

# Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan *Clustering Time Series*

Sekarsari Utami Wijaya<sup>1\*</sup>, Ngatini<sup>2</sup>

<sup>1,2</sup>Universitas Internasional Semen Indonesia; Kompleks PT Semen Indonesia (Persero) Tbk.

Jl. Veteran, Gresik Jawa Timur 61122

<sup>1</sup>Departemen Teknik Logistik UISI Gresik Indonesia

<sup>2</sup>Departemen Informatika UISI Gresik Indonesia

e-mail: <sup>1</sup>sekarsari.wijaya@uisi.ac.id, <sup>2</sup>ngatini@uisi.ac.id

*Diajukan: 10 Oktober 2019, Diperbaiki: 15 Juli 2020, Diterima: 23 Juli 2020*

## Abstrak

Beras merupakan bahan pokok pangan bagi masyarakat Indonesia dan hampir 90% masyarakat Indonesia mengonsumsi beras. Pada awal tahun 2018 harga beras mengalami peningkatan. Kenaikan harga beras ini jika terus dibiarkan akan menyebabkan terjadinya inflasi yang berdampak pada melambatnya pertumbuhan ekonomi nasional serta dampak negatif lainnya. Dalam rangka perumusan kebijakan pengendalian inflasi maka data dan informasi terkait proyeksi keadaan pasar sangat dibutuhkan. Oleh karena itu, pemodelan harga beras di Indonesia sangat perlu dilakukan. Sebagian besar sentra beras di Indonesia berasal dari provinsi di wilayah Indonesia bagian barat, sehingga pada penelitian ini dilakukan pengembangan model harga beras untuk wilayah Indonesia bagian barat dengan menggunakan pendekatan *clustering time series*. Pemodelan dilakukan dengan tahapan pengumpulan data, pemodelan ARIMA pada level provinsi, pemodelan ARIMA pada level klaster dan evaluasi keakuratan model dengan menggunakan MAPE. Hasil penelitian ini menunjukkan model ARIMA level klaster memiliki keakuratan yang lebih tinggi daripada level provinsi.

**Kata Kunci:** *clustering time series*, harga beras, ARIMA, MAPE

## Abstract

*Rice is a staple food for Indonesian people, and around 90% of Indonesians consume rice. In early 2018, the price of rice soared. If the rising price on rice is left uncontrolled, inflation will emerge, slow national economic growth rates and create other negative impacts. In the formulation of inflation control policy, data and information related to the projected market conditions are urgently needed. Therefore, modeling the price of rice in Indonesia is very necessary. Most of the rice production centers in Indonesia are located in provinces in the Western part of Indonesia. This study used the clustering time series approach for the modeling of the price of rice in the Western part of Indonesia. The modeling steps were collecting data, modeling ARIMA at the province level, modeling ARIMA at the cluster level and evaluating model using MAPE. The results showed that the price of rice using modeling ARIMA at the cluster level is more accurate than at the individual level.*

**Keywords:** *clustering time series*, price of rice, ARIMA, MAPE

## 1 Pendahuluan

Beras merupakan bahan pokok pangan bagi sebagian masyarakat Indonesia. Hampir 90% masyarakat Indonesia mengonsumsi beras. Oleh karena itu, beras menjadi sangat penting

dalam stabilitas ekonomi dan politik nasional. Pengalaman telah membuktikan kepada kita bahwa gangguan ketahanan pangan seperti meroketnya kenaikan harga beras pada waktu krisis ekonomi 1997/1998, yang berkembang menjadi krisis multidimensi telah memicu kerawanan sosial yang membahayakan stabilitas ekonomi dan stabilitas nasional [1]. Kenaikan harga beras jika terus dibiarkan akan menyebabkan terjadinya inflasi. Jika harga beras naik Rp100/kg maka dampaknya akan langsung mempengaruhi peningkatan inflasi [2]. Jika inflasi terjadi maka dapat mempengaruhi stabilitas keamanan dan politik. Selain itu, inflasi juga akan menyebabkan pendapatan riil masyarakat menurun, nilai tabungan turun, menyulitkan pelaku usaha untuk melakukan investasi dan produksi, angka kemiskinan meningkat dan pengangguran bertambah. Dampak yang lebih jauh lagi yaitu pertumbuhan ekonomi melambat.

Barometer harga beras nasional di Indonesia hanya ada satu, yaitu Pasar Induk Beras Cipinang (PIBC). Penentuan harga beras yang hanya mengacu pada PIBC ini menjadi tidak relevan untuk wilayah lain, terutama untuk wilayah Indonesia bagian Barat. Ada dua belas provinsi yang merupakan sentra beras di Indonesia, yaitu Jawa Timur, Jawa Tengah, Jawa Barat, Sumatera Selatan, Lampung, Sumatera Utara, Sulawesi Selatan, Nusa Tenggara Barat (NTB), Aceh, Kalimantan Selatan, Kalimantan Barat dan Yogyakarta [3]. Sebagian besar sentra beras tersebut merupakan provinsi di wilayah Indonesia bagian barat. Semua provinsi yang merupakan sentra beras akan mendistribusikan produksi berasnya ke wilayah-wilayah lain untuk memenuhi kebutuhan beras, sehingga harga beras pada provinsi sentra beras akan mempengaruhi harga beras di wilayah yang menjadi tujuan distribusi.

Data dan informasi terkait proyeksi keadaan pasar jangka pendek, misalnya perkiraan kebutuhan, produksi dan proyeksi harga jangka pendek sangat diperlukan untuk perumusan kebijakan pengendalian inflasi [4]. Oleh karena itu, peramalan harga beras di wilayah Indonesia bagian barat perlu dilakukan. Banyak algoritma peramalan yang dapat digunakan untuk estimasi, seperti metode Kalman Filter [5] dan *clustering time series*. Namun dalam penelitian ini, peramalan yang digunakan adalah pendekatan *clustering time series*. Peramalan harga beras akan dilakukan pada setiap provinsi di wilayah Indonesia bagian barat. Berdasarkan pola harga beras dari waktu ke waktu, ada provinsi yang memiliki kemiripan dengan provinsi lainnya. Namun, ada juga provinsi yang tidak memiliki kemiripan dengan provinsi lainnya. Kemiripan pola ini dapat dijadikan pertimbangan dalam melakukan *clustering*.

