

---

## Analisis Dampak Covid-19 Terhadap Indeks Harga Konsumen dengan *K-Means* dan Regresi Berganda

Firli Azizah<sup>1</sup> dan Muhammad Athoillah<sup>2</sup>

<sup>1,2</sup> Prodi Statistika, Fakultas Sains dan Teknologi, Universitas PGRI Adi Buana Surabaya

<sup>1</sup> firliazizah203@gmail.com, <sup>2</sup> athoillah.muhammad@gmail.com

**Abstract.** The Indonesian economy during the global pandemic entered the brink of economic recession. This problem occurs because the state of public consumption has decreased due to the limited space for community movement and sluggish economic activities due to preventing the transmission of Covid-19. This affects the decline in public consumption in economic activities. In this case, it can be seen from the statistical news published by the official website of the Badan Pusat Statistik (BPS) which reports that the inflation rate in the previous months was around 0.10%, while in April 2020 it decreased by 0.08%. Based on these, a *K-means* grouping study was conducted by dividing the cluster into 3 parts and modeling using multiple regression methods. In this study, the variable used was the price index. The results of the *K-means* cluster analysis with the division of 3 clusters, namely cluster 3 (high CPI cluster) consisting of 66 cities, cluster 1 (moderate CPI cluster) consisting of 2 cities, and cluster 2 (low CPI cluster) consisting of 22 cities. Furthermore, the multiple regression results obtained 12 variables that have a significant effect on the Consumer Price Index (CPI). The results of regression modeling are the highest coefficient is food at 0.236 and the lowest coefficients are cigarettes and tobacco at 0.008. Therefore can be concluded that the grouping of the CPI indicator obtained 75% of cities with high index prices, especially in big cities such that economic activity, in general, was still consumptive during the pandemic and multiple regression modeling resulted from 37 indicator variables, only 12 indicator variables had a significant effect on the CPI.

**Keywords:** *k-means*, CPI, multiple regression, and price index

### 1. Pendahuluan

Perekonomian Indonesia saat pandemi secara global memasuki jurang resesi ekonomi. Permasalahan ini disebabkan konsumsi masyarakat menurun karena ruang gerak masyarakat yang dibatasi dan lesunya kegiatan ekonomi karena mencegah penularan Covid-19. Hal itu mempengaruhi penurunan konsumsi masyarakat pada kegiatan ekonomi. Hal tersebut dapat diketahui melalui berita statistik yang dipublikasi oleh *website* resmi Badan Pusat Statistik (BPS) [1] pada bulan-bulan sebelumnya nilai inflasi sebesar 0.10% sedangkan pada bulan April 2020 menurun sebesar 0.08%. Indeks harga merupakan pengukuran kondisi ekonomi secara general. Indeks Harga Konsumen (IHK) merupakan salah satu indikator untuk mengukur pertumbuhan ekonomi sehingga mengetahui fluktuasi kenaikan ataupun penurunan harga untuk mengetahui perkembangan ekonomi. Proses kenaikan harga secara general dan kontinu disebut sebagai inflasi. Salah satu indeks harga untuk mengukur inflasi antara lain IHK. IHK

---

memberikan informasi terkait perkembangan harga barang atau jasa yang dikonsumsi oleh rumah tangga pada umumnya dalam kurun waktu tertentu [2].

Salah satu kerangka dasar kebijakan ekonomi makro di berbagai negara maju dan berkembang antara lain inflasi dengan tingkat fluktuasi yang minimal. Beberapa penelitian yang telah dilakukan menghasilkan adanya hubungan negatif antara inflasi dengan pertumbuhan ekonomi. Inflasi yang melebihi batas ketentuan (*threshold*) secara nyata menghambat pertumbuhan ekonomi dengan *threshold* inflasi negara maju 1-3% sedangkan negara berkembang 11-12 % [3]. Pentingnya kestabilan inflasi bagi kegiatan ekonomi dalam melakukan konsumsi, investasi, dan produksi sehingga menciptakan ketidakpastian pengambilan keputusan yang menyebabkan pendapatan riil masyarakat akan turun akhirnya menjadikan meningkatnya tingkat kemiskinan [4].

