ELINVO (Electronics, Informatics, and Vocational Education), Mei 2023; vol 8(1):86-95
ISSN 2580-6424 (printed), ISSN 2477-2399 (online) DOI: https://doi.org/10.21831/elinvo.v8i1.53208

Soybean Collect Recommender Based on Distance and Productivity Cluster Using K-means Clustering and Simple Addictive Weighting Method

Mega Wahyu Ningtyas[1], Feddy Setio Pribadi[2*]
[1,2]Electronics Engineering Department, Universitas Negeri Semarang
[*]E-mail: feddy.setio@mail.unnes.ac.id

ABSTRACT

Soybeans are an essential agricultural product that is one of the primary food sources in Indonesia, such as tempeh, tofu, soy milk, soy sauce, and other preparations. However, production yields, harvested land area, and soybean productivity in each district or city in Central Java Province vary widely. Differences in soybean productivity in each area are due to production factors such as area, use of fertilizers, seeds, and labor. This study tries mapping the soybean-producing areas based on the distance and productivity of an area using K-means clustering, the haversine formula, and the simple additive weighting method. This study proposes a method for the soybean processing business or customers in places that lack or do not grow soybeans to buy directly from collectors in other areas with strong soybean production and a short distance away. According to the research, four clusters have formed: the first has five members, the second has fourteen, the third has nine, and the fourth has seven. The fourth cluster, which consists of seven members who do not grow soybeans, is advised to take soybeans from the following regions: Kendal Regency, Klaten Regency, Magelang Regency, Batang Regency, and Brebes Regency.

Keywords*:* cluster, K-Means, regency, Simple Additive Weighting, soybean

INTRODUCTION

In Indonesia, soybeans are the primary source of raw material for foods with vegetable protein, including tempeh, tofu, soy milk, soy sauce, and other concoctions [1]. Therefore, the consumption of tofu and tempeh in Indonesia is enormous. For example, Badan Pusat Statistik (BPS) Indonesia recorded that the average consumption of tofu in 2021 per capita is 0.158 kg per week. In the meantime, the standard per capita utilization of tempeh was 0.146 kg per week in 2021, which appears that the standard per capita utilization of tofu in 2021 is 7.58 kg, and tempeh is 4.38. A large amount of consumption of processed foods derived from soybeans has an impact on the increasing demand for soybeans in Indonesia. However, soybean needs are not in line with soybean production in Indonesia, and soybean production in Indonesia cannot meet the high consumer demands [2]. According to an article by Madrim [3,] which discusses the need for soybeans in Indonesia being met through imports, the total demand for national soybeans in 2021 is around 2 million tons, while the national soybean production is only around 240 thousand tons. Therefore, the Indonesian government made efforts to import soybeans from abroad to meet the needs of national soybean consumption [4].

Indonesian soybean imports have increased yearly to meet soybean consumption needs [5]. According to an article by Annur [6], soybean imports reached 2.48 million tons in 2021. That number is up 0.58% compared to the previous year, which was 2.47 million tons. The value of soybean imports also increased by 47.77% from the previous year of US$ 1 billion to US$ 1.48 billion in 2021. The number of soybean imports significantly impacts market conditions; the low border price causes the market to form low expectations, lowering farmer prices and harming farmers [7]. When soybean prices in the international market are high, farmers can benefit. Otherwise, soybean farmers will lose if soybean prices in the international market decrease. This situation makes domestic products, especially soybeans, unable to compete, and farmers are reluctant to grow soybeans.

One of the soybean-producing areas in Indonesia is Central Java Province. Central Java Province is Indonesia's second-largest soybean production center [8]. In 2019 Central Java produced 64,334 tons of soybeans with an area of 37,944 ha of harvested land and productivity of 16.95 quintals/ha. Therefore, central Java can potentially be developed as a center for domestic soybean production to reduce Indonesia's dependence on imported soybeans. However, according to BPS data, soybean productivity results in each region in Central Java Province vary greatly; some areas do not produce soybeans. In addition, soybean productivity varies by region due to production factors such as area, fertilizer, seed, and labor use [9].

Therefore, a suitable distribution pattern for soybeans will be necessary to equally meet everyone's soybean needs, especially for regions that do not grow soybeans. Furthermore, it is essential to close the distance between soybean commodity centers and industrial centers [10]. The correct distribution patterns help farmers distribute soybeans to consumers who are unable to reach them. Marketing activities to deliver producer commodities to consumers require costs, which impact both the prices paid by consumers and the prices received by producers [11]. The correct distribution patterns will lower expenses incurred by bringing customers closer to producers, ensuring that both consumers and producers profit.

The use of K-means on agricultural data has been carried out by Mashita [12] on soybean data and Sudirman [13] on rice data. This research discusses the implementation of K-means clustering to classify agricultural products. K-means is used as a clustering algorithm to classify agricultural products based on provinces in Indonesia. The result is a picture showing the grouping of provinces based on agricultural production.