Beberapa penelitian mengenai pemodelan harga beras di Indonesia telah dilakukan sebelumnya. Penelitian yang dilakukan oleh Ati menggunakan metode ARIMA dan ARCH untuk melakukan pemodelan harga Sembilan Bahan Pokok (Sembako), termasuk harga minyak dan beras di Jawa Tengah [6]. Penelitian serupa juga dilakukan oleh Sumaryanto pada harga beras

---

dengan tingkat agregasi nasional [7]. Salah satu penelitian yang mengimplementasikan ARIMA dan *Clustering Time Series* dilakukan oleh Adinugroho pada pengembangan model peramalan harga minyak di 32 provinsi Indonesia dengan tingkat akurasi error 7 % pada level klaster dan 2,5% pada level individu [8]. Penelitian lain mengenai *clustering time series* adalah penelitian oleh Riyadi, dkk yang menggunakan *cluster time series* dalam peramalan konsumsi listrik [9]. Penelitian tersebut menunjukkan analisis klaster yang dapat digunakan untuk pemodelan dengan menggunakan ARIMA dan ANN [9]. Selain itu, penelitian lain oleh Nazarko, dkk dalam mengestimasi konsumsi daya listrik di Bialystok, Polandia menggunakan *clustering time series* menunjukkan dugaan parameter (koefisien-koefisien) model ARIMA menghasilkan pengklasteran yang sama dengan pendekatan *fuzzy clustering* [10].

Pemodelan yang dilakukan pada penelitian ini menggunakan data pada tahun-tahun sebelumnya. Unit pengamatan yang digunakan adalah data bulanan. Data yang dikumpulkan menurut urutan waktu disebut data deret waktu (*time series*) [11]. Faktor ketidakmiripan antar provinsi menyebabkan pemodelan harga beras di wilayah Indonesia bagian barat harus dilakukan secara efisien, sehingga tidak dapat dilakukan dengan analisis *time series* biasa. Berdasarkan penjelasan sebelumnya, penelitian ini akan melakukan pemodelan ARIMA harga beras dengan level spesifik yaitu tingkat provinsi dan untuk efisiensi pengembangan model harga beras di wilayah Indonesia bagian barat tersebut maka dilakukan pengkajian pengembangan pemodelan ARIMA dengan pendekatan *clustering time series*. Pengkajian tersebut diharapkan dapat memberikan gambaran perbandingan antara pemodelan ARIMA level provinsi dengan level klaster. Simulasi tersebut dilakukan dengan bantuan program R dengan tingkat akurasi berdasarkan nilai MAPE.

## 2 Metode Penelitian

Data yang digunakan dalam penelitian ini adalah harga beras setiap provinsi di wilayah Indonesia bagian barat. Data tersebut merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Data harga beras per provinsi tersebut merupakan data deret waktu dari Januari 2010 sampai dengan Desember 2018 [12]. Tahapan penelitian dapat dilihat pada Tabel 1.

### 2.1 Analisis *Clustering Time Series*

Analisis klaster merupakan teknik menemukan klaster dalam kumpulan data dengan tujuan tertentu, data yang berada pada satu *cluster* mempunyai kemiripan yang dekat, dan memiliki perbedaan yang jelas dengan *cluster* yang lain [13]. Analisis klaster dapat diterapkan pada data deret waktu (*time series*). Selama berkembangnya *clustering* pada data deret waktu, banyak teknik

yang dikembangkan di antaranya adalah penggunaan jarak-jarak pengelompokkan yang sesuai dengan karakteristik data deret waktu [14]. Manfaat pengklasteran di antaranya untuk mengelompokkan variabel yang memiliki pola data yang hampir sama/mirip. Sehingga dalam melakukan peramalan menjadi lebih praktis dan tidak memerlukan pemodelan yang banyak. Berikut tahapan dalam analisis *clustering time series*:

#### 1. Ukuran Ketidakmiripan

Penelitian ini menggunakan ukuran ketidakmiripan *Dynamic Time Wrapping (DTW)* Ukuran ketidakmiripan ini tergolong pendekatan tanpa model tertentu. Jarak DTW adalah jarak minimum antar dua pasang titik dengan memperhitungkan pergeseran [15]. Misalkan terdapat dua deret waktu, yaitu  $R = r_1, r_2, \dots, r_i, \dots, r_n$  dan  $S = s_1, s_2, \dots, s_j, \dots, s_m$ , pertama dibuat matriks berukuran  $n \times m$ , untuk setiap elemen yang berdekatan dengan elemen  $(i, j)$  dimana  $0 < i \leq n$  dan  $0 < j \leq m$ , kita dapat mendefinisikan sebagai:

$$e_{ij} = d_{ij} + \min\{e_{(i-1)(j-1)}, e_{i-1}, e_{j-1}\} \quad (1)$$

dimana  $d_{ij} = (r_i - s_j)^2$  dan  $e_{ij}$  adalah elemen  $(i, j)$ . Untuk menemukan *optimal path*, kita harus menentukan *path* yang memberikan jarak kumulatif terkecil pada  $(n, m)$ . Jarak kumulatif terkecil pada  $(n, m)$  didefinisikan sebagai:

$$D_{DTW}(R, S) = \min \forall w \in P \left\{ \sqrt{\sum_{k=1}^K d_{w_k}} \right\} \quad (2)$$

dimana  $P$  adalah sekumpulan dari semua *wrapping path* yang mungkin,  $w_k$  adalah elemen  $(i, j)$  ke  $k$  pada *wrapping path* dan  $K$  adalah panjang dari *wrapping path* [16].

Tabel 1. Tahapan Metode Penelitian

Tahapan	Deskripsi	Output
Pengumpulan data	Data sekunder harga beras di Indonesia dalam beberapa tahun.	- Data harga beras per bulan selama 9 tahun dari BPS.
Eksplorasi data	- Membagi data menjadi gugus data latih dan gugus data uji. - Membuat grafik pola harga beras tiap provinsi di Indonesia dengan menggunakan data <i>time series</i> . - Memeriksa kestasioneran	- Gugus data latih untuk keperluan pemodelan - Gugus data uji untuk pengukuran tingkat akurasi ketepatan model - Analisis pola pergerakan harga beras dari tahun ke tahun - Data harga beras per provinsi stasioner dalam rata-rata dan ragam.
Pemodelan ARIMA pada level provinsi dengan prosedur iterative Box-Jenkins	- Mengidentifikasi model dengan ACF dan PACF - Menduga parameter model	Model matematis <i>time series</i> per provinsi

Tahapan	Deskripsi	Output
Pemodelan pada level klaster	<ul style="list-style-type: none"> <li>- Membuat ukuran ketidakmiripan DTW</li> <li>- Membuat pengklasteran dengan pautan rata-rata</li> <li>- Menentukan banyaknya klaster yang tepat</li> <li>- Menghitung rata-rata harga beras dari seluruh provinsi yang ada pada setiap klaster</li> <li>- Mengidentifikasi model dengan ACF dan PACF</li> <li>- Menduga parameter model</li> </ul>	Model matematis <i>time series</i> per klaster ( <i>clustering time series</i> )
Evaluasi MAPE hasil pemodelan ARIMA	<ul style="list-style-type: none"> <li>- Meramalkan harga beras sesuai data series</li> <li>- Membandingkan hasil peramalan harga beras dari model ARIMA dengan gugus data uji pada level provinsi dan klaster.</li> </ul>	- Model yang terbaik berdasarkan nilai MAPE