Penelitian yang dilakukan oleh Dewinta [5] mengenai pengelompokan IHK didapatkan hasil 3 klaster dengan klaster 1 tertinggi sebanyak 14 kota, klaster 2 sedang sebanyak 29 kota, dan klaster 3 rendah sebanyak 23 kota. Penelitian mengenai faktor-faktor indeks harga konsumen juga dilakukan oleh Kementerian Perdagangan [6] didapatkan hasil bahwa indeks harga komoditas sumber energi dan makanan mempengaruhi indeks konsumen. Berdasarkan referensi penelitian tersebut peneliti ingin melakukan riset IHK sebagai informasi yang tepat mengenai kondisi konsumsi masyarakat Indonesia di tengah pandemi. Penelitian ini dilakukan untuk menentukan kebijakan pemerintah yang tepat sasaran di tengah pandemi yang serba tidak pasti, seperti menentukan harga komoditas tiap kota dan mengetahui kemampuan daya beli masyarakat tiap kota yang ditandai IHK.

Berdasarkan latar belakang tersebut, dilakukan penelitian untuk melihat 37 variabel indeks harga yang mempengaruhi faktor IHK dan dilakukan pengelompokan 90 kota yang memiliki IHK tertinggi sampai terendah. Pengelompokan menggunakan *K-means* dan pemodelan IHK di Indonesia pada bulan April 2020 di tengah pandemi menggunakan regresi berganda.

## **2. Landasan Teori**

### **2.1. Pemodelan Regresi Berganda.**

Analisis regresi merupakan metode statistika yang menentukan kausalitas variabel prediktor terhadap variabel respon. Analisis regresi berganda digunakan untuk mengetahui variabel prediktor mana saja yang mempengaruhi variabel respon secara signifikansi. Berikut model regresi berganda dalam bentuk matriks.

$$\mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \mathbf{e} \quad (2.1)$$

dengan

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$\mathbf{Y}$  menyatakan vektor variabel prediktor berukuran  $n \times 1$ ,  $\mathbf{X}$  menyatakan matriks variabel prediktor berukuran  $n \times m$ , dengan  $m = k + 1$ ,  $\boldsymbol{\beta}$  menyatakan vektor parameter model berukuran  $m \times 1$  dan  $\mathbf{e}$  menyatakan vektor residu berukuran  $n \times 1$  [7]. Selanjutnya untuk pemodelannya berbentuk persamaan berikut.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i \quad (2.2)$$

dengan

$y$  : variabel respon

$x_i$  : variabel prediktor ( $i = 1, 2, 3, \dots, k$ )

$\beta_0$  : intersep

$\beta_i$  : parameter regresi ( $i = 1, 2, 3, \dots, k$ )

### 2.1.1 Pengujian Serentak

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  (tidak terdapat hubungan signifikan terhadap variabel prediktor secara serentak terhadap variabel respon)

$H_1$  : paling sedikit ada satu  $\beta_i \neq 0$  (terdapat hubungan signifikan terhadap variabel prediktor secara serentak terhadap variabel respon)

Statistik Uji :

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{Y})^2 / k}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - k - 1)} > F_{\alpha, (k, n-k-1)} \quad (2.3)$$

dengan

$\hat{y}_i$  : nilai prediksi dari variabel respon

$\bar{Y}$  : rata-rata dari variabel respon

$k$  : jumlah variabel prediktor

$y_i$  : nilai dari variabel respon

$n$  : jumlah observasi pada baris

$F_{\alpha;k;n-k-1}$  : nilai  $F_{tabel}$

### 2.1.2 Pengujian Parsial

Hipotesis :

$H_0 : \beta_i = 0$  (tidak terdapat hubungan signifikan terhadap variabel prediktor secara parsial terhadap variabel respon)

$H_1 : \beta_i \neq 0$  (terdapat hubungan signifikan terhadap variabel prediktor secara parsial terhadap variabel respon)