The Simple Additive Weighting method was used in a previous study by Romadhona [14] as a decision-making method in the distribution of agricultural assistance. And another study by Yanuari [15] uses Simple Additive Weighting to determine which types of plants are suitable for planting in an area with a certain climate.

Dauni [16] and Husada [17] have used the Haversine Formula to calculate distances on Earth. The Haversine formula is used to find the shortest distance between a user's current location and the nearest school [16]. Meanwhile, [17] used the haversine formula to locate the nearest hospital.

Therefore, this study tries mapping the soybean-producing areas in Central Java based on the distance and productivity of an area using the K-means clustering method, the Haversine Formula, and the Simple Additive Weighting method. The purpose of this research is to map soybean-producing areas and suggest an effective method for getting soybeans from non-producing areas to areas with high soybean productivity in the shortest amount of distance.

METHODS

A. Identification of Soybean Distribution Pattern

There is a certain pattern that has been applied in the community regarding the distribution of soybeans from farmers to the soybean processing industry and consumers. Identification of soybean distribution patterns in this study was conducted to provide an overview of the soybean distribution channel from producers to consumers or the soybean processing industry in Central Java. At this stage, the focus is on collecting and analyzing relevant previous studies to get an overview of the characteristics, patterns, and processes that occur in the distribution of soybeans.

B. Data Collection and Preprocessing

There are two data used in this study. The first data was taken from the Central Java Statistics Agency website [18] with the title "Harvest Area, Production, and Productivity of Corn and Soybeans by Regency/City in Central Java Province". Before the data is used, the data goes through the data cleaning stage, which is the stage to clean the data so that only the data needed for research is obtained. At this stage, the data is tidied up, and unnecessary data is removed. The result is 35 data with four attributes. The data consists of district/city, land area (hectare), production (tons), and productivity (quintals/hectare) which will be used in the research.

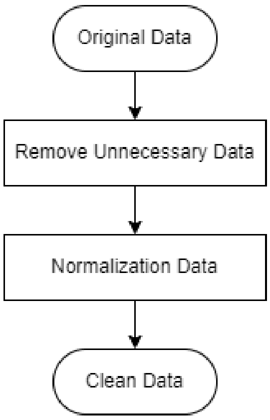

Figure 1. Preprocessing Data

The second data is coordinate point data in the form of the latitude and longitude of each district or city government center in Central Java which is taken manually with the help of google maps. If we right-click on google maps, it will display some information, including latitude and longitude points. The data is obtained by right-clicking on google maps for each district or city government center point in Central Java. The latitude and longitude points are taken and collected in Microsoft Excel software.

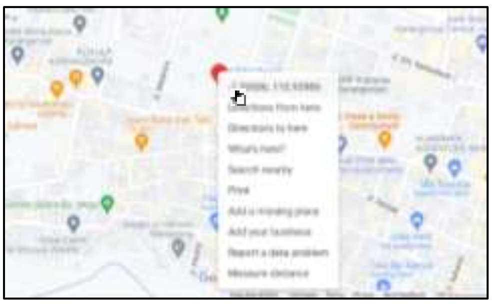

Figure 2. Latitude and Longitude on Google Maps

C. K-means Clustering

The K-means algorithm is included in the partition-based clustering algorithm, which is used to group data by determining the number of k first [19]. K-means is a non-hierarchical clustering method based on the distance that divides data into several clusters and only works on numerical attributes [20][21]. This method attempts to partition existing objects into one or more object clusters based on their characteristics. Cluster similarity is measured by the average value of objects in a cluster, which can be seen as the cluster's center. The K-means algorithm takes k input parameters and partitions a set of n objects into k clusters [22]. Kmeans has been widely used as a grouping tool in various fields. In the health sector, K-means clustering was used in Harliana's research [23] to classify stunting areas, and in the education sector, K-means clustering was used to classify the quality of SMK teachers in Indonesia [24]. The K-means clustering method aims to minimize the objective function in the clustering process by minimizing variations within a cluster and maximizing variations in different clusters.

The stages of clustering using the K-means method are as follows [25]:

1. Set the number of k clusters to be formed
2. Placing data into clusters randomly
3. Using equation (1), calculate the cluster center (centroid) of the data in each cluster, where M is the amount of data in k clusters.

$$C_i = \frac{1}{M}\sum_{j=1}^{m} X_j \quad (1)$$

4. Using the Euclidean formula (2), calculate the distance between each data point and the cluster center (centroid).

$$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \quad (2)$$

5. Repeat step 2 until the members of each cluster have not changed.

The data used in clustering for this research is soybean productivity data. In the K-means method, it is necessary to determine the number of clusters (k). In this study, the number of clusters is set at k=4 to classify data on areas with high, medium, low productivity, and non-productive areas. With the description that k1 is a high-productivity area, k2 is a medium-productivity area, k3 is a low-productivity area, and k4 is an area with no production at all.

