p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
12(1) 2024: 10-23

On Clustering Consumption Patterns In West Java To Determine Sister Cities For The Non-Sampled Area Cost Of Living Survey

Raditya Novidianto [1*], Erwin Tanur [2], Andrea Tri Rian Dani [3], Fachrian Bimantoro Putra [4]
[1] Badan Pusat Statistik Kabupaten Kuningan
[2] Pusat Pendidikan dan Pelatihan (Pusdiklat) BPS
[3,4] Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman

***e-mail**: radit@bps.go.id

Article Info:
Received: March 23, 2024
Accepted: May 29, 2024
Available Online: June 8, 2024

Keywords:
ACF, Biclustering, CID, Euclidean, Similarity
.

Abstract: In the Cost of Living Survey activities in the districts / cities affected by the sample, the data collected is data on how much expenditure is needed by households in daily life as a commodity package. This commodity package is close to the socio-economic survey (SUSENAS) for non-sample districts / cities in determining "Sister Cities" which do not have certain indicators such as inflation. This inflation is very important to use as a basis for calculating the poverty line, so care needs to be taken in determining "Sister cities" regarding the addition of inflation cities in West Java. With the help of machine learning, the similarity level method using distance measures, namely Euclidean distance, CID distance, and ACF distance, can help districts / cities find sister cities. Furthermore, grouping was carried out using a biclustering algorithm to see the characteristic variables in West Java from the same consumption pattern data. The biclustering parameter with tuning parameter δ=0.1 is the best bicluster with a total of 3 biclusters with a value of MSR/V=0.02433 with identical characteristic variables, namely Average Fish Consumption (X3), Average Meat Consumption (X4), Average Consumption of Eggs and Milk (X5), Average Consumption of Vegetables (X6), Average Consumption of Fruit (X8), Average Consumption of Oil and Coconut (X9), Average Consumption of Housing and Household Facilities (X15), Average Consumption of Various Goods and Services and Average Consumption of Taxes (X16), Levies and Insurance (X19).

1. INTRODUCTION

Inflation is an indicator that represents changes in the average price of a group of goods and services represented in the consumer or producer shopping basket [1]. These measurements are carried out in a Cost of Living Survey (SBH) weighing diagram [2]. In other words, inflation results in a reduction in the purchasing power of the currency because goods and services become more expensive [3]. West Java had 7 inflation cities in 2018, namely Bogor City, Sukabumi City, Bandung City, Cirebon City, Bekasi City, Depok City, and Tasikmalaya City [4] plus 3 new inflation districts / cities in 2022, namely Bandung districts, Majalengka

[DOI: 10.14710/JSUNIMUS. 12.1.2024.10-23]

districts and Subang districts. Meanwhile, West Java has 17 districts / cities that are not SBH samples, so they do not have inflation figures.

Inflation figures are really needed in SBH non-sample districts / cities because the calculation of several important indicators such as poverty, Wholesale Price Index, Large Price Index, Producer Price Index, Gross Domestic Product Deflator, and Asset Price Index requires the role of inflation [5]. For example, in calculating districts / cities poverty figures, inflation figures should be included because multiplying the previous year's poverty line with year-to-year inflation will produce a provisional poverty line for the year concerned [6]. Therefore, to determine the inflation value in districts / cities that do not carry out SBH activities, BPS uses a sister city approach. [7]. This approach uses a weighing diagram of a city implementing SBH that has the same consumption pattern and is geographically close [8]. The impact of errors in determining sister cities is on the policies that will be taken, for example, an increase in the minimum wage, which has implications for the level of workers' welfare, where the determination requires inflation data [7]. So, a separate study is needed regarding the determination of sister cities, especially on inflation. So far, in determining sister cities, they usually only use a geographic proximity approach, but it is necessary to consider using data patterns to calculate the sister city approach [9].

In SBH activities in the districts / cities affected by the sample, the data collected is data on how much expenditure is required by individuals or households in daily life [10]. SBH will produce commodity packages that will be used as objects of observation for every price change every month so that inflation figures appear [11]. Therefore, in determining sister cities, what is used to see the level of similarity is community consumption patterns, GDP per capita, population, and distance from the city of inflation. Consumption patterns themselves are important to see any anomalies that occur in an area, so there is a need for special policies related to this [12].

Therefore, measurements were carried out using a simple method, machine learning, to determine the level of similarity of several variables, such as average expenditure on commodity consumption patterns, population, GDP per capita, and distance from the city of inflation used in 2022. In addition, the Grouping of consumption patterns is carried out to provide insight to the government to make policies based on the characteristic variables that form these groupings using the Cheng and Chen biclustering algorithm.

2. LITERATURE REVIEW

2.1. Distance Measure

Distance measurement is one of the key components of clustering. This data can be in various forms, including raw values from equal or unequal euclides, euclidean pairs of feature values, transition matrices, etc. Several distances will be used in time series cluster analysis according to Liao, T.W [13], namely as follows: *Eclidean distance, root mean square distance, and Mikowski distance, Pearson's correlation coefficient and related, Short time series distance (STS), Dynamic time warping distance (DTW), Probability-based distance function for data with error, Kullback-Liebler distance, J divergence and symmetric information divergence, Dissimilarity index based on the cross-correlation function between two time series*.