Statistik Uji :

$$t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} > t_{\alpha/2, n-k-1} \quad (2.4)$$

dengan

$\hat{\beta}_i$  : parameter terhadap variabel respon

$se(\hat{\beta}_i)$  : standard error dari koefisien  $\hat{\beta}$

$t_{\alpha/2, n-k-1}$  : nilai  $t_{tabel}$

### 2.1.3 Uji Asumsi Klasik

a. Uji Asumsi Normalitas

Menurut Montgomery [8] uji normalitas merupakan metode uji untuk mengetahui kenormalan suatu distribusi data pada sebuah set data atau variabel yang diambil dari populasi normal. Perumusan analisis uji normalitas dapat dilihat sebagai berikut :

i. Hipotesis

$H_0$  : Residu berdistribusi normal

$H_1$  : Residu tidak berdistribusi normal

ii. Statistik Uji

$$D_{hitung} = \sup_x |F_n(x) - F_0(x)| \quad (2.5)$$

iii. Daerah Kritis :  $H_0$  ditolak apabila  $D_{hitung} > D_0$

dengan

$D_{hitung}$  : nilai maksimum dari  $|F_n(x) - F_0(x)|$

$F_n(x)$  : nilai distribusi kumulatif

$F_0(x)$  : nilai distribusi kumulatif normal

b. Uji Asumsi Autokorelasi

Uji autokorelasi digunakan untuk mengetahui adanya hubungan korelasi antar residu dengan menghitung residu pada model untuk mengetahui korelasi pada observasi satu dengan observasi lainnya [8]. Salah satu uji yang dapat dilakukan adalah uji Durbin Watson dengan perumusan analisis sebagai berikut

i. Hipotesis

$H_0 : \rho = 0$  (Residu tidak terdapat autokorelasi)

$H_1 : \rho \neq 0$  (Residu terdapat autokorelasi)

ii. Statistik Uji

$$d_{hitung} = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (2.6)$$

dengan

$d_{hitung}$  : nilai uji Durbin Watson

$e$  : residu

$d_L$  : batas bawah pada tabel Durbin Watson

$d_U$  : batas atas pada tabel Durbin Watson

Kriteria pengujian Durbin-Watson:

1. Apabila nilai  $d_{hitung}$  lebih kecil daripada nilai  $d_L$  atau lebih besar daripada  $(4 - d_L)$ , berarti menolak  $H_0$  yang artinya terdapat autokorelasi.
2. Apabila  $d_{hitung}$  berada di antara  $d_U$  dan  $(4 - d_U)$ , berarti gagal tolak  $H_0$  yang artinya tidak terdapat autokorelasi
3. Apabila  $d_{hitung}$  berada di antara  $d_L$  dan  $d_U$  atau di antara  $(4 - d_U)$  dan  $(4 - d_L)$ , berarti tidak memiliki kesimpulan yang dapat ditarik.

c. Uji Asumsi Identik atau Homogenitas

Uji homogenitas merupakan uji model regresi yang bertujuan mendeteksi ketidaksamaan varian residu dari observasi satu ke observasi lain. Varian residu mengalami konstan dari satu observasi ke observasi lain, maka disebut homokedastisitas dan sebaliknya disebut heterokedastisitas [8]. Model regresi yang baik adalah model yang homoskedastisitas (tidak terjadi heteroskedastisitas). Masalah heteroskedastisitas dapat dideteksi dengan menggunakan uji *Breusch Pagan Godfrey* (BPG) sebagai berikut :

i. Hipotesis :

$$H_0 : \text{Variansi residu bersifat homoskedastisitas atau } \text{var}(e_i) = \sigma^2$$

$$H_1 : \text{Variansi residu bersifat heteroskedastisitas atau } \text{var}(e_i) \neq \sigma^2$$

Statistik Uji :

$$\phi = \frac{1}{2} \times SS_{\text{regresi}} \quad (2.7)$$

ii. Daerah Kritis :

$$H_0 \text{ ditolak jika } \phi_{\text{hitung}} > \chi_{df(k-1)}^2$$

dengan

$\phi$  : nilai statistik uji *Breusch Pagan*

$SS_{\text{regresi}}$  : *sum of square* dari regresi

$\chi_{df(k-1)}^2$  : nilai tabel *chi-square* dengan derajat bebas  $k-1$ ,  $k$  berarti jumlah variabel prediktor