D. Haversine Formula

The distance from one point on the Earth's surface to another is affected by curvature [16]. The Haversine Formula uses the length of a straight line between latitude and longitude to calculate the distance between two points [26]. The advantages of the Haversine formula over other geodetic distance calculations are that it is more accurate than methods, has a lower error rate for analysis speed, and is easier to perform [17]. Also, the Haversine formula has its law. All equations used are based on the Earth's spherical shape, ignoring that the Earth is slightly elliptical [16]. Calculation of the haversine formula in equation (3) [26]:

$$d = 2.R.\arcsin\sqrt{sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat2).\cos(lat1).sin^2\left(\frac{\Delta long}{2}\right)} \quad (3)$$

Where:

d = distance (km)
Δlat = latitude2 – latitude1
Δlong = longitude2 – longitude1
R = radius of the Earth (6371 km)

In this study, the latitude and longitude points used are the central points of government for each region, with the calculation that the production of each region is collected in the city center or government center. The data is obtained by manually taking the coordinates of each region's central government with the help of Google Maps. Each address point listed on Google Maps contains latitude and longitude information. Table 1 displays the latitude and longitude data of the central government of each regency or city in Central Java Province.

Table 1. Latitude and Longitude Data for Each Region

Region Name	Latitude	Longitude
Kabupaten Cilacap	-7.724468331	109.0093655
Kabupaten Banyumas	-7.423015171	109.2301023
Kabupaten Purbalingga	-7.387679411	109.3640083
Kabupaten Banjarnegara	-7.396129494	109.696158
Kabupaten Kebumen	-7.668634601	109.6535832
Kabupaten Purworejo	-7.714173773	110.0082884
Kabupaten Wonosobo	-7.35886488	109.9046456
Kabupaten Magelang	-7.592559715	110.2194273
Kabupaten Boyolali	-7.545876791	110.6105477
Kabupaten Klaten	-7.712718706	110.5921512
Kabupaten Sukoharjo	-7.664368536	110.8360201
Kabupaten Wonogiri	-7.815478455	110.925845
Kabupaten Karanganyar	-7.595822197	110.9398767
Kabupaten Sragen	-7.42668696	111.0231924
Kabupaten Grobogan	-7.082050575	110.9179295
Kabupaten Blora	-6.970024142	111.4148092
Kabupaten Rembang	-6.703920363	111.3425043
Kabupaten Pati	-6.752594305	111.040085
Kabupaten Kudus	-6.806380585	110.8414313
Kabupaten Jepara	-6.590301358	110.6680674
Kabupaten Demak	-6.891590979	110.6383948
Kabupaten Semarang	-7.129175341	110.4040973
Kabupaten Temanggung	-7.315542571	110.1818715
Kabupaten Kendal	-6.922974009	110.2033029
Kabupaten Batang	-6.910795003	109.7297392
Kabupaten Pekalongan	-7.025026642	109.5911614
Kabupaten Pemalang	-6.893389684	109.3806242
Kabupaten Tegal	-6.995668619	109.127759
Kabupaten Brebes	-6.870721333	109.0392465
Kota Magelang	-7.504075916	110.2213583
Kota Surakarta	-7.569454176	110.8293366
Kota Salatiga	-7.330775634	110.500604
Kota Semarang	-6.982089462	110.4127213
Kota Pekalongan	-6.897000316	109.6620809
Kota Tegal	-6.8703746	109.1374548

E. Simple Additive Weighting Method (SAW)

The Simple Additive Weighting (SAW) method, also known as the weighted summing method, is one of many methods used to assist with decision-making problems [27]. The basic idea of this method is to get the total weight of the performance evaluation of each alternative overall attribute [28]. Furthermore, the SAW method needs to normalize the decision matrix to a comparable scale to all available alternative ratings [29]. The stage of the SAW method is as follows [30][31]:

1. Set the criteria that will be used as a basis for decision-making. The SAW method's

criteria are divided into benefit and cost. Benefit criteria category, if the criterion has a more excellent value, the better, while the lower cost criterion, the better. For example, this study uses cluster members and distance as criteria, and both are cost criteria.

2. Set the appropriate rating for each alternative for each attribute
3. Normalization of the criteria matrix is based on the equation adjusted to the type of attribute (benefit or cost attribute) to obtain a normalized matrix. The matrix normalization calculation is shown in equation (4)

$$r_{ij} = \begin{cases} \frac{x_{ij}}{max_{(xij)}} & j \text{ is benefit attribut} \\ \frac{min_{xij}}{x_{ij}} & j \text{ is cost attribut} \end{cases} \quad (4)$$

With:
r_{ij} = normalized performance rating score
x_{ij} = attribute value of each criterion
max x_{ij} = the largest value of each criterion
min x_{ij} = the smallest value of each criterion