The distance measures used in this research are *uclidean distance*, *autocorrelation function based distance (ACF)*, *complexity-invariant distance (CID)*, and *dynamic time warping (DTW)*. To determine the best distance measure, cophenetic correlation will be used.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

a. Euclidean distance

Euclidean distance is calculated from raw data, not standard data [14]. The advantage of using Euclidean distance is that adding a new object does not affect the distance between 2 objects, even if the new object is an outlier. However, the distance is greatly influenced by the scale (if the scale is changed, for example, from cm to mm, the cluster analysis results may differ). Euclidean distance is a distance matrix x_i dan v_j each becomes a vector *P- dimensional*. The Euclidean Distance formula is calculated as follows [15].

$$dist(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

b. Autocorrelation Function (ACF) Distance

A matrix based on estimating the autocorrelation function (ACF) for time series data with a correlation structure [16]. For example $\widehat{\boldsymbol{p}}_{\boldsymbol{x}} = (\widehat{\boldsymbol{p}}_{\boldsymbol{1x}}, \widehat{\boldsymbol{p}}_{\boldsymbol{2x}}, \dots, \widehat{\boldsymbol{p}}_{\boldsymbol{Lx}})^{\boldsymbol{t}}$, and $\widehat{\boldsymbol{p}}_{\boldsymbol{y}} = (\widehat{\boldsymbol{p}}_{\boldsymbol{1y}}, \widehat{\boldsymbol{p}}_{\boldsymbol{2y}}, \dots, \widehat{\boldsymbol{p}}_{\boldsymbol{Ly}})^{\boldsymbol{t}}$ are autocorrelation vectors resulting from the estimation of time series *X* and *Y* for some *L* like $\widehat{\boldsymbol{p}}_{\boldsymbol{ix}} \approx 0$ dan $\widehat{\boldsymbol{p}}_{\boldsymbol{ix}} \approx 0$ for $i > L$. The distance between two time series data can then be formed as follows:

$$d_{ACF} = \sqrt{\left(\widehat{\boldsymbol{p}}_{\boldsymbol{x}} - \widehat{\boldsymbol{p}}_{\boldsymbol{y}}\right)' \Omega \left(\widehat{\boldsymbol{p}}_{\boldsymbol{x}} - \widehat{\boldsymbol{p}}_{\boldsymbol{y}}\right)} \tag{2}$$

Where $\boldsymbol{\Omega}$ is the weighting matrix.

c. Complexity-Invariant Distance (CID) Distance

Complexity-invariant distance (CID) is a measure that uses the difference in data variation between two-time series data to correct the basic distance used [17]. The basic distance commonly used is the Euclidean distance. The complexity-invariant Distance (CID) distance calculation is as follows:

$$CID(u,v) = ED(u,v).CF(u,v) \tag{3}$$

with:
CID(u,v) : Complexity-invariant distance between data *u* and data *v*
ED(u,v) : Euclidean distance between data *u* and data *v*
CF(u,v) : Correction factor between *u* and data *v*
The correction factor (*CF*) will separate data based on data variations. The higher the variation in the data being compared, the greater the *CF* value. If the data being compared has the same variation, then the *CID* distance will be closer to the Euclidean distance.

$$CF(u,v) = \frac{\max\left(CE(u), CE(v)\right)}{min\left(CE(u), CE(v)\right)} \tag{4}$$

$CE(u) = \sqrt{\sum_{t=1}^{n-1}(u_t - u_{t+1})^2}$, Where *t* is the *t-th* data

CE(u) estimates variations in time series data by straightening the time series data. Data with high variation will have longer lines than data with low variation. For the resulting *CE(u)* value to be consistent, the time series data must meet 3 conditions: the same number of data series, sampling rate, and standardized amplitude. The CID distance tends to be smaller for data of the same complexity and should increase when compared with data of different complexity.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

2.2. Biclustering

Then, after measuring the similarity distance, further analysis is carried out using the machine learning method, namely Bicluster [18]. Bicluster is defined as part of a row and part of a column with a high similarity value based on the Cheng and Church (CC) algorithm [19]. This algorithm aims to find biclusters with a Mean Square Residual (MSR) smaller than the specified limit [20].

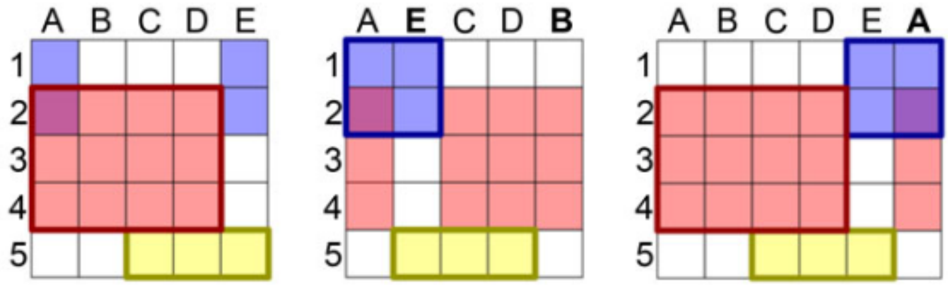

Figure 1 Representation of outcomes from clustering and biclustering

The biclustering analysis used in this research is the CC algorithm (Cheng & Church) or δ - biclustering [19]. The CC algorithm is greedy and tries to find the maximum bicluster with high similarity. A collection of rows and columns is called a bicluster if it has a mean squared residual value below the level (δ) determined by researchers [21].

3. METHODOLOGY

3.1 Data

The data used in this research are consumption patterns derived from the 2022 National Socio-Economic Survey (SUSENAS), GDP per capita, population, and a distance matrix to determine sister cities of non-sample SBH districts / cities from SBH cities in the West Java region, while for The grouping only uses 2022 Susenas data. The year 2022 was chosen considering changes in the base year for inflation, previously 2018 [22]. In this study, the Euclidean distance, CID distance, and ACF distance from non-sample districts / cities to the SBH sample in 2023 will be calculated to determine the characteristics from the Level of Similarity based on the heatmap [23]. The heatmap is formed from a distance measurement matrix. After that, a two-way grouping was created using the Cheng and Church (CC) bicluster algorithm to determine the characteristics of accumulation based on household consumption patterns, which is useful as a basis for policymakers through characteristic variables. districts / cities use 23 food and non-food consumption pattern variables to determine sister cities. The data used in this study are as follows.