**2.2. Analisis Klaster Menggunakan K-means.** *Clustering* adalah mengelompokkan objek kedalam beberapa kelompok yang terletak dalam suatu kelompok yang memiliki sifat relatif homogen. Selain itu *clustering* dan grup merupakan dua hal yang berbeda karena pada grup pengelompokan anggotanya kondisinya harus sama, sedangkan pada *clustering* pengelompokan berdasarkan jarak kedekatan dari suatu karakteristik sampel observasi [9]. Menurut Malik dan Tuckfield [10] langkah awal algoritma *K-means* dengan menentukan nilai  $k$  secara random,  $k$  merupakan banyaknya klaster yang ingin dibentuk dengan nilai secara random atau sesuai dengan penelitian dan nilai tersebut menjadi pusat klaster atau *centroid*. Setelah mendapatkan nilai *centroid* selanjutnya menentukan jarak terdekat setiap data terhadap masing-masing *centroid* dengan menggunakan rumus jarak *Euclidean*. Berikut merupakan rumus jarak *Euclidean* [10].

---

$$D(x) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (2.8)$$

Selanjutnya untuk mengetahui  $k$  optimum untuk meminimalisir kekurangan algoritma  $K$ -means [11] dengan rumus

$$K = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (2.9)$$

dengan

$D(x)$  : nilai *Euclidean Distance*

$i$  : baris data

$x_i$  : data ke- $i$

$c_i$  : *centroid* data

$D(x)^2$  : jarak *Euclidean*

$\sum_{x \in X} D(x)^2$  : jumlah jarak *Euclidean*.

Ringkasan dari langkah-langkah algoritma  $K$ -means menurut Emily [10] sebagai berikut :

1. Memilih  $k$  pusat kluster pertama secara random.
2. Mengamati  $x$  dari setiap data dan menghitung jarak terdekat  $D(x)$  menuju pusat kluster.
3. Memilih kluster baru diantara titik data dengan probabilitas  $x$  secara proporsi jarak  $D(x)$ .
4. Mengulangi langkah 2 dan 3 sehingga  $k$  pusat kluster terpilih.

### 3. Hasil dan Pembahasan

Penelitian ini menggunakan data katalog Indikator Ekonomi bulan April 2020 yang dipublikasi setiap bulan oleh *website* resmi Badan Pusat Statistika (BPS) [1]. Variabel yang digunakan yaitu indikator Indeks Harga Konsumen (IHK) yang terdiri dari 37 variabel indikator dengan 11 kriteria. Adapun variabel penelitian dapat dilihat pada Tabel 1.

Untuk mengetahui faktor-faktor indeks harga yang mempengaruhi Indeks Harga Konsumen (IHK) secara signifikan menggunakan analisis regresi berganda, sehingga dapat mengetahui faktor harga komoditi apa saja yang mempengaruhi harga-harga secara umum yang mengalami penurunan (deflasi) atau kenaikan harga (inflasi).