4. The result is obtained from the matrix multiplication ranking process, which is the sum of normalized r with a weight vector to obtain the largest value as the best solution. Calculating the last alternative value can be done using equation (5)

$$V_i = \sum_{j=1}^{n} W_j \, r_{ij} \quad (5)$$

With:
V_i = rating for each alternative
W_j = weighting value of each criterion

A larger V_i value indicates that the alternative is preferred

RESULT AND DISCUSSION

In this study, the identification of soybean distribution patterns was carried out by analyzing relevant previous studies to get an overview of the characteristics, patterns, and processes that occur in the distribution of soybeans. Research conducted by Fiona, Soetoro, and Normansyah [32] in the Ciamis district, West Java revealed that soybean distribution channels were carried out by farmers, collectors, wholesalers, and industrial consumers. The distribution pattern can be seen in the following figure:

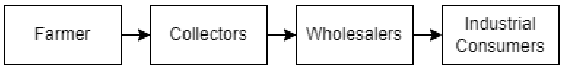

Figure 3. Soybeans Distribution Pattern 1

Another research done in the Grobogan district, Central Java by Kristanti and Almuntaka [33] indicated that there were different distribution patterns of soybeans, with the most common being from farmers, collectors, cooperatives, wholesalers, small dealers, and consumers. The following diagram illustrates the soybean distribution flow:

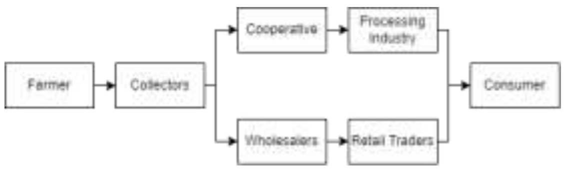

Figure 4. Soybeans Distribution Pattern 2

This study tries mapping the soybean-producing areas in Central Java based on the distance and productivity of an area using the K-means clustering method, the Haversine Formula, and the Simple Additive Weighting method. The K-means method is used to classify regions of Central Java based on soybean productivity. With the number of the cluster (k) = 4, where k1 is a highly productive region, k2 is a medium productive region, k3 is a low productive region, and k4 is a shallow productive region. The K-means clustering method is used because, in K-means clustering, an area that belongs to a group with high, low, or medium productivity is clearly defined compared to the usual ordering method. With K-means clustering, an area belonging to a specific group is determined by measuring the distance between the data and the center of each group; therefore, a more definite grouping is obtained—the K-means method groups data by minimizing variations within a cluster and maximizing variations in different clusters. The results of clustering can be seen in Table 2.

Table 2. Clustering Results

Region Name	Land Area (ha)	Production (ton)	Productivity (quintal/ha)
K1			
Kabupaten Purworejo	640	1277	19.95
Kabupaten Klaten	1563	3410	21.82
Kabupaten Blora	1264	2940	23.26
Kabupaten Kendal	723	1466	20.29
Kabupaten Brebes	1607	3537	22.00
Total	**5797**	**12630**	**107.32**
K2			
Kabupaten Purbalingga	1656	3102	18.73
Kabupaten Banjarnegara	491	852	17.35
Kabupaten Wonosobo	25	40	16.21
Kabupaten Magelang	4	7	17.81
Kabupaten Boyolali	695	1131	16.28
Kabupaten Karanganyar	277	488	17.61
Kabupaten Sragen	1665	2772	16.65
Kabupaten Grobogan	7495	13961	18.63
Kabupaten Kudus	264	464	17.54
Kabupaten Jepara	52	91	17.33
Kabupaten Demak	3318	6235	18.79
Kabupaten Pekalongan	200	337	16.86
Kabupaten Pemalang	195	344	17.63
Kabupaten Tegal	73	130	17.77
Total	**16410**	**29954**	**245.19**
K3			
Kabupaten Cilacap	4941	6526	13.21
Kabupaten Banyumas	817	1037	12.69
Kabupaten Kebumen	1712	2210	12.91
Kabupaten Sukoharjo	1101	1690	15.36
Kabupaten Wonogiri	1933	2874	14.87
Kabupaten Rembang	2135	2948	13.81
Kabupaten Pati	1900	2781	14.64
Kabupaten Semarang	93	124	13.38
Kabupaten Batang	1106	1559	14.10
Total	**96621**	**21749**	**124.97**
K4			
Kabupaten Temanggung	-	-	-
Kota Magelang	-	-	-
Kota Surakarta	-	-	-
Kota Salatiga	-	-	-
Kota Semarang	-	-	-
Kota Pekalongan	-	-	-
Kota Tegal	-	-	-
Total	-	-	-

The clustering results using the K-means method show that five regions, namely Purworejo Regency, Klaten Regency, Blora Regency, Kendal Regency, and Brebes Regency, are included in cluster 1, which means that the area has high soybean productivity. Cluster 2 areas that have sufficient or moderate soybean productivity are Purbalingga Regency, Banjarnegara Regency, Wonosobo Regency, Magelang Regency, Boyolali Regency, Karanganyar Regency, Sragen Regency, Grobogan Regency, Kudus Regency, Jepara Regency, Demak Regency, Pekalongan Regency, Pemalang Regency, and Tegal Regency. Nine areas included in cluster 3 are Cilacap Regency, Banyumas Regency, Kebumen Regency, Sukoharjo Regency, Wonogiri Regency, Rembang Regency, Pati Regency, Semarang Regency, and Batang Regency. Meanwhile, Temanggung Regency, Magelang City, Surakarta City, Salatiga City, Semarang City, Pekalongan City, and Tegal City are included in cluster 4, which means that the seven regions have very low soybean productivity or no production at all.