Table 1 Variables used in research

Variable	Definition	Unit
X1	Average Grain Consumption	Rupiah
X2	Average Consumption of Tubers	Rupiah
X3	Average Fish Consumption	Rupiah
X4	Average Meat Consumption	Rupiah
X5	Average Consumption of Eggs and Milk	Rupiah
X6	Average Consumption of Vegetables	Rupiah
X7	Average Consumption of Nuts	Rupiah

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

Variable	Definition	Unit
X8	Average Fruit Consumption	Rupiah
X9	Average Consumption of Oil and Coconut	Rupiah
X10	Average Consumption of Beverage Ingredients	Rupiah
X11	Average Consumption of Spices	Rupiah
X12	Average Consumption Other Consumption	Rupiah
X13	Average Consumption of Prepared Food and Beverages	Rupiah
X14	Average Cigarette and Tobacco Consumption	Rupiah
X15	Average Consumption of Housing and Household Facilities	Rupiah
X16	Average Consumption of Various Goods and Services	Rupiah
X17	Average Consumption of Clothing, Footwear and Headgear	Rupiah
X18	Average Consumption of Durable Goods	Rupiah
X19	Average Consumption of Taxes, Levies and Insurance	Rupiah
X20	Average Consumption for Parties and Ceremonies/Kenduri	Rupiah
X21	Total population	Soul
X22	GRDP Per Capita	Thousands of Rupiah
X23	Inter- districts / cities Distance Matrix	Km

Euclidean distance is one of the most frequently used and popular distance measures to measure the similarity [24] between non-SBH districts / cities and SBH districts / cities. However, similarity measurement models will also be compared using CID and ACF distance to determine sister cities. This distance measure is usually used by clustering algorithms, such as K-means, K-medoids, and hierarchical clustering [25]. However, this method is only one way to see the level of similarity between the research object and the variables; a machine learning algorithm is used, namely bicluster. So, the distance matrix will be used for cities with inflation against non-sample Cost of Living Survey cities to form a grouping due to the level of similarity. The method used in this study is finding the closest distance between non-sample districts / cities and the Cost of Living Survey (SBH) sample between West Java regions. The data structure used in this study is as follows.

Regency/City	X1	X2	X3	...	X22
3201_BOGOR	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	...	$x_{1,22}$
3202_SUKABUMI	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	...	$x_{2,22}$
3203_CIANJUR	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	...	$x_{3,22}$
3204_BANDUNG	$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	...	$x_{4,22}$
3205_GARUT	$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	...	$x_{5,22}$
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
3279_BANJAR	$x_{27,1}$	$x_{27,2}$	$x_{27,3}$	...	$x_{27,22}$

Figure 2 Data structure used in the research

The distance is calculated from non-SBH districts / cities to SBH sample districts / cities. In 2022, there will be 10 SBH sample districts / cities in West Java, namely Bogor City, Sukabumi City, Bandung City, Cirebon City, Bekasi City, Depok City, and Tasikmalaya City, Bandung districts, Majalengka districts and Subang districts while the remaining 17 districts / cities, is non-SBH. The distance similarity measure is one of the most frequently used and popular distance measures to measure the similarity between non-SBH districts / cities and SBH districts / cities.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

3.2 Analysis Steps

The steps of the data analysis method using machine learning methods to identify sister cities are explained below:

1. Exploring the data with descriptive statistics and creating a spatial mapping for each variable used to identify SBH's non-sample sister cities.
2. As a first step before using machine learning methods to determine the level of similarity and group data, standardize the data using Z-Score. Standardization is used to scale the data for each variable within the same range.
3. Distance measures were calculated using Euclidean distance, CID distance, and ACF distance to identify the level of similarity.
4. Apply the bicluster algorithm to the grouping process and tune the parameters δ =0.1, 0.2, 0.3,..., 1 to determine the number of biclusters and characteristic variables for each bicluster formed.
5. Calculate MSR/V for each tuning parameter to get the best group. The best MSR/V value has the smallest value.
6. Profiling the results of the districts / cities grouping results from measuring the level of similarity and grouping in West Java by visualizing it with spatial mapping.

4. RESULTS AND DISCUSSION

4.1 Descriptive

Each districts / cities must have an inflation rate when calculating poverty rates [26]. Inflation figures are used to calculate the temporary poverty line to determine households in the bottom 40 percent [27]. After that, the unit price of a calorie is calculated and then converted to the standard calorie, namely 2100 Cal, to obtain a poverty line for cutting points for how many households live below the poverty line [28]. Following the SBH non-sample regencies/cities, the use of sister cities in the base year of inflation is the same as in 2018. It is not only inflation that requires sister cities but also GDP calculations as well. However, this research was focused on looking for sister cities from non-SBH cities where this inflation indicator is a determinant of calculating the poverty rate. The following is a table of cities that are not SBH.

Table 2 Sister City in each non-SBH sample districts / cities in the 2018 base year

districts / cities	Sister City people's welfare	Sister City Geographic	Sister City PDRB
Bogor	Bogor city	Bogor city	Bogor city
Sukabumi	Sukabumi City	Sukabumi City	Sukabumi City
Cianjur	Sukabumi City	Sukabumi City	Sukabumi City
Garut	Tasikmalaya City	Tasikmalaya City	Tasikmalaya City
Tasikmalaya	Tasikmalaya City	Tasikmalaya City	Tasikmalaya City
Ciamis	Tasikmalaya City	Tasikmalaya City	Tasikmalaya City
Kuningan	Cirebon City	Cirebon City	Cirebon City
Cirebon	Cirebon City	Cirebon City	Cirebon City
Sumedang	Bandung	Cirebon City	Cirebon City
Indramayu	Cirebon City	Cirebon City	Cirebon City

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

districts / cities	Sister City people's welfare	Sister City Geographic	Sister City PDRB
Purwakarta	Bandung	Cirebon City	Cirebon City
Karawang	Bekasi city	Bekasi city	Bekasi city
Bekasi	Bekasi city	Bekasi city	Bekasi city
Bandung Barat	Bandung	Bandung	Bandung
Pangandaran	Tasikmalaya City	Tasikmalaya City	Tasikmalaya City
Kota Cimahi	Bandung	Bandung	Bandung
Kota Banjar	Tasikmalaya City	Tasikmalaya City	Tasikmalaya City

The table 2 shows that the determination of sister cities is based more on the conditions surrounding the non-SBH sample districts / cities. So, distance is something that is taken into consideration when determining sister cities in non-SBH sample districts / cities. The following is the distribution of each variable, as depicted in the next image.