Tabel 1. Operasional Variabel Indeks Harga Konsumen April 2020

Kriteria	Variabel	Keterangan
Makanan, Minuman, dan Tembakau	$X_1$	Makanan
	$X_2$	Minuman yang tidak beralkohol
	$X_3$	Rokok dan tembakau
Pakaian dan Alas Kaki	$X_4$	Pakaian
	$X_5$	Alas Kaki
	$X_6$	Sewa dan Kontrak Rumah
Perumahan, Air, Listrik, dan Bahan Bakar Rumah Tangga	$X_7$	Pemeliharaan, Perbaikan, dan Keamanan Tempat Tinggal
	$X_8$	Penyediaan Air dan Layanan Perumahan Lainnya
	$X_9$	Listrik dan Bahan Bakar Rumah Tangga
	$X_{10}$	Furnitur, Perlengkapan, dan Karpet
	$X_{11}$	Tekstil Rumah Tangga
	$X_{12}$	Peralatan rumah tangga
	$X_{13}$	Barang Pecah Belah dan Peralatan Makan dan Minum
Perlengkapan, Peralatan, dan Pemeliharaan Rutin Rumah tangga	$X_{14}$	Peralatan dan Perlengkapan Perumahan dan Kebun
	$X_{15}$	Barang dan Layanan Untuk Pemeliharaan Rumah Tangga Rutin
	$X_{16}$	Obat-Obatan dan Produk Kesehatan
	$X_{17}$	Jasa Rawat Jalan
Kesehatan	$X_{18}$	Jasa Rawat Inap
	$X_{19}$	Jasa Kesehatan lainnya
	$X_{20}$	Pembelian Kendaraan
	$X_{21}$	Pengoperasian Peralatan Transportasi Pribadi
	$X_{22}$	Jasa Angkutan Penumpang
	$X_{23}$	Jasa Pengiriman Barang



Lanjutan Tabel 1. Operasional Variabel Indeks Harga Konsumen April 2020

Kriteria	Variabel	Keterangan
Informasi, Komunikasi, dan Jasa Keuangan	X <sub>24</sub>	Peralatan Informasi dan Komunikasi
	X <sub>25</sub>	Layanan informasi dan Komunikasi
	X <sub>26</sub>	Jasa keuangan
Rekreasi, Olahraga, dan Budaya	X <sub>27</sub>	Barang Rekreasi lainnya dan Olahraga
	X <sub>28</sub>	Layanan Rekreasi dan Olahraga
	X <sub>29</sub>	Koran, Buku dan Perlengkapan Sekolah
	X <sub>30</sub>	Pendidikan Dasar dan Anak Usia Dini
Pendidikan	X <sub>31</sub>	Pendidikan Menengah
	X <sub>32</sub>	Pendidikan Tinggi
	X <sub>33</sub>	Pendidikan Lainnya
Penyediaan Makanan dan Minuman	X <sub>34</sub>	Jasa Pelayanan Makanan dan Minuman
Perawatan Pribadi dan Jasa Lainnya	X <sub>35</sub>	Perawatan Pribadi
	X <sub>36</sub>	Perawatan Pribadi Lainnya
	X <sub>37</sub>	Jasa Lainnya

Hasil analisis regresi berganda metode backward didapatkan 13 variabel dari 37 variabel yang berpengaruh signifikan terhadap IHK dan didapatkan pemodelan regresi terbaik sebagai berikut :

$$\hat{y} = -5,144 + 0,256x_1 + 0,008x_3 + 0,031x_5 + 0,087x_6 + 0,124x_9 + 0,129x_{21} + 0,026x_{22} + 0,077x_{25} + 0,016x_{30} + 0,27x_{32} + 0,014x_{34} + 0,133x_{35} + 0,023x_{36} \quad (3.1)$$

Pada model 3.1 dapat disimpulkan bahwa saat pandemi indeks harga komoditi yang meningkatkan IHK terdapat pada koefisien tinggi pada variabel makanan, listrik dan bahan bakar rumah tangga, pengoperasian peralatan transportasi pribadi dan perawatan pribadi. Hasil uji serentak (Uji *F*) didapatkan nilai  $F_{hitung}$  sebesar 16,332 dengan *p-value* sebesar 0,000 yang berarti secara serentak variabel independen berpengaruh signifikan terhadap variabel dependen. Sedangkan hasil uji parsial (uji *t*) dapat dilihat pada Tabel 2.

Selanjutnya uji asumsi klasik regresi berganda yaitu uji normalitas, uji autokorelasi, dan uji heteroskedastisitas dengan hasil ditunjukkan pada Tabel 3.