Cluster 4 members are advised to take soybeans from clusters 1, 2, or 3 with the closest distance, as calculated by the SAW method. In the calculation with the SAW method, the number of cluster members in each region weighs 50%, and the distance between regions weighs 50%. This weight choice is because it is assumed that both the amount of productivity and the distance from the area are equally crucial for the availability of soybeans. Distance and soybean productivity are used as decision-making parameters in the SAW method decision-making system because soybean availability necessitates high productivity and short distances. High soybean productivity in an area is assumed to have a good amount of soybean production compared to a large area of land. While the short distance can reduce transportation costs, it does not affect prices drastically. The results of the calculation of the SAW method can be seen in Table 3 and Table 4.

Table 3. SAW Method Calculation Results

Name of Region from Cluster Results	Kabupaten Temanggung	Kota Magelang	Kota Surakarta
K1			
Kabupaten Purworejo	0.821	0.649	0.558
Kabupaten Klaten	0.745	0.605	0.673
Kabupaten Blora	0.609	0.534	0.557
Kabupaten Kendal	0.854	0.576	0.553
Kabupaten Brebes	0.614	0.533	0.525
K2			
Kabupaten Purbalingga	0.421	0.302	0.282
Kabupaten Banjarnegara	0.535	0.333	0.292
Kabupaten Wonosobo	0.750	0.378	0.301
Kabupaten Magelang	0.748	0.750	0.329
Kabupaten Boyolali	0.538	0.364	0.468
Kabupaten Karanganyar	0.423	0.312	0.672
Kabupaten Sragen	0.415	0.305	0.449
Kabupaten Grobogan	0.432	0.305	0.346
Kabupaten Kudus	0.418	0.298	0.312
Kabupaten Jepara	0.410	0.294	0.298
Kabupaten Demak	0.474	0.310	0.318
Kabupaten Pekalongan	0.463	0.306	0.285
Kabupaten Pemalang	0.405	0.293	0.280
Kabupaten Tegal	0.377	0.287	0.277
K3			
Kabupaten Cilacap	0.280	0.203	0.193
Kabupaten Banyumas	0.313	0.212	0.197
Kabupaten Kebumen	0.387	0.242	0.207
Kabupaten Sukoharjo	0.356	0.237	0.667
Kabupaten Wonogiri	0.323	0.225	0.347
Kabupaten Rembang	0.273	0.199	0.214
Kabupaten Pati	0.303	0.207	0.223
Kabupaten Semarang	0.649	0.273	0.245
Kabupaten Batang	0.397	0.224	0.204

Table 4. SAW Method Calculation Results

Name of Region from Cluster Results	Kota Salatiga	Kota Semarang	Kota Pekalongan	Kota Tegal
K1				
Kabupaten Purworejo	0.680	0.588	0.539	0.540
Kabupaten Klaten	0.784	0.598	0.528	0.529
Kabupaten Blora	0.614	0.574	0.520	0.522
Kabupaten Kendal	0.722	0.841	0.564	0.546
Kabupaten Brebes	0.573	0.554	0.555	1.000
K2				
Kabupaten Purbalingga	0.349	0.316	0.310	0.336
Kabupaten Banjarnegara	0.389	0.340	0.319	0.314
Kabupaten Wonosobo	0.439	0.367	0.316	0.304
Kabupaten Magelang	0.542	0.365	0.289	0.288
Kabupaten Boyolali	0.713	0.373	0.280	0.280
Kabupaten Karanganyar	0.469	0.341	0.274	0.275
Kabupaten Sragen	0.462	0.348	0.274	0.275
Kabupaten Grobogan	0.481	0.394	0.277	0.277
Kabupaten Kudus	0.429	0.410	0.279	0.279
Kabupaten Jepara	0.397	0.408	0.283	0.282
Kabupaten Demak	0.493	0.555	0.285	0.283
Kabupaten Pekalongan	0.367	0.340	0.485	0.352
Kabupaten Pemalang	0.343	0.322	0.373	0.451
Kabupaten Tegal	0.330	0.308	0.314	0.638
K3				
Kabupaten Cilacap	0.240	0.213	0.199	0.223
Kabupaten Banyumas	0.255	0.225	0.217	0.254
Kabupaten Kebumen	0.290	0.239	0.211	0.218
Kabupaten Sukoharjo	0.404	0.259	0.191	0.193
Kabupaten Wonogiri	0.340	0.242	0.189	0.191
Kabupaten Rembang	0.273	0.243	0.187	0.189
Kabupaten Pati	0.308	0.278	0.192	0.192
Kabupaten Semarang	0.667	0.667	0.211	0.205
Kabupaten Batang	0.295	0.275	0.667	0.249

As a result of the calculation, cluster 4 members are advised to take soybeans from clusters 1, 2, or 3, as shown in Table 5. The higher the calculated number, the higher the ranking; contrary, the lower the calculated number, the lower the ranking. Temanggung Regency is encouraged to bring soybeans from Kendal, Purworejo, Wonosobo, Magelang, and Klaten Regencies. Magelang City is recommended for bringing soybeans from Magelang Regency, Purworejo Regency, Klaten Regency, Kendal Regency, and Blora Regency. It is suggested that soybeans be brought to Surakarta City from Klaten Regency, Karanganyar Regency, Sukoharjo Regency, Purworejo Regency, and Blora Regency.