## 2. Ukuran Ketepatan Banyaknya *Cluster*

Banyaknya *cluster* yang tepat perlu diketahui agar pemodelan yang dilakukan memiliki tingkat keakuratan yang tinggi atau eror yang rendah. Salah satu ukuran ketepatan banyaknya klaster pada *time series* adalah koefisien *silhouette* dengan rumus sebagai berikut [13]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

dengan  $a(i)$  adalah rata-rata jarak antar anggota dalam klaster dan  $b(i)$  adalah jarak terkecil antar anggota klaster yang berdekatan (*nearest neighbour*). Banyaknya klaster yang tepat ditandai dengan nilai  $s(i)$  yang semakin mendekati 1.

## 3. Pemodelan *Autoregressive Integrated Moving Average* (ARIMA)

Model ARIMA merupakan salah satu model yang digunakan dalam peramalan data deret waktu. Suatu deret waktu  $\{Y_t\}$  adalah suatu model campuran *autoregressive integrated moving average* apabila pada pembedaan ke- $d$   $W_t = (1 - B)^d Y_t$  merupakan proses ARMA yang stasioner. Jadi, kita akan memperoleh

$$\phi_p(B)(1 - B)^d Y_t = \theta_0 + \theta_q(B)a_t \quad (4)$$

dengan operator AR yang stasioner  $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$  dan operator MA  $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ . Parameter  $\theta_0$  memiliki peran yang berbeda untuk  $d = 0$  dan  $d > 0$ . Ketika  $d = 0$ , prosesnya bersifat stasioner dan  $\theta_0$  berkaitan dengan rata-rata dari proses, yaitu  $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$ . Ketika  $d \geq 1$ ,  $\theta_0$  disebut tren yang deterministik dan sering tidak diikutkan ke model jika tidak benar-benar dibutuhkan. Model yang tidak stasioner dan bersifat homogen pada Persamaan (4) disebut model *autoregressive integrated moving average* dengan order  $(p, d, q)$  dan dinotasikan dengan ARIMA  $(p, d, q)$ . Ketika  $p = 0$ , model ARIMA  $(p, d, q)$  juga disebut model *integrated moving average* dengan order  $(d, q)$  dan dinotasikan sebagai model IMA  $(d, q)$  [11].

## 2.2 Ukuran Akurasi Ketepatan Model

Ukuran ketepatan yang sering digunakan untuk mengetahui ketepatan suatu metode peramalan dalam memodelkan data *time series*, yaitu MAPE (*Mean Absolute Percentage Error*), MSD (*Mean Squared Deviation*), dan MAD (*Mean Absolute Deviaton*). Ukuran ketepatan MSD dan MAD memiliki kelemahan. Kelemahan dari dua ukuran ini adalah tidak memudahkan perbandingan antar deret berskala yang berbeda dan untuk selang waktu yang berlainan karena MSD dan MAD merupakan ukuran absolut yang sangat tergantung pada skala dari data deret waktu. Selain itu, interpretasi nilai MSD tidak bersifat intuitif, karena ukuran ini menyangkut pengkuadratan sederetan nilai [17]. Oleh karena itu, ukuran ketepatan peramalan yang akan digunakan pada penelitian ini adalah MAPE. Pengukuran akurasi ini dipilih karena alat ukur ini terbebas dari skala yang digunakan pada amatan asli [8]. Definisi MAPE adalah [18]:

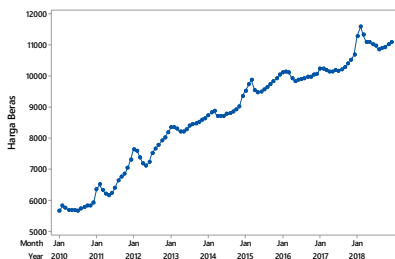
$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| 100\% \quad (5)$$

dengan  $y_t$  = harga beras pada periode ke- $t$  dan  $\hat{y}_t$  = ramalan harga beras pada periode ke- $t$ .

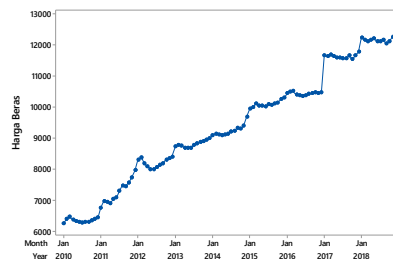
## 3 Hasil dan Pembahasan

### 3.1 Eksplorasi Data

Eksplorasi data yang dilakukan dalam penelitian ini merupakan penerapan dari statistika deskriptif. Statistika deskriptif hanya memberikan informasi mengenai data yang dimiliki [19]. Berdasarkan plot *time series*, pola harga beras seluruh provinsi di wilayah Indonesia bagian barat dapat dikelompokkan menjadi dua. Gambar 1 merupakan plot harga beras pada dua belas provinsi. Secara eksploratif, harga beras mengalami fluktuasi. Namun, fluktuasi yang terjadi tidak terlalu signifikan. Harga beras di wilayah ini cenderung selalu meningkat secara perlahan-lahan dari tahun ke tahun. Dua belas provinsi tersebut meliputi Provinsi Aceh, Riau, Sumatera Utara, Jambi, Bengkulu, Lampung, DKI Jakarta, Jawa Barat, Jawa Tengah, Daerah Istimewa Yogyakarta, Jawa Timur dan Banten. Harga beras mencapai nilai tertinggi pada Februari 2018.