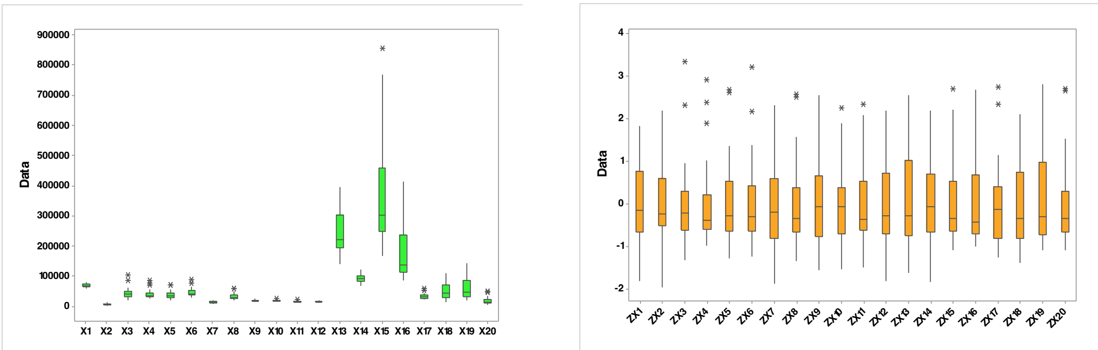

Figure 3 Boxplot of Monthly Household Consumption Expenditure Data in West Java Based on Expenditure Groups

Data on research variables can be depicted in a box plot. Each consumption variable certainly has different units, causing the distribution of data to vary in the characteristics of each variable. Therefore, standardization is carried out. If the data does have significantly different units, a standardization process is carried out by changing the existing data to Z-Score [29]. The standardization process will mean that two data with vast unit differences will automatically become narrower. Data scaling or standardization is the process of changing numerical values in a data set into a general scale without changing the data distribution [30]. This is illustrated in the boxplot above.

4.2 Distance Measure

In this research, distance measurements were calculated. The purpose of this distance measurement is to determine sister cities through data patterns. Several methods were carried out so that these methods would be compared so that this gap emerged in this research. This determination of sister cities can be done by looking at the consistency of several methods used.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

	3204_BANDUNG	3210_MAJALENGKA	3213_SUBANG	3271_BOGOR	3272_SUKABUMI	3273_BANDUNG	3274_CIREBON	3275_BEKASI	3276_DEPOK	3278_TASIKMALAYA
3201_BOGOR	8.922	11.578	8.657	5.223	6.900	10.476	11.917	10.979	10.084	11.880
3202_SUKABUMI	6.292	9.967	7.817	6.817	5.278	10.541	10.759	13.221	12.979	9.824
3203_CIANJUR	5.037	8.126	7.698	9.147	7.533	12.424	9.089	15.980	15.881	8.019
3205_GARUT	5.863	7.637	8.773	12.205	10.372	12.363	8.563	16.946	17.109	6.040
3206_TASIKMALAYA	7.074	6.883	9.449	12.688	11.316	12.638	7.456	17.119	17.446	4.396
3207_CIAMIS	7.683	6.084	8.574	12.405	11.057	10.715	6.946	15.414	15.899	4.329
3208_KUNINGAN	8.086	3.905	7.776	12.130	11.435	11.215	5.327	15.706	16.081	5.104
3209_CIREBON	6.507	3.006	6.160	10.809	10.140	10.640	4.677	15.019	15.527	5.166
3211_SUMEDANG	6.293	5.822	7.360	10.659	8.723	7.725	7.482	13.527	14.107	6.067
3212_INDRAMAYU	8.150	6.232	6.286	10.712	10.206	9.927	7.120	13.752	13.897	7.102
3214_PURWAKARTA	6.286	7.449	5.801	7.652	6.529	8.394	8.409	12.176	12.482	8.722
3215_KARAWANG	6.471	7.540	5.481	7.351	7.029	9.040	7.928	12.563	12.804	8.320
3216_BEKASI	8.792	10.933	8.454	5.710	7.123	6.585	9.990	8.183	8.346	10.927
3217_BANDUNG_BARAT	4.878	7.216	6.949	8.778	6.630	8.340	8.525	12.921	13.292	7.451
3218_PANGANDARAN	9.244	8.763	10.116	13.085	12.064	12.635	8.798	16.430	16.480	6.524
3277_CIMAHI	7.803	9.672	9.311	8.885	7.066	3.902	8.472	8.922	10.041	8.997
3279_BANJAR	8.259	6.677	8.724	12.827	11.444	10.724	7.162	15.732	16.339	5.223

Figure 4 Euclidean distance heatmap of West Java consumption

The image above shows the results of data processing using Euclidean distance. The more intense the color produced, the greater the distance between objects. This means that the level of similarity in consumption patterns is also increasingly the same. The same thing is done with calculations using ACF distance. The results are as follows.

