada tingkat anggota gerombol. Integrasi pendekatan *feature-based clustering time series* dan SARIMAX terbukti efektif dalam menangkap dinamika harga antargerombol serta meningkatkan efisiensi komputasi karena hanya memerlukan empat model peramalan.

Kata kunci : cabai merah, CSPA, *feature-based*, penggerombolan deret waktu, SARIMAX



## ABSTRACT

HARIYOL. Clustering and Forecasting of Red Chili Prices Across Provinces in Indonesia Using an Ensemble Feature-Based Clustering Time Series Approach and SARIMAX. Supervised by YENNI ANGRAINI and AGUS MOHAMAD SOLEH.

Red chili is a strategic horticultural commodity in Indonesia, characterized by highly fluctuating price movements and elevated volatility. This study aims to cluster Indonesian provinces based on red chili price movement characteristics using a *feature-based clustering time series* approach, model the time series data of red chili prices within each formed cluster using the SARIMAX approach, and forecast red chili prices for each cluster based on the best model obtained. The data used comprise weekly provincial level red chili prices from the National Strategic Food Price Information Center (PIHPS) of Bank Indonesia covering the period January 2019 to December 2025, with the clustering procedure involving feature extraction using the *Autocorrelation Function* (ACF), hierarchical clustering with five linkage methods, ensemble integration via the *Cluster-based Similarity Partitioning Algorithm* (CSPA) and the *Walktrap* algorithm, followed by SARIMAX modeling with major religious holidays as exogenous variables, and performance evaluation using the *Mean Absolute Percentage Error* (MAPE). The results indicate the formation of four optimal consensus clusters, with major religious holiday exogenous variables having a significant effect on the models for *prototypes* of Clusters 1 and 4 but not for Clusters 2 and 3, with MAPE values ranging from 9.53% to 16.61%, where Cluster 4 falls into the highly accurate category (MAPE < 10%) and Clusters 1, 2, and 3 fall into the good category (10%–20%), with *prototype*-level MAPE consistently lower than member level MAPE. The integrated *feature-based clustering time series* and SARIMAX approach proved effective in capturing inter provincial price dynamics and improving computational efficiency by requiring only four forecasting models.

**Keywords:** clustering time series, CSPA, feature-based, red chili, SARIMAX



## @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.

# **PENGGEROMBOLAN DAN PERAMALAN HARGA CABAI MERAH PROVINSI DI INDONESIA DENGAN PENDEKATAN *ENSEMBLE FEATURE-BASED CLUSTERING TIME SERIES* DAN SARIMAX**

**HARIYOL**

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  
2026**



@Hak cipta milik IPB University

IPB University

Tim Penguji pada Ujian Skripsi:  
1 Prof. Dr. Ir. Hari Wijayanto, M.Si



IPB University  
— Bogor Indonesia —

- 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 : Penggerombolan dan Peramalan Harga Cabai Merah Provinsi di  
Indonesia dengan Pendekatan *Ensemble Feature-based*  
*Clustering Time Series* dan SARIMAX

Nama : Haiyol  
NIM : G1401221009

Disetujui oleh

Pembimbing 1:

Dr. Yenni Angraini, S.Si., M.Si

---

Pembimbing 2:

Dr. Agus Mohamad Soleh, S.Si., M.T.

---

Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:

Dr. Bagus Sartono, S.Si., M.Si  
NIP. 197804112005011002

---

Tanggal Ujian:  
(23 Juni 2026)

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 Desember 2025 sampai bulan Juni 2026 ini ialah “Penggerombolan dan Peramalan Harga Cabai Merah Provinsi di Indonesia dengan Pendekatan *Ensemble Feature-based Clustering Time Series* dan SARIMAX”. 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. Ibu Dr. Yenni Angraini, S.Si., M.Si dan Bapak Dr. Agus Mohamad Soleh, S.Si., M.T selaku dosen pembimbing yang telah memberikan bimbingan, arahan, nasihat, saran, dan semangat kepada penulis selama proses penyusunan karya ilmiah ini.
2. Pusat Informasi Harga Pangan Strategis Nasional (PIHPS) yang berada di bawah Bank Indonesia, selaku penyedia data harga cabai merah yang digunakan dalam karya ilmiah ini.
3. Prof. Dr. Ir. Hari Wijayanto, M.Si selaku dosen penguji luar yang telah memberikan saran dan masukkan dalam penulisan karya ilmiah ini.
4. Keluarga besar Program Studi Statistika dan Sains Data yang telah memberikan ilmu, pengalaman, dan pembelajaran yang bermanfaat selama masa perkuliahan dan penyusunan karya ilmiah ini.
5. Ibu Kasmawati dan Yoliarman selaku keluarga penulis yang senantiasa memberikan doa, dukungan, motivasi, dan kasih sayang kepada penulis.
6. Teman-teman dan sahabat Statistika IPB 59, terutama Deden Ahmad Rabani, R. Mugni Chairil Arbi Asyari, Haidar Ramdhani, Muhammad Abdan Rofi, Delita Nur Hasanah, Sabrina Afifah Putri Utami, Shalma Kaisya Candradewi, dan Raihana Asma Amani, yang telah kebersamai, memberikan dukungan, serta berjuang bersama selama masa perkuliahan.
7. Seluruh pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan, dukungan, dan doa dalam penyelesaian karya ilmiah ini.