**Gambar 1.** Pola Harga Beras 12 Provinsi

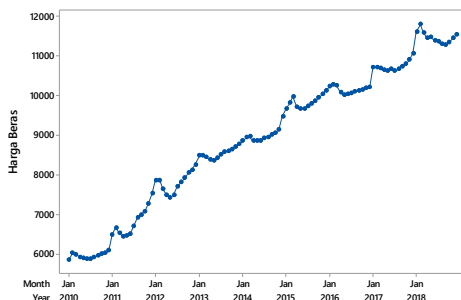


**Gambar 2.** Pola Harga Beras 6 Provinsi

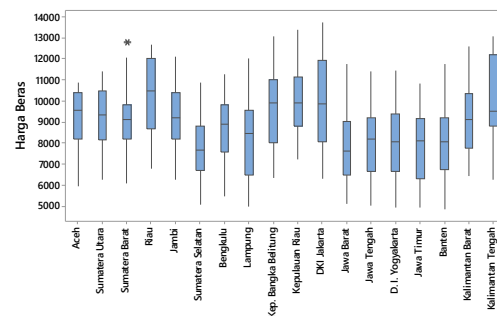
Gambar 2 memiliki pola yang cenderung hampir sama dengan Gambar 1, yaitu harga beras mengalami fluktuatif yang tidak terlalu signifikan dan cenderung naik secara perlahan-lahan. Namun, ada perbedaan yaitu pola harga beras pada enam provinsi ini mengalami peningkatan yang cukup signifikan pada Januari 2017, yaitu dari harga Rp9.870 menjadi Rp10.949. Enam provinsi ini, meliputi Provinsi Sumatera Barat, Sumatera Selatan, Bangka Belitung, Kepulauan Riau, Kalimantan Barat dan Kalimantan Tengah. Secara umum, pola harga beras seluruh provinsi di wilayah Indonesia bagian barat dari awal tahun 2010 sampai akhir tahun 2018 cenderung mengalami peningkatan. Meskipun pada bulan-bulan tertentu harga beras sempat mengalami penurunan.

Berdasarkan Gambar 3, harga beras berada di titik terendah pada bulan Mei di hampir setiap tahunnya. Sedangkan harga beras berada di titik tertinggi pada bulan Maret di hampir setiap tahunnya. Persediaan beras pada Februari dan awal Maret semakin menipis sehingga menyebabkan harga beras meningkat pada bulan Maret. Hal tersebut disebabkan oleh adanya periode puncak panen raya padi yang terjadi pada Maret dan April. Setelah panen raya terjadi, persediaan beras semakin banyak sehingga menyebabkan harga beras mulai menurun pada Mei.

Secara keseluruhan data dari 2010 hingga 2018, harga beras terendah terletak di provinsi Jawa Barat, yaitu sebesar Rp7.623. Sedangkan harga beras tertinggi terletak di provinsi Riau yaitu sebesar 10.485. Berdasarkan diagram kotak pada Gambar 4, jarak antara  $Q_1$  dan  $Q_3$  di provinsi DKI Jakarta sangat jauh. Hal tersebut menunjukkan ragam/varians di provinsi tersebut sangat tinggi atau data harga beras sangat heterogen/bervariasi. Sedangkan jarak antara  $Q_1$  dan  $Q_3$  di Provinsi Sumatera Barat sangat dekat. Hal tersebut menunjukkan ragam di provinsi tersebut kecil atau data harga beras cenderung homogen/tidak bervariasi. Provinsi yang memiliki pencilan/*outlier* atas, yaitu Provinsi Sumatera Barat.



**Gambar 3.** Pola Harga Beras di Indonesia



**Gambar 4.** Diagram Kotak Harga Beras

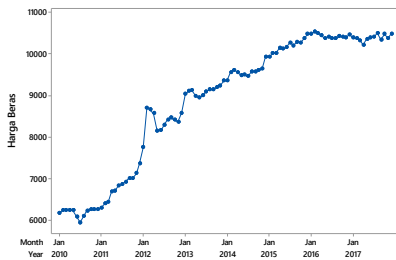
### 3.2 Pemodelan ARIMA pada Level Provinsi

Sebelum pemodelan harga beras setiap provinsi dilakukan, data harga beras dibagi menjadi dua gugus terlebih dahulu. Data tersebut dibagi menjadi gugus data latih dan gugus data uji. Gugus

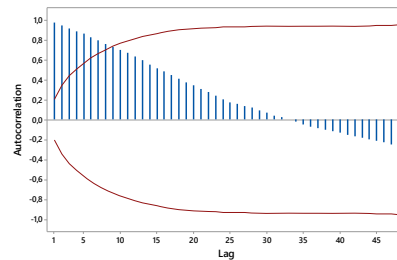
data latih digunakan untuk membuat pemodelan dan pengklasteran, sedangkan gugus data uji digunakan untuk mengukur tingkat akurasi model yang dibuat. Pola harga beras 2010-2013 cenderung tidak stabil. Oleh karena itu, data pada tahun tersebut tidak diikutsertakan dalam melakukan pemodelan. Data harga beras pada 2014-2017 dijadikan gugus data uji, sedangkan data beras pada 2018 dijadikan gugus data latih. Setelah data latih ditentukan, pemeriksaan kestasioneran dari data deret waktu perlu dilakukan. Pemeriksaan ketidastasioneran dapat dilakukan dengan membuat plot deret waktu.

Pembuatan model pertama kali dilakukan pada Provinsi Aceh. Berdasarkan plot data harga beras (Gambar 5), harga beras di Provinsi Aceh berfluktuasi dan cenderung menunjukkan adanya peningkatan. Fluktuasi ini ditandai oleh adanya perubahan harga beras secara bertahap pada setiap tahunnya baik naik maupun turun, sehingga data harga beras di provinsi tersebut tidak stasioner.

Selain plot deret waktu, kestasioneran dapat diperiksa dengan menggunakan plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*). Berdasarkan plot ACF pada Gambar 6, Provinsi Aceh memiliki nilai otokorelasi yang menurun lambat secara eksponensial. Pola ACF tersebut menunjukkan bahwa data harga di Provinsi Aceh tidak stasioner. Pola yang menurun lambat secara eksponensial pada plot ACF menyebabkan plot PACF tidak perlu dibuat. Plot ACF setiap provinsi dengan data harga beras semua provinsi di wilayah Indonesia bagian barat belum stasioner.



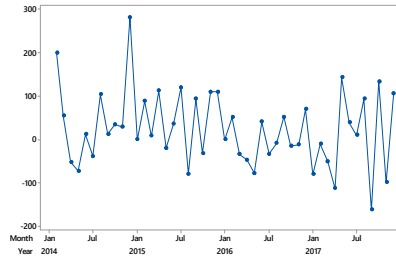
**Gambar 5.** Plot Harga Beras Provinsi Aceh



**Gambar 6.** Plot ACF Provinsi Aceh

Berdasarkan plot deret waktu pada Gambar 5, data harga beras di provinsi Aceh tidak stasioner dalam rata-rata. Hal tersebut ditandai dengan adanya peningkatan harga secara perlahan. Namun, data tersebut sudah stasioner dalam ragam karena lebar terompeta tidak terlalu mengalami perubahan. Ketidastasioneran dalam rata-rata dapat diatasi dengan melakukan pembedaan (*differencing*). Plot pembedaan pertama harga beras di Provinsi Aceh dapat dilihat pada Gambar 7. Gambar tersebut menunjukkan data harga beras sudah stasioner dalam rata-rata karena tidak menunjukkan adanya pola meningkat/menurun. Selain itu, data tersebut juga sudah stasioner dalam ragam karena lebar terompeta relatif sama.