	3204_BANDUNG	3210_MAJALENGKA	3213_SUBANG	3271_BOGOR	3272_SUKABUMI	3273_BANDUNG	3274_CIREBON	3275_BEKASI	3276_DEPOK	3278_TASIKMALAYA
3201_BOGOR	0,84	0,55	0,5	0,5	0,62	0,96	0,39	1,2	1,24	0,57
3202_SUKABUMI	0,71	0,57	0,52	0,66	0,6	1,08	0,47	1,36	1,43	0,48
3203_CIANJUR	0,3	0,63	0,68	1,3	1,23	1,75	0,77	2,06	2,13	0,61
3205_GARUT	0,36	0,52	0,6	1,14	1,09	1,59	0,66	1,87	1,93	0,4
3206_TASIKMALAYA	0,43	0,52	0,6	1,08	1,04	1,52	0,61	1,8	1,86	0,32
3207_CIAMIS	0,72	0,52	0,55	0,66	0,7	1,14	0,48	1,32	1,38	0,37
3208_KUNINGAN	0,63	0,32	0,47	0,72	0,7	1,13	0,33	1,41	1,46	0,37
3209_CIREBON	0,39	0,17	0,38	0,83	0,78	1,26	0,32	1,59	1,63	0,32
3211_SUMEDANG	0,84	0,6	0,67	0,61	0,52	0,93	0,62	1,19	1,24	0,62
3212_INDRAMAYU	1,16	0,89	0,98	0,95	0,91	1,17	0,87	1,35	1,4	0,86
3214_PURWAKARTA	0,77	0,55	0,59	0,87	0,79	1,15	0,58	1,49	1,55	0,6
3215_KARAWANG	0,31	0,4	0,48	1,07	1,04	1,48	0,59	1,82	1,89	0,49
3216_BEKASI	1,17	0,94	0,88	0,5	0,74	0,81	0,85	1,08	1,01	0,97
3217_BANDUNG BARAT	0,66	0,6	0,59	0,76	0,7	1,18	0,58	1,44	1,48	0,48
3218_PANGANDARAN	0,59	0,49	0,55	0,9	0,86	1,35	0,53	1,56	1,64	0,3
3277_CIMAHI	1,6	1,41	1,39	0,95	0,81	0,71	1,3	1,1	1,11	1,38
3279_BANJAR	0,53	0,4	0,4	0,78	0,77	1,21	0,39	1,5	1,56	0,32

Figure 5 Heatmap ACF distance resulting from processing

The image above is the result of calculations using ACF distance. When compared with the results from the Euclidean distance, there are differences in the position of the sister cities. Bogor districts, which originally had a sister city, Bogor City, became Cirebon City. Sukabumi districts, which has a sister city, Sukabumi City, became Cirebon City. Sumedang districts, which initially had the sister city Majalengka, became Sukabumi City. Several other Districts can be depicted on the heatmap. Furthermore, there is also a distance measure using the CID distance as follows.

	3204_BANDUNG	3210_MAJALENGKA	3213_SUBANG	3271_BOGOR	3272_SUKABUMI	3273_BANDUNG	3274_CIREBON	3275_BEKASI	3276_DEPOK	3278_TASIKMALAYA
3201_BOGOR	14,92	13,62	10,48	5,6	8,89	15,47	12,56	11,94	10,75	13,47
3202_SUKABUMI	8,18	10,89	8,3	8,18	5,29	12,11	13,12	15,62	17,79	11,14
3203_CIANJUR	5,28	12,11	11,15	14,96	10,25	14,75	15,11	25,75	29,69	12,4
3205_GARUT	8,18	7,78	8,86	13,65	11,15	15,24	9,73	18,66	21,86	6,38
3206_TASIKMALAYA	10,73	7,35	10,37	13,05	13,23	16,93	7,79	17,34	20,5	4,52
3207_CIAMIS	12,71	7,08	10,27	13,15	14,1	15,66	7,25	16,59	17,14	4,86
3208_KUNINGAN	14,67	4,98	10,21	14,1	15,98	17,96	6,09	18,53	16,36	6,28
3209_CIREBON	9,09	3,06	6,23	12,08	10,91	13,12	5,31	16,53	19,83	5,45
3211_SUMEDANG	7,96	6,54	8,04	13,15	8,95	8,63	9,38	16,43	19,88	7,07
3212_INDRAMAYU	11	6,56	6,44	12,39	10,61	11,83	8,37	15,66	18,36	7,76
3214_PURWAKARTA	6,36	10,71	8,11	12,08	8,57	9,61	13,49	18,93	22,51	13,01
3215_KARAWANG	6,7	11,1	7,84	11,88	9,45	10,6	13,02	20	23,65	12,71
3216_BEKASI	10,18	13,42	10,09	7,7	7,99	6,73	13,69	10,87	12,86	13,92
3217_BANDUNG BARAT	5,15	9,72	9,1	12,98	8,15	8,95	12,81	18,82	22,46	10,41
3218_PANGANDARAN	15,6	10,4	12,35	14,15	15,68	18,82	9,36	18,03	17,42	7,46
3277_CIMAHI	8,11	13,22	12,38	13,34	8,82	4,25	12,92	13,19	17,22	12,76
3279_BANJAR	15,74	8,95	12,03	15,66	16,8	18,05	8,61	19,5	17,46	6,75

Figure 6 CID distance heatmap resulting from processing

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

In the image above, when compared with the Euclidean and ACF heatmaps, they have the same pattern. The signals given by non-SBH sample districts / cities tend to have a thick color in the SBH sample districts / cities, namely Bekasi City and Depok City. The three pictures explain that of the 10 inflation districts / cities in West Java, Bekasi City and Depok City, the level of similarity in the SBH non-sample districts / cities data patterns is quite far. This means that these two cities cannot be used as a reference as sister cities. If you look at their position, the two cities tend to follow the consumption patterns in DKI Jakarta Province. However, geographically, it is in the West Java region.