Tabel 2. Hasil Uji Parsial Terhadap Faktor-Faktor IHK

Variabel	$t_{hitung}$	$p$ -value	Kesimpulan
$X_1$ (Makanan)	13,773	0,000	Signifikan
$X_3$ (Rokok dan tembakau)	2,085	0,042	Signifikan
$X_5$ (Alas kaki)	2,205	0,032	Signifikan
$X_6$ (Sewa dan kontrak rumah)	2,311	0,025	Signifikan
$X_9$ (Sewa dan kontrak rumah)	3,216	0,002	Signifikan
$X_{21}$ (Pengoperasian peralatan transportasi pribadi)	3,897	0,000	Signifikan
$X_{22}$ (Jasa Angkutan Penumpang)	4,760	0,000	Signifikan
$X_{25}$ (Layanan informasi dan komunikasi)	1,992	0,049	Signifikan
$X_{30}$ (Pendidikan dasar dan anak usia dini)	2,285	0,026	Signifikan
$X_{32}$ (Pendidikan tinggi)	2,764	0,008	Signifikan
$X_{34}$ (Jasa pelayanan makanan dan minuman)	2,436	0,018	Signifikan
$X_{35}$ (Perawatan pribadi)	4,471	0,000	Signifikan
$X_{36}$ (Perawatan pribadi lainnya)	2,587	0,013	Signifikan

Tabel 3. Hasil Uji Asumsi Klasik

Uji Asumsi Klasik	Nilai Hitung	$p$ -value	Kesimpulan
Uji Normalitas ( <i>Kolmogorov Smirnov</i> )	0,670	0,200	Gagal tolak $H_0$ yang menyimpulkan data berdistribusi normal dan asumsi normalitas terpenuhi.
Uji Autokorelasi ( <i>Durbin Watson</i> )	1,889	0,267	Nilai $d_U$ dan $d_L$ pada tabel Durbin-Watson ( $n = 90, k = 13$ ) sebesar 1,3693 dan 1,9659, dengan begitu nilai autokorelasi di antara $1,3693 < 1,8892 < 1,9659$ yang menyimpulkan bahwa tidak terjadi autokorelasi sehingga asumsi terpenuhi.
Uji Heteroskedastisitas ( <i>Breusch Pagan</i> )	32,595	0,675	Gagal tolak $H_0$ yang menyimpulkan tidak terjadi heteroskedastisitas sehingga asumsi terpenuhi

Tabel 3 menunjukkan bahwa uji asumsi klasik regresi berganda terpenuhi. Selanjutnya dilakukan klusterisasi dengan metode  $K$ -means berdasarkan variabel yang berpengaruh signifikan terhadap IHK. Objek-objek yang diamati dikelompokkan berdasarkan jarak terdekat dengan menggunakan jarak *Euclidean*. Berdasarkan hasil perhitungan normalisasi data dan jarak *Euclidean* maka dilakukan klusterisasi dengan nilai rata-rata  $K$ -means tertinggi merupakan kluster tertinggi dan sebaliknya yang rendah

merupakan klaster terendah. Hasil visualisasi klaster ditunjukkan pada Gambar 1 dan pembagian klaster yang terbentuk pada Tabel 4.

Tabel 4. Hasil Pembagian Klaster Indeks Harga 90 Kota di Indonesia

Klaster	Anggota Hasil Klasterisasi 90 Kota di Indonesia
1	Tanjung Pinang dan Singkawang
2	Banda Aceh, Pekanbaru, Bungo, Bandar Lampung, Metro, Bogor, Bekasi, Semarang, Yogyakarta, Cilegon, Tangerang, Serang, Singaraja, Wangaipu, Maumere, Sintang, Pontianak, Tarakan, Manado, Kotamobagu, Luwuk, Palu, Bulukumba, Watampone, Makassar, Paolopo, Mamuju, Tual, Ternate, Manokwari, Merauke, dan Jayapura
3	Meulaboh, Padangsidempuan, Lhoksumawe, Pematangsiantar, Gunungsitoli, Pangkal Pinang, Batam, Tanjung Pinang, Sibolga, Medan Padang, Tembilahan, Bukittinggi, Dumai, Palembang, Jambi, Lubuklinggau, Bengkulu, Tanjung Pandan, Pangkal Pinang, Batam, Bogor, Sukabumi, Jakarta, Bekasi, Cirebon, Tasikmalaya, Depok, Cilacap, Purwokerto, Surakarta, Kudus, Semarang, Yogyakarta, Tegal, Banyuwangi, Jember, Sumenep, Malang, Kediri, Probolinggo, Surabaya, Madiun, Denpasar, Mataram, Bima, Kupang, Sampit, Palangkaraya, Kotabaru, Tanjung, Banjarmasin, Balikpapan, Samarinda, Tanjung Selor, Pare-pare, Kendari, Gorontalo, Baubau, Ambon, Sorong dan Timika,