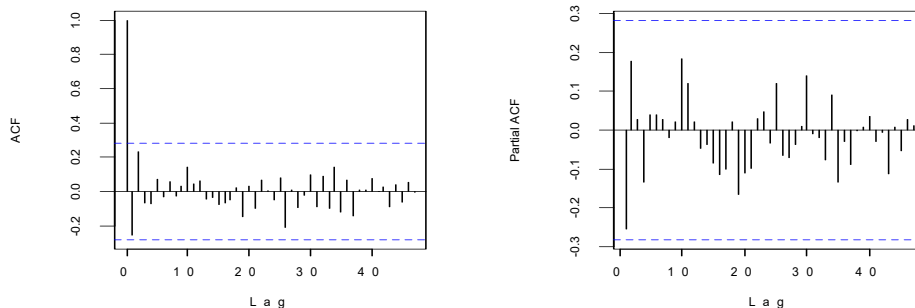




**Gambar 7.** Plot Perbedaan Pertama Harga Beras di Provinsi Aceh

Pemeriksaan kestasioneran juga dapat dilakukan secara inferensia, yaitu uji Augmented Dickey-Fuller (ADF) dan uji Kwiatkowski Phillips Schmidt Shin (KPSS). Taraf nyata yang digunakan sebesar 5%. Nilai-p dari uji ADF yang diperoleh sebesar 0,04. Nilai-p yang dihasilkan lebih kecil daripada taraf nyata sehingga kesimpulan yang diperoleh adalah perbedaan pertama data harga beras di Provinsi Aceh mengandung *unit root* atau sudah stasioner. Nilai-p dari uji KPSS sebesar 0,1. Nilai-p yang dihasilkan lebih besar dari taraf nyata, sehingga perbedaan pertama data harga beras di Aceh sudah stasioner. Uji ADF dan KPSS untuk setiap provinsi menunjukkan bahwa semua provinsi sudah stasioner dalam rata-rata dan ragam. Kesimpulan dari uji ADF dan KPSS sudah sesuai dengan plot deret waktu perbedaan pertama. Oleh karena data perbedaan pertama sudah stasioner dalam rata-rata dan ragam, langkah selanjutnya adalah membuat model.

Pemodelan ARIMA menggunakan prosedur iterasi Box-Jenkins. Tahapan prosedur Box-Jenkins meliputi identifikasi model, pendugaan parameter dan evaluasi model. Tahap identifikasi model menggunakan plot ACF dan PACF perbedaan pertama data harga beras Provinsi Aceh. Berdasarkan Gambar 8, otokorelasi pada lag musiman tidak terlihat sehingga kandidat model yang teridentifikasi IMA (1,1), ARI (1,1) dan ARIMA (1,1,1). Sedangkan model *overfitting* yang dipilih adalah IMA (1,2), ARI (2,1), ARIMA (2,1,1), ARIMA (1,1,2) dan ARIMA (2,1,2).



**Gambar 8.** Plot ACF dan PACF Perbedaan Pertama Harga Beras Provinsi Aceh

Berdasarkan uji statistik Ljung Box, nilai-p dari semua sisaan kandidat model ARIMA pada lag 12, 24, dan 36 lebih dari 5%. Hal tersebut menunjukkan bahwa sisaan bersifat *white noise*. Langkah selanjutnya adalah mencari nilai MAPE terkecil untuk menentukan model ARIMA yang

tepat. Semakin kecil nilai MAPE maka semakin akurat model yang dibuat. Berdasarkan Tabel 2, nilai MAPE antar kandidat model tidak terlalu berbeda jauh. Namun, model ARIMA (0,1,2) memiliki nilai MAPE terkecil, yaitu sebesar 2,32%.

Tabel 2. Pendugaan Model dan Pemeriksaan Sisaan ARIMA di Provinsi Aceh

No.	ARIMA			MAPE (%)	Ljung Box Statistics (Nilai-p)		
	p	q	D		12	24	36
1	0	1	1	2,94	0,733	0,942	0,771
2	1	1	1	3,01	0,769	0,961	0,921
3	1	1	0	2,75	0,726	0,933	0,830
4	0	1	2	2,32	0,887	0,993	0,970
5	2	1	0	2,35	0,880	0,990	0,976
6	2	1	1	2,83	0,820	0,985	0,968
7	2	1	2	3,22	0,845	0,988	0,973

Model ARIMA untuk provinsi Aceh dengan sisaan yang *white noise* dan MAPE terkecil adalah ARIMA (0,1,2). Setelah identifikasi model dilakukan, parameter model ARIMA perlu diduga. Model tersebut dapat dinotasikan  $W_t^{Aceh} = -0,095e_{t-1} + 0,391e_{t-2}$  dimana  $W_t^{Aceh} = Y_t^{Aceh} - Y_{t-1}^{Aceh}$ , sehingga model IMA (1,2) Provinsi Aceh dapat dinotasikan sebagai berikut:

$$Y_t^{Aceh} = Y_{t-1}^{Aceh} + e_t - 0,095e_{t-1} + 0,391e_{t-2} \quad (6)$$

dengan  $Y_t$  adalah harga beras waktu ke- $t$  dan  $t$  adalah indeks bulan. Pemodelan ARIMA setiap provinsi di wilayah Indonesia bagian barat dapat dilihat pada Tabel 3. Ukuran akurasi ketepatan model dapat dilihat dari nilai MAPE. Semakin rendah nilai MAPE menunjukkan model semakin akurat.