Table 3 Results of Sister City Selection for each Distance Measure

Non-SBH Districts / Cities	***EUC Distance***	***ACF Distance***	***CID Distance***
3201_BOGOR	3271_BOGOR	3274_CIREBON	3271_BOGOR
3202_SUKABUMI	3272_SUKABUMI	3274_CIREBON	3272_SUKABUMI
3203_CIANJUR	3204_BANDUNG	3204_BANDUNG	3204_BANDUNG
3205_GARUT	3204_BANDUNG	3204_BANDUNG	3278_TASIKMALAYA
3206_TASIKMALAYA	3278_TASIKMALAYA	3204_BANDUNG	3278_TASIKMALAYA
3207_CIAMIS	3278_TASIKMALAYA	3278_TASIKMALAYA	3278_TASIKMALAYA
3208_KUNINGAN	3210_MAJALENGKA	3210_MAJALENGKA	3210_MAJALENGKA
3209_CIREBON	3210_MAJALENGKA	3210_MAJALENGKA	3210_MAJALENGKA
3211_SUMEDANG	3210_MAJALENGKA	3272_SUKABUMI	3210_MAJALENGKA
3212_INDRAMAYU	3213_SUBANG	3278_TASIKMALAYA	3210_MAJALENGKA
3214_PURWAKARTA	3213_SUBANG	3210_MAJALENGKA	3204_BANDUNG
3215_KARAWANG	3213_SUBANG	3204_BANDUNG	3204_BANDUNG
3216_BEKASI	3271_BOGOR	3272_SUKABUMI	3273_BANDUNG
3217_BANDUNG_BARAT	3204_BANDUNG	3278_TASIKMALAYA	3204_BANDUNG
3218_PANGANDARAN	3278_TASIKMALAYA	3278_TASIKMALAYA	3278_TASIKMALAYA
3277_CIMAHI	3273_BANDUNG	3273_BANDUNG	3272_SUKABUMI
3279_BANJAR	3278_TASIKMALAYA	3278_TASIKMALAYA	3278_TASIKMALAYA

The table 3 shows several regencies/cities from SBH non-sample distance (desimilarity) measurements. Each method has gaps related to calculating data patterns. However, there is also a methodology that produces the same sister city. If at least two are the same, then several districts / cities should experience changes in determining sister cities. The districts that should experience change is Cianjur Districts, which was previously Sukabumi City, to become Bandung Districts. Kuningan Districts and Cirebon Districts, which previously had the sister city of Cirebon City, became Majalengka Districts. Kerawang Districts and Bekasi Districts, which previously had the sister city of Bekasi City, became Bandung Districts and Bandung City. Sumendang Districts and West Bandung Districts, which previously had the sister city of Bandung City, became Majalengka Districts and Bandung Districts.

Policies related to consumption patterns in West Java as a result of inflation, carried out further analysis by forming clusters related to consumption patterns. Data processing in this research used R version 4.3.2. to form groupings in the form of biclusters. The process of forming a bicluster begins with tuning the parameters by determining the value of δ. then the best parameters will be seen at the smallest MSR/V value.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

Table 4 Results of parameter tuning (Parameter Tuning)

δ	Bicluster	MSR/V
0.1	**3**	**0.028433**
0.2	2	0.032765
0.3	2	0.057901
0.4	1	0.047549
0.5	1	0.00000
0.6	1	0.00000
0.7	1	0.00000
0.8	1	0.00000
0.9	1	0.00000
1	1	0.00000

The table 4 shows that there is no linear relationship between MSR/V and the parameter tuning results. The table shows that gamma 0.1 is the smallest value produced, so it can be said to be the best grouping.

Table 5 Number of Provinces and Variables from the best bicluster

Bicluster	Number of Provinces	Number of Variables
1	19	12
2	6	11
3	2	16

Number of Variables The table 5 shows the number of best biclusters, namely in bicluster 1, the number of groups and the number of variables that form the groups. Characteristic variables are identical variables from a bicluster. All Districts/cities in the best bicluster can enter the group that has been formed. Following are the details of the grouping.

Table 6 Details of Characteristic Variables for Each Bicluster Member

***Bicluster* 1**	***Bicluster* 2**	***Bicluster* 3**
Variable: **X3 X4 X5 X6 X8 X9 X11 X12 X15 X16 X17 X19**	Variable: **X3 X4 X5 X6 X7 X8 X9 X11 X15 X16 X19**	Variable: **X1 X3 X4 X5 X6 X7 X8 X9 X10 X14 X15 X16 X17 X18 X19 X20**

The grouping results are shown in the table 6. There are three biclusters with their characteristic variables. These characteristic variables form districts / cities groups based on consumption patterns. Consumption patterns are used to find out which characteristic variables are used as a reference for policy-making regarding fulfilling consumption in society, especially in West Java. The grouping results can be seen in the following image.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

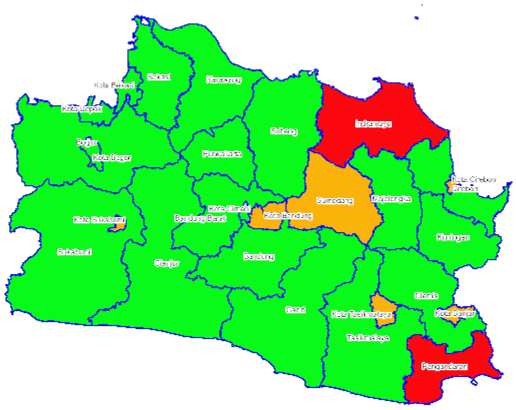

Figure 7 Clustering map based on the bicluster algorithm

The formation process can be seen visually on the heatmap between the variables and the research object. The darker the color produced, the stronger the resulting relationship. The characteristic variables are fish, meat, eggs, vegetables, fruit, coconut oil, housing, various goods and services, as well as taxes, levies and insurance. These eight variables are variables that appear in each group and characterize the bicluster that is formed.