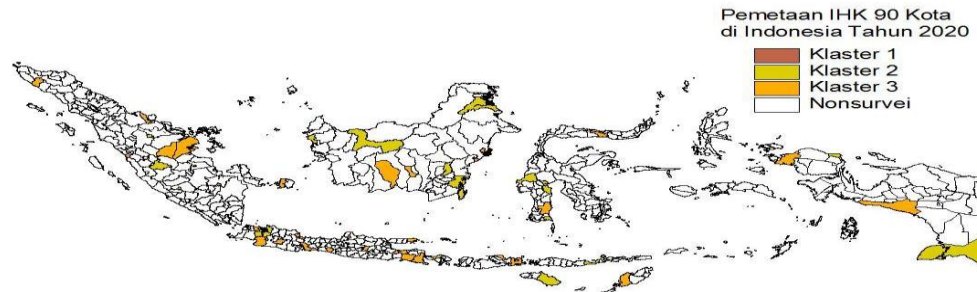
Selanjutnya menghitung rata-rata masing-masing klaster terhadap variabel untuk menentukan klaster yang memiliki indeks harga tinggi, sedang dan rendah. Hasil rata-rata masing-masing klaster ditunjukkan pada Tabel 5.

Tabel 5. Rata-rata Klaster Terhadap Variabel

Klaster	$X_1$	$X_3$	$X_5$	$X_6$	$X_9$	$X_{21}$	$X_{22}$
1	0,229	0,121	-0,380	0,868	2,028	1,395	0,662
2	-0,359	-0,458	0,171	-0,049	0,255	0,339	0,214
3	0,109	0,149	-0,035	-0,037	-0,214	-0,203	-0,114
Klaster	$X_{25}$	$X_{30}$	$X_{34}$	$X_{35}$	$X_{36}$		
1	0,791	1,037	-0,646	-1,813	0,750		
2	-1,128	-0,324	0,921	0,158	-0,493		
3	0,338	0,046	-0,276	0,058	0,122		

Variabel  $X_1$  rata-rata tertinggi pada klaster 1, sedang pada klaster 3, dan rendah pada klaster 2 begitu seterusnya hingga variabel  $X_{35}$ . Untuk rata-rata tertinggi lebih banyak pada klaster 1, rata-rata sedang pada klaster 3 dan rata-rata rendah pada klaster 2

sehingga dihasilkan bahwa pengelompokan klaster 90 kota menjadi 3 kelompok yaitu kategori tinggi (klaster 1), kategori sedang (klaster 3), dan kategori rendah (klaster 2).