Tabel 3. Model ARIMA Setiap Provinsi di Wilayah Indonesia Bagian Barat

No.	Provinsi	Model ARIMA	MAPE
1.	Aceh	$Y_t^{Aceh} = Y_{t-1}^{Aceh} + e_t - 0,095e_{t-1} + 0,391e_{t-2}$	2,32%
2.	Sumatera Utara	$Y_t^{Sumut} = Y_{t-1}^{Sumut} + e_t + 0,566e_{t-1}$	2,93%
3.	Sumatera Barat	$Y_t^{Sumbar} = Y_{t-1}^{Sumbar} + e_t + 0,238e_{t-1}$	4,47%
4.	Riau	$Y_t^{Riau} = 0,191Y_{t-1}^{Riau} + 0,809Y_{t-2}^{Riau} + e_t - 0,078e_{t-1} - 0,875e_{t-2}$	1,41%
5.	Jambi	$Y_t^{Jambi} = 2,178Y_{t-1}^{Jambi} - 1,356Y_{t-2}^{Jambi} + e_t + 0,178e_{t-3} - 0,730e_{t-2} - 0,338e_{t-3}$	4,77%
6.	Sumatera Selatan	$Y_t^{Sumsel} = 1,185Y_{t-1}^{Sumsel} - 0,185Y_{t-2}^{Sumsel} + e_t$	3,32%
7.	Bengkulu	$Y_t^{Bengkulu} = 1,226Y_{t-1}^{Bengkulu} - 0,226Y_{t-2}^{Bengkulu} + e_t$	2,38%
8.	Lampung	$Y_t^{Lampung} = Y_{t-1}^{Lampung} + e_t - 0,372e_{t-1}$	6,54%
9.	Bangka Belitung	$Y_t^{Babel} = Y_{t-1}^{Babel} + e_t - 0,03e_{t-1}$	5,53%
10.	Kepulauan Riau	$Y_t^{Kepri} = 0,801Y_{t-1}^{Kepri} + 0,199Y_{t-2}^{Kepri} + e_t$	4,09%
11.	DKI Jakarta	$Y_t^{DKI} = 0,18Y_{t-1}^{DKI} + 0,82Y_{t-2}^{DKI} + e_t + 0,834e_{t-1} - 0,125e_{t-2}$	3,24%
12.	Jawa Barat	$Y_t^{Jabar} = Y_{t-1}^{Jabar} + e_t + 0,248e_{t-1}$	4,78%
13.	Jawa Tengah	$Y_t^{Jateng} = 1,998Y_{t-1}^{Jateng} - 0,998Y_{t-2}^{Jateng} + e_t - 0,591e_{t-1} - 0,375e_{t-2}$	2,44%
14.	DI Yogyakarta	$Y_t^{DIY} = 1,365Y_{t-1}^{DIY} - 0,365Y_{t-2}^{DIY} + e_t$	3,33%
15.	Jawa Timur	$Y_t^{Jatim} = Y_{t-1}^{Jatim} + e_t + 0,489e_{t-1}$	2,98%
16.	Banten	$Y_t^{Banten} = Y_{t-1}^{Banten} + e_t - 0,067e_{t-1}$	5,80%

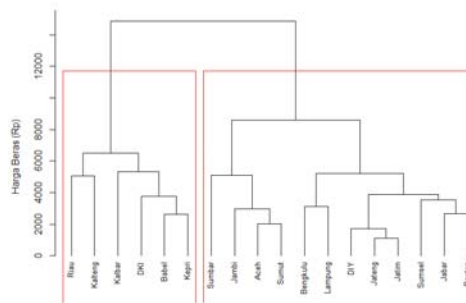
No.	Provinsi	Model ARIMA	MAPE
17.	Kalimantan Barat	$Y_t^{Kalbar} = 1,9998Y_{t-1}^{Kalbar} - 0,9998Y_{t-2}^{Kalbar} + e_t - 0,989e_{t-1}$	1,85%
18.	Kalimantan Tengah	$Y_t^{Kalteng} = 1,067Y_{t-1}^{Kalteng} - 0,67Y_{t-2}^{Kalteng} + e_t$	1,23%

### 3.3 Clustering Time Series

Data harga beras 2010-2018 pada masing-masing provinsi di wilayah Indonesia bagian barat memiliki pola deret waktu yang berbeda-beda. Berdasarkan plot deret waktu yang telah dijelaskan sebelumnya, ada dua pola yang berbeda sehingga terbentuk dua kluster yang dilakukan secara manual. Dua kluster tersebut dapat dilihat pada Gambar 1 dan Gambar 2. Pengklasteran yang telah dilakukan pada penjelasan sebelumnya memerlukan waktu yang lama dan bersifat subjektif karena objek yang diamati berjumlah besar. Oleh karena itu, perlu adanya suatu metode yang digunakan untuk melakukan pengklasteran. Teknik pengklasteran yang digunakan dalam penelitian ini adalah pengklasteran berhierarki. Metode pengklasteran berhierarki yang digunakan adalah pautan rata-rata.

#### Diagram Pohon (Dendrogram)

Tahapan pertama dalam melakukan analisis *clustering time series* adalah menentukan matriks jarak untuk mengetahui ukuran ketidakmiripan. Ukuran ketidakmiripan yang digunakan untuk data deret waktu adalah DTW. Setelah matriks jarak DTW terbentuk, diagram pohon dibuat untuk mengklasterkan provinsi-provinsi yang memiliki pola harga beras yang mirip. Penentuan banyaknya kluster berdasarkan nilai koefisien *silhouette*. Berdasarkan nilai koefisien *silhouette*, pembentukan dua kluster merupakan pengklasteran terkuat karena memiliki nilai koefisien di atas 0,7. Gambar 9 menunjukkan nama-nama provinsi yang termasuk ke dalam kluster satu dan kluster dua. Kluster satu terdiri dari Provinsi Riau, Kalimantan Tengah, Kalimantan Barat, DKI Jakarta, Bangka Belitung dan Kepulauan Riau, sedangkan kluster kedua terdiri dari Provinsi Sumatera Barat, Jambi, Aceh, Sumatera Utara, Bengkulu, Lampung, DI Yogyakarta, Jawa Tengah, Jawa Timur, Sumatera Selatan, Jawa Barat, dan Banten.



**Gambar 9.** Diagram Pohon dengan DTW

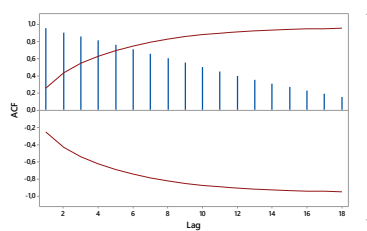
Setelah banyaknya kluster dan anggota kluster ditentukan, rata-rata harga beras dihitung dari seluruh anggota kluster. Nilai rata-rata ini digunakan untuk pembuatan model ARIMA.

Berdasarkan Gambar 10, harga beras pada masing-masing kluster semakin lama semakin meningkat. Harga beras pada kluster 1 dan Kluster 2 berfluktuatif. Pola tersebut mengidentifikasi adanya ketidakstasioneran.

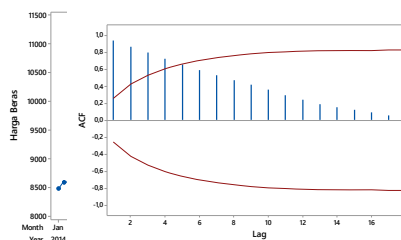
Untuk lebih meyakinkan bahwa pola data tidak stasioner, plot ACF kedua kluster tersebut perlu dibuat. Plot ACF pada Gambar 11 menunjukkan nilai otokorelasi menurun secara perlahan. Hal tersebut menunjukkan harga beras pada seluruh kluster tidak stasioner. Oleh karena itu, perlu dilakukan pembedaan pertama untuk mengatasi ketidakstasioneran.