Salatiga Town It is advised to bring soybeans from Klaten Regency, Kendal Regency, Boyolali Regency, Purworejo Regency, and Semarang Regency. Semarang City is advised to bring soybeans from the following regencies: Kendal, Semarang, Klaten, Purworejo, and Blora. Pekalongan City is advised to bring soybeans from the Batang, Kendal, Brebes, Purworejo, and Klaten regencies. Meanwhile, Tegal City is advised to bring soybeans from the following regencies: Brebes, Tegal, Kendal, Purworejo, and Klaten.

Table 5. SAW Results Ranking

Region Without Soybean Production	Soybean Producing Region	Cluster	Rank
Kabupaten Temanggung	Kabupaten Kendal	1	1
	Kabupaten Purworejo	1	2
	Kabupaten Wonosobo	2	3
	Kabupaten Magelang	2	4
	Kabupaten Klaten	1	5
Kota Magelang	Kabupaten Magelang	2	1
	Kabupaten Purworejo	1	2
	Kabupaten Klaten	1	3
	Kabupaten Kendal	1	4
	Kabupaten Blora	1	5
Kota Surakarta	Kabupaten Klaten	1	1
	Kabupaten Karanganyar	2	2
	Kabupaten Sukoharjo	3	3
	Kabupaten Purworejo	1	4
	Kabupaten Blora	1	5
Kota Salatiga	Kabupaten Klaten	1	1
	Kabupaten Kendal	1	2
	Kabupaten Boyolali	2	3
	Kabupaten Purworejo	1	4
	Kabupaten Semarang	3	5
Kota Semarang	Kabupaten Kendal	1	1
	Kabupaten Semarang	3	2
	Kabupaten Klaten	1	3
	Kabupaten Purworejo	1	4
	Kabupaten Blora	1	5
Kota Pekalongan	Kabupaten Batang	3	1
	Kabupaten Kendal	1	2
	Kabupaten Brebes	1	3
	Kabupaten Purworejo	1	4
	Kabupaten Klaten	1	5
Kota Tegal	Kabupaten Brebes	1	1
	Kabupaten Tegal	2	2
	Kabupaten Kendal	1	3
	Kabupaten Purworejo	1	4
	Kabupaten Klaten	1	5

The results of this study are expected to cut the distribution channel of soybeans to be shorter, which means that the goods will be quickly distributed to consumers, thus reducing the risk of damage, and immediately consumed by consumers to be processed into other foodstuffs. According to prior studies by Kristanti and Almuntaka [33] and Fiona, Soetoro, and Normansyah [32], the distribution pattern of soybeans begins with farmers, then moves to collectors, wholesalers, small merchants, and finally industrial customers. Based on the findings of this study, the present distribution pattern may be directly divided into farmers, collectors, and industrial customers.

As shown in Table 5, the soybean processing industry or consumers in areas that lack or do not produce soybeans can directly buy soybeans from collectors in other areas with good soybean production and a short distance away. This can cut marketing costs so that the difference in selling prices and buying prices between farmers and consumers or soybean industry players is not too big.

CONCLUSION

Based on the research conducted, it can be concluded that clustering Central Java Province based on soybean productivity in 2019 using K-means produced four clusters: the first with five members, the second with 14 members, the third with nine members, and the fourth with seven members. The seven members of Cluster 4 are advised to get soybeans from other nearby sources and to choose Kendal Regency as a preferable destination for soybean retrieval from Temanggung Regency and Semarang City. Furthermore, Klaten Regency became the primary suggestion for soybean collecting for the cities of Surakarta and Salatiga. Then there's Magelang Regency, which is the Magelang city's major suggestion for soybean harvesting. And Batang Regency, which is the primary soybean-picking location for Pekalongan City. Finally, for Tegal City, the major recommended place for soybean collecting is Brebes Regency. This study used the SAW method, which has weaknesses in terms of inconsistent weight allocation. The subjective nature of these weights significantly influences the ultimate result. To obtain more accurate and unbiased weights, it is recommended for future research to reassess the weighting utilized in the SAW approach, involving stakeholders or experts in the relevant fields.