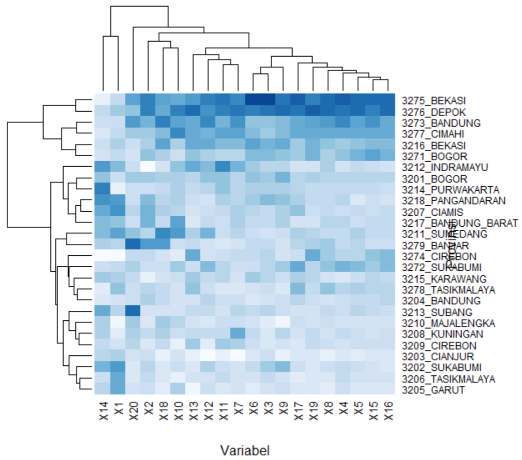

Figure 8 Heatmap of consumption pattern variables in

The image 8 shows the signals that arise from the research object and the variables that will be grouped. The hope is that by looking at this picture, there will be a need to consider policies related to consumption patterns in each group, as depicted in the spider chart below.

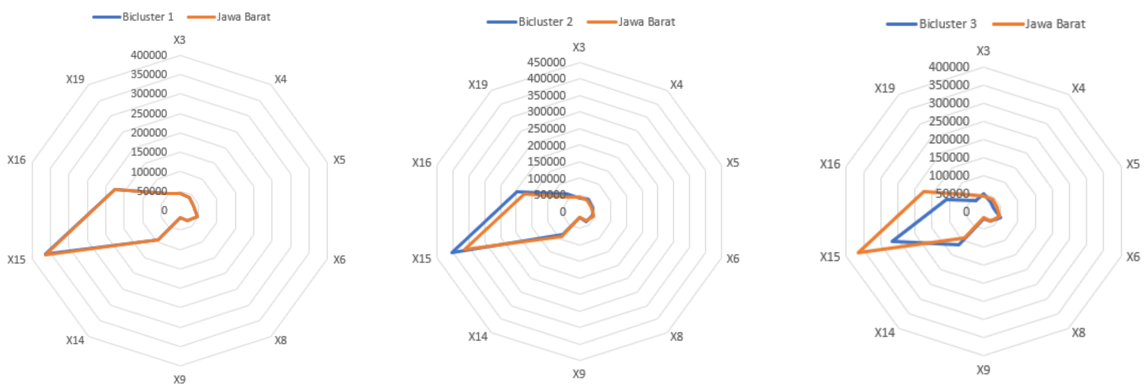

Figure 9 Best Bicluster Radar Chart

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

The spider chart above shows that bicluster 1 is a reflection of consumption patterns in West Java. The lines formed on the spider chart coincide with the lines formed by the conditions in West Java. Bicluster 2 shows a pattern that tends to be above the West Java average consumption pattern. Meanwhile, bicluster 3 shows that consumption patterns tend to be below the average for West Java.

5. CONCLUSION

The conclusions that can be drawn from this research are that first, determining sister cities can consider the use of consumption pattern variables, population, GRDP per capita, and distance effects. The two machine learning methods, namely Euclidean distance, CID distance, and ACF distance, can be used to determine sister cities. Still, the cultural consumption patterns of non-sample Districts/cities and SBH samples must be considered. The three distance measurements found a fairly high gap between non-SBH sample Districts/cities and Bekasi City and Depok City as SBH sample cities, so these two cities should not be used as a reference to become sister cities in West Java. The fourth analysis to obtain characteristic variables using the biclustering method shows that Districts/cities in West Java have three biclustering groups with average consumption patterns, where grouping using biclustering can group Districts/cities based on variable characteristics, namely biclusters with the same characteristics as average. West Java average, bicluster below the West Java average, and bicluster above the West Java average. Characteristic variables that can be used as regional dividers from biclusters are Average Fish Consumption (X3), Average Meat Consumption (X4), Average Egg and Milk Consumption (X5), Average Vegetable Consumption (X6), Average -Average Consumption of Fruit (X8), Average Consumption of Oil and Coconut (X9), Average Consumption of Housing and Household Facilities (X15), Average Consumption of Various Goods and Services and Average Consumption of Taxes (X16), Levy and Insurance (X19).

ACKNOWLEDGMENT

This research was conducted because of scientific writing training held by PUSDIKLAT BPS. Thank you for holding activities that are very useful for your capacity as a civil servant in the future.

REFERENCES

[1] G. Strasser, T. Messner, F. Rumler, and M. Ampudia, “Inflation heterogeneity at the household level,” *ECB Occas. Pap.*, no. 2023/325, 2023.

[2] S. Sarbaini and N. Nazaruddin, “Pengaruh Kenaikan BBM Terhadap Laju Inflasi di Indonesia,” *J. Teknol. Dan Manaj. Ind. Terap.*, vol. 2, no. I, pp. 25–32, 2023.

[3] A. M. Taylor and M. P. Taylor, “The purchasing power parity debate,” *J. Econ. Perspect.*, vol. 18, no. 4, pp. 135–158, 2002.

[4] S. Indonesia, “Encyclopedia of Social and Economic Indicators,” Statistics Indonesia.

[5] A. B. Santosa, “Analisis Inflasi di Indonesia,” 2017.

[6] A. Adji, T. Hidayat, H. Tuhiman, S. Kurniawati, and A. Maulana, “Pengukuran Garis Kemiskinan di Indonesia: Tinjauan Teoretis dan Usulan Perbaikan,” *Jakarta Tim Nas. Percepatan Penanggulangan Kemiskin.*, 2020.