### Identifikasi Model

Setelah dilakukan pembedaan pertama, plot data harga beras pada kluster 1 dan 2 cenderung stasioner dalam rata-rata karena titik tengah grafik berada pada nilai yang stabil. Meskipun pada

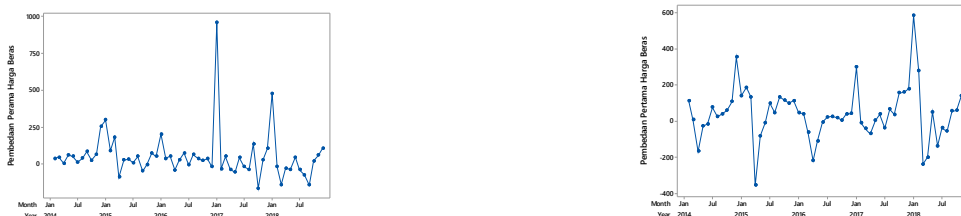


**Gambar 10.** Plot Deret Waktu Harga Beras Kluster 1 (kiri) dan Kluster 2 (kanan)



**Gambar 11.** Plot ACF Kluster 1 (kiri) dan Kluster 2 (kanan)

Gambar 12 ada terompet yang memiliki lebar yang berbeda, secara keseluruhan lebar plot tersebut sebagian besar memiliki lebar yang sama. Oleh karena itu, plot deret waktu setelah dilakukan pembedaan pertama dapat dikatakan stasioner dalam rata-rata dan ragam. Nilai ADF dan KPSS untuk masing-masing kluster dapat dilihat pada Tabel 4. Nilai-p uji ADF pada kluster 1 dan 2 memberikan kesimpulan bahwa pembedaan pertama harga beras pada masing-masing kluster tidak mengandung *unit root* atau data stasioner. Nilai-p uji KPSS pada kluster 1 dan 2 juga memberikan kesimpulan bahwa kedua kluster tersebut sudah stasioner, baik dalam rata-rata maupun ragam.

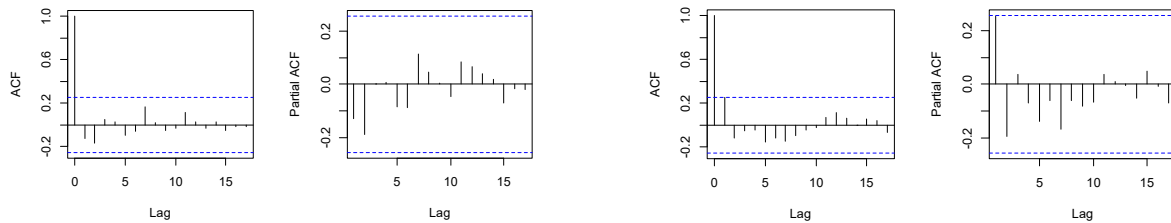


**Gambar 12.** Plot Deret Waktu Pembedaan Pertama Kluster 1 (kiri) dan Kluster 2 (kanan)

Tabel 4. Uji ADF dan KPSS Pembedaan Pertama Klaster 1 dan Klaster 2

Klaster	Uji ADF		Uji KPSS	
	Level	Nilai-p	Level	Nilai-p
1	-3,904	0,020	0,164	0,100
2	-3,859	0,022	0,053	0,100

Plot ACF dan PACF pada Gambar 13 digunakan untuk mengidentifikasi model ARIMA untuk masing-masing klaster. Plot ACF pada kedua klaster tersebut *cut off* (terpotong) setelah lag ke-1 dan plot PACF menurun mengikuti pola sinus, sehingga kandidat model ARIMA yang teridentifikasi untuk klaster 1 dan 2 adalah ARIMA (0,1,1). Model tersebut akan di-*overfitting* menjadi ARIMA (1,1,1), ARIMA (0,1,2), ARIMA (1,1,2), ARIMA (0,2,1), ARIMA (1,2,1), ARIMA (0,2,2) dan ARIMA (1,2,2).



**Gambar 13.** Plot ACF dan PACF Pembedaan Pertama Klaster 1 (kiri) dan 2 (kanan)

### Pendugaan Parameter dan Peramalan

Setelah mengidentifikasi kandidat model ARIMA berdasarkan plot ACF dan PACF, langkah selanjutnya adalah mengestimasi parameter kandidat model ARIMA. Estimasi parameter ini digunakan untuk melakukan peramalan harga beras pada tahun berikutnya sesuai dengan tahun yang ada di gugus uji. Selanjutnya, peramalan harga beras ini digunakan untuk menentukan nilai MAPE. Nilai MAPE ini akan dijadikan pertimbangan dalam pemilihan model ARIMA terbaik. Semakin kecil nilai MAPE, maka semakin baik model tersebut. Model ARIMA yang memiliki MAPE paling kecil untuk klaster 1 adalah ARIMA (0,2,1), sedangkan klaster 2 adalah ARIMA (0,2,2). Model ARIMA dapat ditulis secara matematis pada Tabel 5.

Tabel 5. Model ARIMA Klaster 1 dan 2

Klaster	Model ARIMA	MAPE
1	$Y_t^{G1} = 2Y_{t-1}^{G1} - Y_{t-2}^{G1} + e_t - e_{t-1}$	1,97%
2	$Y_t^{G2} = 2Y_{t-1}^{G2} - Y_{t-2}^{G2} + e_t - 0,534e_{t-1} - 0,461e_{t-2}$	2,12%

### 3.4 Perbandingan MAPE Model ARIMA Level Provinsi dan Klaster

Provinsi Riau, Kalimantan Tengah, Kalimantan Barat, DKI Jakarta, Bangka Belitung dan Kepulauan Riau memiliki pola harga beras yang mirip, sehingga provinsi tersebut berada pada klaster yang sama yaitu klaster satu. Sedangkan Provinsi Sumatera Barat, Jambi, Aceh, Sumatera Utara, Bengkulu, Lampung, DI Yogyakarta, Jawa Tengah, Jawa Timur, Sumatera Selatan, Jawa Barat, dan Banten memiliki pola harga beras yang mirip, sehingga provinsi tersebut berada pada klaster yang sama yaitu klaster dua. Namun, klaster satu memiliki pola data yang berbeda dengan klaster dua. Tingkat ketepatan model dari masing-masing klaster dapat dilihat dari nilai MAPE pada Tabel 6.