Gambar 1. Pemetaan Berdasarkan Hasil Pengklasteran Menggunakan *K-means*

#### 4. Kesimpulan

Analisis regresi untuk mengetahui indeks harga yang berpengaruh signifikan terhadap Indeks Harga Konsumen (IHK) adalah  $X_1$  (makanan),  $X_3$  (rokok dan tembakau),  $X_5$  (alas kaki),  $X_6$  (sewa dan kontrak rumah),  $X_9$  (listrik dan bahan bakar rumah tangga),  $X_{21}$  (pengoperasian peralatan transportasi pribadi),  $X_{22}$  (jasa angkutan penumpang),  $X_{25}$  (layanan komunikasi dan informasi),  $X_{30}$  (pendidikan dasar dan anak usia dini),  $X_{32}$  (pendidikan tinggi),  $X_{34}$  (jasa pelayanan makanan dan minuman),  $X_{35}$  (perawatan pribadi), dan  $X_{36}$  (perawatan pribadi lainnya). Berdasarkan analisis klaster *K-means* didapatkan 2% kota klaster tinggi, 68% kota klaster sedang, 30% kota klaster rendah. Klasterisasi yang telah dilakukan relevan dengan klasterisasi yang dilakukan Dewinta [5] pada data tahun 2012 di 66 kota di Indonesia. Ada perbedaan proporsi klaster tinggi, sedang dan rendah. Pada penelitian Dewinta [5] menghasilkan proporsi klaster tinggi 21%, klaster sedang 43% dan klaster rendah 36%. Kota yang sama posisinya pada klaster sedang adalah Lhoksumawe, Padang, Batam, Jambi, Bengkulu, Cirebon, Tasikmalaya, Purwokerto, Surakarta, Yogyakarta, Kediri, Malang, Denpasar, Mataram, Kupang, Sampit, Banjarmasin, Balikpapan, Ambon. Kota yang sama posisinya pada klaster rendah adalah Bogor, Tangerang, dan Watampone. Perbedaan ini terjadi karena perbedaan tahun penelitian, variabel yang digunakan, dan pandemi Covid-19 yang menurunkan tingkat konsumsi masyarakat sehingga adanya ketidakseimbangan antara produksi dan konsumsi yang menyebabkan IHK cenderung mengelompok pada klaster sedang. Variabel komoditas indeks harga yang berpengaruh secara signifikan dapat

diseimbangkan lagi produksi dan konsumsi untuk menjaga stabilitas harga komoditi agar harga tersebut dapat dijangkau masyarakat dengan harga yang wajar. Hal tersebut dilakukan karena jika produksi barang meningkat tidak diikuti dengan tingkat konsumsi akan menyebabkan lemahnya pasar modal. Sebaliknya, jika produksi turun sedangkan konsumsi tinggi akan menyebabkan kenaikan harga yang tidak stabil dan terjadi inflasi.

### Daftar Pustaka

- [1] Badan Pusat Statistik. *Indikator Ekonomi April 2020*. Jakarta. BPS RI. 2020.
- [2] Syahyuti. *Indeks Harga Konsumen*. URL : <http://syahyutivariabel.blogspot.co.id/2011/03/indeks-hargakonsumen.html>, diakses pada 27 September 2020.
- [3] Khan, M. dan Senhadji, A. Threshold Effect in Relationship Between Inflation and Growth. *IMF Working Paper*. 2000.
- [4] Bank Indonesia. URL: <http://www.bi.go.id/id/moneter/inflasi/pengenalan/Contents/Pentingnya.aspx>, diakses pada 15 Oktober 2020.
- [5] Sinaga, M. D., Windarto P. A., Hartama, D., dan Saifullah. Pengelompokan Indeks Harga Konsumen Menurut Kota dengan Data Mining Clustering. *Prosiding Seminar Nasional Sains & Teknologi Informasi (SENSASI)*. Pematangsiantar, Indonesia. 328-336. 2019.
- [6] Kementerian Perdagangan. *Laporan Akhir Analisis Kebijakan Pemerintah terhadap Inflasi dan Kemiskinan*. 2015.
- [7] Lindley, D. V. Regression and Correlation Analysis. *Journal of New Palgrave : A Dictionary of Economics*. London. 1987.
- [8] Montgomery, D. and Peck, E. *Introduction to Linear Regression Analysis*. John Wiley & Sons Inc. New York. 2006.
- [9] Satriyanto, E. *Kombinasi Algoritma Agglomerative Clustering dan K-means untuk Segementasi Pengunjung Website*. Skripsi. Universitas Telkom. 2011.
- [10] Malik, A. and Tuckfield, B. *Applied Unsupervised Learning with R*. Packt Publishing. University of Birmingham. United Kingdom. 2019.
- [11] Sukarhat, G. *Analisis dan Implementasi Algoritma K-means pada Clustering*. Skripsi. Universitas Telkom. 2011.
- [12] Aicher, C. and Emily, B. Scalable Clustering of Correlated Time Series using Expectation Propagation. *Journal Semantic Scholar*. University of Washington. 2016.