REFERENCES

[1] G. Setyawan and S. Huda, "Analisis Pengaruh Produksi Kedelai, Konsumsi Kedelai, Pendapatan per Kapita . dan Kurs Terhadap Impor Kedelai di Indonesia," J. Ekon. dan Manaj., vol. 19, no. 2, pp. 215–225, 2022, doi: 10.29264/jkin.v19i2.10949.

[2] M. Nahrul, F. Ristanto, and S. N. Sarifah, "Analisis Determinan Volume Impor Kedelai Indonesia menggunakan Metode ECM (Error Correction Model) Tahun 1991- 2020," J. Ekon. Bisnis, Manaj. dan Akunt., vol. Im, pp. 18–30, 2022.

[3] S. Madrim, "Menteri Pertanian: 90 Persen Kebutuhan Kedelai Dipenuhi Lewat Impor," VOA Indonesia, 2022. https://www.voaindonesia.com/a/menteri-pertanian-90-persen-kebutuhan-kedelai-dipenuhi-lewat-impor/6410366.html (accessed Jun. 05, 2022).

[4] B. Kharisma, "Determinan Produksi Kedelai di Indonesia dan Implikasi Kebijakannya," E-Jurnal Ekon. dan Bisnis Univ. Udayana, vol. 3, p. 679, 2018, doi: 10.24843/eeb.2018.v07.i03.p03.

[5] I. Hardianti and N. D. Setiawina, "Faktor - Faktor Mempengaruhi Impor Kedelai di Indonesia," E-Jurnal Ekon. Pembang. Univ. Udayana, vol. 10, no. 6, pp. 2313–2340, 2021.

[6] C. M. Annur, "Nilai Impor Kedelai Indonesia Naik Jadi US$ 1,48 Juta pada 2021," databoks, 2022.https://databoks.katadata.co.id/datapublish/2022/03/07/nilai-impor-kedelai-indonesia-naik-jadi-us-148-juta-pada-2021.

[7] D. W. Laily, "Faktor-Faktor Yang Berpengaruh Terhadap Produksi Kedelai Nasional," J. Agrinka, vol. 2, no. 02, 2018.

[8] S. Rokhimah, T. Widjojoko, and A. N. Mandamdari, "Analisis Peramalan Produksi, Luas Panen, dan Harga Kedelai di Provinsi Jawa Tengah," in Prosiding Seminar Nasional Hasil Penelitian Agribisnis, 2022, vol. 6, pp. 124–130.

[9] M. Ivanni, N. Kusnadi, and Suprehatin, "Efisiensi Teknis Produksi Kedelai Berdasarkan Varietas dan Wilayah Produksi di Indonesia," J. Agribisnis Indones., vol. 7, no. 1, pp. 27–36, 2019.

[10] R. Anggreati, "Indonesia Dinilai Sanggup Kurangi Ketergantungan Impor Kacang Kedelai," medcom.id, 2022. https://www.medcom.id/ekonomi/bisnis/GNlWgwmK-indonesia-dinilai-sanggup-kurangi-ketergantungan-impor-kacang-kedelai (accessed Jul. 26, 2022).

[11] A. D. Fadhlullah, T. Ekowati, and Mukson, "Analisis Rantai Pasok (Supply Chain) Kedelai di UD Adem Ayem Kecamatan Pulokulon Kabupaten Grobogan," J. Pendidik. Bisnis dan Ekon., vol. 4, no. 2, pp. 1–10, 2018, [Online]. Available: https://jurnal.uns.ac.id/bise/article/download/22826/18374.

[12] Masitha, Solikhun, D. Suhendro, I. S. Damanik, and M. Fauzan, "Implementasi K-Means Clustering Untuk Mengelompokkan Hasil Pertanian Kacang Kedelai (Ha) Berdasarkan Provinsi," Senaris, vol. 2, no. 5, pp. 192–199, 2020.

[13] Sudirman, A. P. Windarto, and A. Wanto, "Data mining tools | rapidminer : K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia," 2018, doi: 10.1088/1757-899X/420/1/012089.

[14] H. Rhomadhona et al., "Sistem Pendukung Keputusan Distribusi Bantuan Pertanian Menggunakan Simple Additice Weighting," J. Syst. Comput. Eng. ISSN, vol. 2, no. 1, p. 1, 2021.

[15] Y. Yanuari, M. G. Husada, and D. B. Utami, "Aplikasi Rekomendasi Jenis Tanaman Pangan Menggunakan Metode Simple Additive Weighting (SAW," JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 3, no. 1, 2018, doi: 10.31328/jointecs.v3i1.495.

[16] P. Dauni, M. D. Firdaus, R. Asfariani, M. I. N. Saputra, A. A. Hidayat, and W. B. Zulfikar, "Implementation of Haversine formula for school location tracking," J. Phys. Conf. Ser., vol. 1402, no. 7, 2019, doi: 10.1088/1742-6596/1402/7/077028.

[17] C. Husada, K. D. Hartomo, and H. P. Chernovita, "Implementasi Haversine Formula untuk Pembuatan SIG Jarak Terdekat ke RS Rujukan COVID-19," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 4, no. 5, pp. 874–883, 2020, doi: 10.29207/resti.v4i5.2255.