Nilai MAPE tanpa klaster diperoleh dengan cara mencari nilai rata-rata MAPE pada setiap provinsi di wilayah Indonesia bagian barat. Berdasarkan Tabel 6, model ARIMA dengan pengklasteran memiliki MAPE yang lebih rendah daripada model ARIMA setiap provinsi. Skala penilaian keakurasian ramalan meliputi MAPE kurang dari 10% dikategorikan sangat kuat, MAPE antara 11%-20% adalah ramalan bagus, MAPE antara 21%-50% adalah ramalan yang masuk akal, dan MAPE lebih dari 51% adalah ramalan tidak akurat [20]. Oleh karena itu, pemodelan pada level klaster dapat dikategorikan lebih akurat dibandingkan pemodelan level provinsi. Sehingga, peramalan harga beras untuk tahun berikutnya dapat dilakukan lebih efisien dengan menggunakan model ARIMA *clustering time series*.

Tabel 6. Perbandingan MAPE Model ARIMA Setiap Provinsi dan Klaster

Jenis	Klaster	MAPE
Tanpa Klaster	Tanpa Klaster	3,52%
Dengan Klaster	Klaster 1	1,97%
	Klaster 2	2,12%

## 4 Simpulan

Simpulan yang diperoleh dari penelitian yang telah dilakukan antara lain:

1. Model ARIMA setiap provinsi di wilayah Indonesia bagian barat sebanyak 18 persamaan dengan nilai MAPE terkecil sebesar 1,23% dan MAPE terbesar sebesar 6,54%. Hal tersebut menunjukkan bahwa model yang diperoleh memiliki tingkat akurasi yang tinggi.
2. Pengklasteran yang terbentuk sebanyak dua klaster. Klaster 1 terdiri atas 6 provinsi, sedangkan klaster 2 terdiri dari 12 provinsi. Berdasarkan nilai MAPE, model ARIMA kedua klaster tersebut memiliki tingkat keakurasian yang tinggi.
3. Berdasarkan nilai MAPE, model ARIMA pada level klaster memiliki keakuratan yang lebih tinggi daripada level provinsi. Meskipun nilai MAPE pada setiap provinsi ada yang lebih

rendah dari MAPE peramalan dengan pengklasteran, pemodelan ARIMA dengan pengklasteran tetap lebih efisien.

## 5 Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia atas kesempatan dan pendanaan dalam melakukan penelitian ini. Serta penulis juga mengucapkan terima kasih kepada Universitas Internasional Semen Indonesia atas dukungannya selama ini.

## 6 Daftar Pustaka

- [1] M. Ismet, “Tantangan Mewujudkan Kebijakan Pangan Nasional yang Kuat,” *J. Pangan*, vol. 16, no. 1, 2007, doi: <https://doi.org/10.33964/jp.v16i1.271>.
- [2] T. D. Komalasari, “Inflasi, Makna dan Pengaruhnya untuk Kita,” 2018. <https://www.pikiran-rakyat.com/ekonomi/pr-01293148/inflasi-makna-dan-pengaruhnya-untuk-kita-419017> (accessed Jul. 21, 2018).
- [3] H. A. Al Hikam, “12 Provinsi akan Jadi Tumpuan Produksi Beras RI,” 2019. <https://finance.detik.com/berita-ekonomi-bisnis/d-4428624/12-provinsi-akan-jadi-tumpuan-produksi-beras-ri> (accessed Jul. 21, 2018).
- [4] Kementerian Perdagangan, *Kajian Pengembangan Model Proyeksi Jangka Pendek Pasar Komoditas Pangan Pokok*. Jakarta: Kementerian Perdagangan, 2009.
- [5] Ngatini, E. Apriliani, and H. Nurhadi, “Ensemble and Fuzzy Kalman Filter for position estimation of an autonomous underwater vehicle based on dynamical system of AUV motion,” *Expert Syst. Appl.*, vol. 68, pp. 29–35, Feb. 2017, doi: 10.1016/j.eswa.2016.10.003.
- [6] E. N. L. Ati, “Analisis Volatility Forecasting 9 Bahan Pokok Menggunakan Metode GARCH dengan Program R,” Universitas Negeri Semarang, 2015.
- [7] Sumaryanto, “Analisis Volatilitas Harga Eceran Beberapa Komoditas Pangan Utama Dengan Model Arch/Garch,” *J. Agro Ekon.*, vol. 27, no. 2, pp. 135–163, 2009.
- [8] W. Adinugroho, “Pengembangan Model Peramalan Harga Minyak Goreng Dengan Pendekatan Clustering Time Series Widyayanto Adinugroho,” IPB, 2018.
- [9] D. Riyadi, M. A. A., Fithriasari, K., “Data Mining Peramalan Konsumsi Listrik dengan Pendekatan Clustering Time Series sebagai Preprocessing,” *J. Sains dan Seni ITS*, vol. 5, no. 1, pp. D121–D126, 2016.
- [10] Z. W. Nazarko J, Jurczuk A, “ARIMA Models in Load Modelling with Clustering Approach,” 2005, doi: 10.1109/PTC.2005.4524719.
- [11] W. Wei, *Time Series Analysis: Univariate and Multivariate Methods*. Second penyunt. Temple (US): Pearson Addison Wesley., 2006.
- [12] Badan Pusat Statistik (BPS), “Statistik Harga Konsumen Perdesaan Kelompok Makanan Tahun 2010-2018,” 2019. <https://www.bps.go.id/publication/2019/05/08/0815f41af6d51e3a59ba7957/statistik-harga-konsumen-perdesaan-kelompok-makanan-2010-2018.html>.

- [13] P. J. Kaufman, L. & Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey (US): John Wiley and Sons Inc., 1990.
- [14] W. T. Liao, "Clustering of time series data - A survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, 2005, doi: 10.1016/j.patcog.2005.01.025.
- [15] D. J. Berndt and J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," in *Knowledge Discovery in Databases Workshop*, 1994, pp. 359–370.
- [16] C. A. Niennattrakul, V. & Ratanamahatana, "On Clustering Multimedia Time Series Data Using K-Means and Dynamic Time Warping," 2007.
- [17] R. T. Sungkawa, I. & Megasari, "Penerapan Ukuran Ketepatan Nilai Ramalan Data Deret Waktu dalam Seleksi Model Peramalan Volume Penjualan PT Satriamandiri Citramulia," *ComTech Comput. Math. Eng. Appl.*, vol. 2, no. 2, pp. 636–645, 2011.
- [18] M. Montgomery, D. C., Jennings, C. L. & Kulahci, *Introduction to Time Series Analysis and Forecasting*. Third penyunt. New Jersey (US): John Wiley & Sons Inc., 2008.
- [19] S. U. Wijaya, "Pendugaan Parameter dan Pengujian Hipotesis pada Bivariate Poisson Inverse Gaussian Regression," ITS, 2017.
- [20] C. D. Lewis, *Industrial and Business Forecasting methods: A Radical Guide to Exponential Smoothing and Curve Fitting*. Boston (US): Butterworth Scientific., 1982.