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

Bogor, Juli 2026

*Hariyol*



## DAFTAR ISI

|                                                                   |    |
|-------------------------------------------------------------------|----|
| DAFTAR TABEL                                                      | x  |
| DAFTAR GAMBAR                                                     | x  |
| DAFTAR LAMPIRAN                                                   | x  |
| I PENDAHULUAN                                                     | 1  |
| 1.1 Latar Belakang                                                | 1  |
| 1.2 Rumusan Masalah                                               | 2  |
| 1.3 Tujuan                                                        | 3  |
| 1.4 Manfaat                                                       | 3  |
| II TINJAUAN PUSTAKA                                               | 4  |
| 2.1 <i>Feature-based Clustering Time Series</i>                   | 4  |
| 2.2 Metode Penggerombolan Hierarki Aglomeratif                    | 4  |
| 2.3 Ukuran Ketepatan Jumlah Gerombol                              | 9  |
| 2.4 <i>Cluster-based Similarity Partitioning Algorithm (CSPA)</i> | 12 |
| 2.5 Pemodelan Data Deret Waktu                                    | 13 |
| 2.6 Evaluasi Ketepatan Model                                      | 14 |
| III METODE                                                        | 15 |
| 3.1 Data                                                          | 15 |
| 3.2 Prosedur Analisis Data                                        | 15 |
| IV HASIL DAN PEMBAHASAN                                           | 17 |
| 4.1 Praproses dan Eksplorasi Data                                 | 17 |
| 4.2 Penggerombolan Deret Waktu                                    | 19 |
| 4.3 Pemodelan Deret Waktu                                         | 24 |
| 4.4 Peramalan Harga Cabai Merah 52 Periode Mendatang              | 30 |
| V SIMPULAN DAN SARAN                                              | 34 |
| 5.1 Simpulan                                                      | 34 |
| 5.2 Saran                                                         | 34 |
| DAFTAR PUSTAKA                                                    | 36 |
| LAMPIRAN                                                          | 39 |
| RIWAYAT HIDUP                                                     | 54 |



## DAFTAR TABEL

|    |                                                                                                              |    |
|----|--------------------------------------------------------------------------------------------------------------|----|
| 1  | Kategori penggerombolan berdasarkan nilai koefisien <i>silhouette</i>                                        | 10 |
| 2  | Kategori Kebaikan MAPE                                                                                       | 14 |
| 3  | Jumlah provinsi yang memiliki nilai ACF signifikan                                                           | 19 |
| 4  | Jumlah Gerombol optimal pada setiap metode pautan                                                            | 20 |
| 5  | Matriks keanggotaan biner metode <i>average linkage</i>                                                      | 21 |
| 6  | Hasil matriks kesamaan ( <i>similarity</i> )                                                                 | 21 |
| 7  | Hasil penggerombolan 4 Gerombol                                                                              | 22 |
| 8  | Hasil pemeriksaan kestasioneran ragam dengan transformasi <i>Box-Cox</i> pada masing-masing <i>prototype</i> | 26 |
| 9  | Hasil uji ADF data <i>prototype</i> setiap Gerombol setelah <i>differencing</i>                              | 27 |
| 10 | Model terbaik setiap <i>prototype</i> berdasarkan nilai AIC terkecil                                         | 28 |
| 11 | Hasil pengujian diagnostik model dari setiap <i>prototype</i>                                                | 28 |
| 12 | Hasil evaluasi model data <i>prototype</i> setiap Gerombol                                                   | 29 |
| 13 | Perbandingan evaluasi data uji <i>prototype</i> dan data uji anggota Gerombol                                | 30 |

## DAFTAR GAMBAR

|   |                                                                                                       |    |
|---|-------------------------------------------------------------------------------------------------------|----|
| 1 | Plot deret waktu harga cabai merah mingguan setiap provinsi                                           | 17 |
| 2 | Plot kotak garis harga cabai merah mingguan setiap provinsi                                           | 18 |
| 3 | Ekstraksi fitur ACF untuk provinsi Aceh                                                               | 20 |
| 4 | Plot deret waktu harga cabai merah (a) Gerombol 1, (b) Gerombol 2, (c) Gerombol 3, dan (d) Gerombol 4 | 24 |
| 5 | Plot deret waktu harga cabai merah <i>prototype</i> Gerombol 1                                        | 25 |
| 6 | Plot data deret waktu <i>prototype</i> Gerombol 1 setelah dilakukan transformasi <i>Box-Cox</i>       | 26 |
| 7 | Plot hasil peramalan harga cabai merah (a) Gerombol 1, (b) Gerombol 2, (c) Gerombol 3, (d) Gerombol 4 | 33 |

## DAFTAR LAMPIRAN

|   |                                                                                   |    |
|---|-----------------------------------------------------------------------------------|----|
| 1 | Hasil ekstraksi fitur ACF pada setiap provinsi                                    | 39 |
| 2 | Matriks keanggotaan biner setiap pautan                                           | 41 |
| 3 | Plot deret waktu harga cabai merah setiap <i>prototype</i> Gerombol               | 45 |
| 4 | Hasil pengujian stasioneritas dalam ragam                                         | 47 |
| 5 | Plot data deret waktu setiap <i>prototype</i> setelah transformasi <i>Box-Cox</i> | 49 |
| 6 | Hasil pengujian stasioneritas dalam rata-rata                                     | 50 |
| 7 | QQ plot setiap <i>prototype</i>                                                   | 52 |