[18] Badan Pusat Statistik, "Luas Panen, Produksi, dan Produktivitas Jagung dan Kedelai Menurut Kabupaten/Kota di Provinsi Jawa," Badan Pusat Statistik, 2021. https://jateng.bps.go.id/statictable/2021/04/15/2450/luas-panen-produksi-dan-produktivitas-jagung-dan-kedelai-menurut-kabupaten-kota-di-provinsi-jawa-2019.html (accessed Feb. 10, 2022).

[19] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," IEEE Access, voyl. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[20] S. Abadi et al., "Application model of k-means clustering: Insights into promotion strategy of vocational high school," Int. J. Eng. Technol., vol. 7, no. 2.27 Special Issue 27, pp. 182–187, 2018, doi: 10.14419/ijet.v7i2.11491.

[21] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, "A dynamic K-means clustering for data mining," Indones. J. Electr. Eng. Comput. Sci., vol. 13, no. 2, pp. 521–526, 2019, doi: 10.11591/ijeecs.v13.i2.pp521-526.

[22] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. Morgan Kaufmann Publishers, 2006.

[23] H. Harliana, R. M. Herdian Bhakti, O. Saeful Bachri, and F. Sofian Efendi, "Optimasi K-Means dengan Particle Swarm Optimization pada Pengelompokkan Daerah Stunting," J. Ilm. Intech Inf. Technol. J. UMUS, vol. 3, no. 02, pp. 95–101, 2021, doi: 10.46772/intech.v3i02.457.

[24] A. E. Wibowo and T. Habanabakize, "K-Means Clustering Untuk Klasifikasi Standar Kualifikasi Pendidikan Dan Pengalaman Kerja Guru," J. Din. Vokasional Tek. Mesin, vol. 7, pp. 152–163, 2022.

[25] A. H. Nasyuha, Zulham, and I. Rusydi, "Implementation of K-means algorithm in data analysis," Telkomnika (Telecommunication Comput. Electron. Control., vol. 20, no. 2, pp. 307–313, 2022, doi: 10.12928/TELKOMNIKA.v20i2.21986.

[26] R. A. Azdy and F. Darnis, "Use of Haversine Formula in Finding Distance between Temporary Shelter and Waste End Processing Sites," J. Phys. Conf. Ser., vol. 1500, no. 1, 2020, doi: 10.1088/1742-6596/1500/1/012104.

[27] N. Aminudin et al., "Higher Education Selection using Simple Additive Weighting," Int. J. Eng. Technol., vol. 7, no. 2.27, p. 211, Aug. 2018, doi: 10.14419/ijet.v7i2.27.11731.

[28] N. Setiawan et al., "Simple Additive Weighting As Decision Support System For Determining Employees Salary," Int. J. Eng. Technol., vol. 7, no. 2.14 Special Issue 14, pp. 309–313, 2018.

[29] A. Fitrul Hadi, R. Permana, and H. Syafwan, "Decision Support System in Determining Structural Position Mutations Using Simple Additive Weighting (SAW) Method," J. Phys. Conf. Ser., vol. 1339, no. 1, p. 012015, Dec. 2019, doi: 10.1088/1742-6596/1339/1/012015.

[30] Fauzi, Nungsiyati, T. Noviarti, M. Muslihudin, R. Irviani, and A. Maseleno, "Optimal Dengue Endemic Region Prediction using Fuzzy Simple Additive Weighting based Algorithm," Int. J. Pure Appl. Math., vol. 118, no. 7 Special Issue, 2018.

[31] A. Cahyapratama and R. Sarno, "Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) Methods In Singer Selection Process," in 2018 International Conference on Information and Communications Technology (ICOIACT), Mar. 2018, pp. 234–239, doi: 10.1109/ICOIACT.2018.8350707.

[32] Y. Fiona., Soetoro., and Z. Normansyah, "Analisis Pemasaran Kedelai (Suatu Kasus di Desa Langkapsari Kecamatan Banjarsari Kabupaten Ciamis)," J. Ilm. Mhs. Agroinfo Galuh, vol. 1, no. 2, pp. 137–142, 2015.

[33] N. E. Kristanti and I. S. Almuntaha, "Penentuan Saluran Pemasaran terhadap Tingkat Harga pada Rantai Pasok Kedelai (Glycine maxL.) Merr.) di Kabupaten Grobogan Provinsi Jawa Tengah," Agritech, vol. 37, no. 4, p. 443, 2018, doi: 10.22146/agritech.24808.

[34] Badan Pusat Statistik, "Rata-Rata Konsumsi per Kapita Seminggu Beberapa Macam Bahan Makanan Penting, 2007-2021," Badan Pusat Statistik, 2021. https://www.bps.go.id/statictable/2014/09/08/950/rata-rata-konsumsi-per-kapita-seminggu-beberapa-macam-bahan-makanan-penting-2007-2021.html. (accessed Feb. 10, 2022).