[7] P. D. Pickupana, P. H. P. Jati, and M. Sukin, “Penentuan Sister City Untuk Pembentukan Diagram Timbang Di Nusa Tenggara Timur Dengan Algoritma K-Means,” *J. Stat.*

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

Terap. (ISSN 2807-6214), vol. 1, no. 2, pp. 14–24, 2021.

[8] A. Fadlurohman and T. W. Utami, “Pemodelan Generalized Space Time Autoregressive With Variable Exogenous (Gstar-X) Pada Inflasi Enam Kota Survei Biaya Hidup Di Jawa Tengah,” in *Prosiding Seminar Nasional Indonesian R Summit*, 2020.

[9] A. Mahendra, “Analisis Pengaruh Pertumbuhan Ekonomi, Pendapatan Perkapita, Inflasi Dan Pengangguran Terhadap Jumlah Penduduk Miskin Di Provinsi Sumatera Utara,” *J. Ris. Akunt. Keuang.*, pp. 123–148, 2016.

[10] A. DURROTUSSA’ADAH, “Pembangunan Sistem Pakar untuk Manajemen Pengetahuan pada Kegiatan Peninjauan dan Pengeditan Data Survei Biaya Hidup.” Program Studi Komputasi Statistik Program Diploma IV, 2016.

[11] E. Pujiati, D. Yuniarti, and R. Goejantoro, “Peramalan Dengan Menggunakan Metode Double Exponential Smoothing Dari Brown,” *J. Eksponensial*, vol. 7, no. 1, pp. 33–40, 2017.

[12] A. Harumeka, “Pemanfaatan Data Survei Sosial Ekonomi Nasional Untuk Memilih Sister City Pada Kabupaten/Kota Non-Sampel Survei Biaya Hidup Di Jawa Timur: Utilization of National Socio-Economic Survey Data to Select Sister Cities in Non-Sample Districts/Cities of the Cos,” *J. Ilm. Komputasi dan Stat.*, vol. 2, no. 2, pp. 26–31, 2023.

[13] T. W. Liao, “Clustering of time series data—a survey,” *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, 2005.

[14] A. A. Mattjik, I. Sumertajaya, G. N. A. Wibawa, and A. F. Hadi, “Sidik peubah ganda dengan menggunakan SAS.” 2011.

[15] A. Budi and H. Maulana, “Pengenalan Citra Wajah Sebagai Identifier Menggunakan Metode Principal Component Analysis (PCA),” *J. Tek. Inform.*, vol. 9, no. 2, 2016.

[16] P. Galeano, D. Peña, and R. S. Tsay, “Outlier detection in multivariate time series by projection pursuit,” *J. Am. Stat. Assoc.*, vol. 101, no. 474, pp. 654–669, 2006.

[17] R. Ardalova, “Analisis Harga Cabai Merah Besar di Pasar Eceran Jakarta dengan Menggunakan Analisis Gerombol Deret Waktu”.

[18] R. M. KUSAIRI, “An Improved Biclustering Algorithm With Overlapping Control For Identification Of Informative Genes And Pathways,” 2021.

[19] S. C. Madeira and A. L. Oliveira, “Biclustering algorithms for biological data analysis: a survey,” *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 1, no. 1, pp. 24–45, 2004.

[20] S. Dutta, M. Hore, F. Ahmad, A. Saba, M. Kumar, and C. Das, “SBi-MSREimpute: A Sequential Biclustering Technique Based on Mean Squared Residue and Euclidean Distance to Predict Missing Values in Microarray Gene Expression Data,” in *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 2*, Springer, 2019, pp. 673–685.

[21] L. A. Mina and G. W. Sledge Jr, “Rethinking the metastatic cascade as a therapeutic target,” *Nat. Rev. Clin. Oncol.*, vol. 8, no. 6, pp. 325–332, 2011.

[22] D. J. Dhandio, M. A. Simanjuntak, S. Martha, and S. Supandi, “Peramalan Inflasi Kota Pontianak dengan Metode Double Exponential Smoothing: Pontianak City Inflation Forecasting Using the Double Exponential Smoothing Method,” in *Jurnal Forum Analisis Statistik (FORMASI)*, 2023, pp. 51–66.

[23] Ł. Pawluczuk and M. Iskrzyński, “Food web visualisation: Heat map, interactive graph and animated flow network,” *Methods Ecol. Evol.*, vol. 14, no. 1, pp. 57–64, 2023.

[24] M. Kirişci, “New cosine similarity and distance measures for Fermatean fuzzy sets and TOPSIS approach,” *Knowl. Inf. Syst.*, vol. 65, no. 2, pp. 855–868, 2023.

[25] V. A. P. Sangga, “Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015,” 2018.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]

[26] B. P. Statistik, “Data dan Informasi kemiskinan kabupaten/kota tahun 2018,” *Jakarta Badan Pus. Stat.*, 2019.
[27] S. Sudirman, I. G. Indradi, A. Sriyono, and A. Prayitno, “Analisis Determinan Dan Program Pengentasan Kemiskinan Rumahtangga Petani Dalam Rangka Mendukung Arahan Kebijakan Reforma Agraria Dalam Mengatasi Kemiskinan Petani (Studi Di Desa Bogem Kecamatan Bayat, Klaten Jawa Tengah),” 2011.
[28] B. Pudjianto and M. Syawie, “Kemiskinan dan pembangunan manusia,” *Sosio Inf. Kaji. Permasalahan Sos. dan Usaha Kesejaht. Sos.*, vol. 1, no. 3, 2015.
[29] R. F. Pratama, “Menerapkan Algoritma Support Vector Machine (SVM) di Klasifikasi Masyarakat Tanjung Lowland di Lampung Timur,” *J. Portal Data*, vol. 2, no. 10, 2022.
[30] M. Raymond, “Ilmu Peluang dan statistika untuk insinyur dan ilmuwan,” 2016.

[DOI: 10.14710/JSUNIMUS.12.1.2024.10-23]