Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production

Nuri Omurbek¹, Onur Akcakaya²,* and Ezgi Dilan Urmak Akcakaya²

¹ Faculty of Economics and Administrative Sciences, Suleyman Demirel University, Isparta, Turkey
E-mail: ⟨nuriomurbek@sdu.edu.tr⟩

² Faculty of Economics and Administrative Sciences, University of Ardahan, Ardahan, Turkey
E-mail: ⟨{onurakcakaya, ezgidilanurmak}@ardahan.edu.tr⟩

Abstract. This study aims to cluster Turkish cities based on their local agricultural production and rank them in terms of performance by combining cluster analysis and multi-criteria decision-making (MCDM) methods. In this context, a three-phase methodology is developed. In the first phase, Ward’s method is utilized to cluster cities according to agricultural production characteristics. In the second phase, the objective criteria weights are determined using the Criteria Importance Through Intercriteria Correlation technique (CRITIC). In the third phase, to rank the clusters in terms of performance, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method is applied. Due to the results, the 81 cities are divided into six clusters in terms of agricultural production features. The cluster with the highest performance is Cluster 6, in which Konya is alone. Cluster 4, which includes Antalya and Mersin, follows this cluster. Cluster 1 with 25 cities and Cluster 2 with 19 cities are the clusters with the lowest results. The results show that only a few cities such as Konya, Antalya, and Mersin are generating more than tens of them in combination. These findings reveal that local governments should reconsider their agricultural programs and develop new strategies under the direction of the central government.

Keywords: Cluster analysis, local governments, multi-criteria decision-making, performance measurement, urban agriculture

Received: March 21, 2021; accepted: September 6, 2021; available online: December 23, 2021
DOI: 10.17535/crorr.2021.0009

1. Introduction

Agriculture has long been regarded as a rural activity. However, this has started to change since the discovery of its potential. Agriculture provides significant advantages to the urban economy such as bringing money to farmers and providing food for citizens. It also makes a critical contribution to urban food security [35].

The identification of the basic drivers of agricultural production is absolutely required to create effective public policy [29]. Measuring performance is also important in this regard. The primary goal of performance evaluation is to achieve effective and efficient project performance by providing an information flow to project management at each stage [28]. This helps decision-makers to decide how to increase effectiveness and efficiency with the available resources [14].

In this context, this paper aims to cluster Turkish Cities according to their local agricultural production and to rank them in terms of performance by using cluster analysis and multi-criteria decision-making (MCDM) methods in an integrated way. This aim includes two main objectives.

*Corresponding author.
The initial step is to look for cities with similar production characteristics. As a result, units in the same cluster will be able to cooperate more effectively in the future in terms of decision-making and implementation. The second objective is to determine the best performance among all clusters. Thus, it is expected that the performance distribution of agricultural production throughout the country can be defined by local units. This may provide the local units with the advantage of determining their current positions, as well as strengthen the competition among themselves.

To achieve these two goals, first, we construct the nine agricultural production criteria, based on the literature, expert opinions, and data availability. The data set for each criterion is collected from the database of the Turkish Statistical Institute (TUIK). It includes values belonging to nine agricultural production criteria between 2007-2018. At the clustering phase, we converted the annual data into average because using only one year’s data would not be logical for the evaluation.

The availability of data is an additional factor in determining the criteria because it is not easy to find quality and regular data for certain criteria. For instance, the criterion of the amount of vegetables and fruits produced in the greenhouse is not accessible for all cities. Hence, we had to eliminate some criteria. Also, no data exists for some agricultural activities such as the fishing industry.

In this context, we develop a three-step approach that combines Ward’s, the Criteria Importance Through Intercriteria Correlation Technique (CRITIC), and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) methods. In the first phase, Ward’s method is used to find the most similar couple of clusters. In the second phase, CRITIC is utilized for determining the objective weights of agricultural production indicators, and in the third phase, TOPSIS is used to evaluate in terms of their performance.

The remainder of this paper is organized as follows: Section 2 includes a review of the literature on related studies. Section 3 presents the information about the data used in the paper. Section 4 lays out a detailed description of the methodology. The empirical application is given in Section 5 by utilizing the application steps. In this section, results are also presented. Finally, Section 6 discusses the findings, provides concluding remarks, and makes suggestions for future studies.

2. Literature Review

Similar to this study, some decision-making problems require additionally classifying alternatives based on similar characteristics, so additional methods may be required. Cluster analysis is a useful method for such situations because it can predetermine the clusters that consist of similar objects [11]. Cluster analysis has been used in studies from across various research fields, such as in reviewing applications of cluster analysis in marketing and presenting alternative methods of cluster analysis [34]; exploring host communities’ perceptions of tourism [33]; comparing spectral, hierarchical and k-clustering in e-nose data-sets [16]; determining the efficiency of at regulation on greenhouse gas emissions [2]; determining the impact of dimensionality reduction on stock in distinct market situations [13]; and the analysis of water quality [10]. One of the widespread hierarchical clustering techniques is Ward’s method [39]. Ward’s method has been applied in many studies, such as in comparing various clustering methods for the clustering of mixed-mode data [9], analyzing healthcare systems [24], comparing two clustering methods in portfolio management [20], developing software for data analysis [26], and researching combinations of distance metrics and hierarchical clustering criteria [42].

MCDM methods are applicable to solving decision-making problems [23]. Recently, they have become increasingly popular in decision-making with multi-dimensional attributes [38]. TOPSIS is one of the MCDM methods and it is widely used, either separately or integrated with other methods, to solve at multitude of decision-making problems. It has been utilized in a
Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production

A large number of studies, such as with compromise solutions [30], researching the applications of TOPSIS [5], measuring the efficiency of economic sectors [4], selecting suppliers [22], selecting a location for a liquefied natural [3], assessing sustainable housing affordability [25], evaluating the ecological-economic efficiency of innovations in green technology [37], selecting an optimal road safety composite index [36], assessing integrated flood vulnerability [41], and selecting biological nano-materials [43].

In multi-criteria decision-making models, determining criteria weight coefficients is a critical step. The CRITIC approach is one of the most extensively utilized and well-known objective methods. It is a correlation approach that determines criteria contrasts by using the standard deviation of ranking criteria values of alternatives per column, as well as the correlation coefficients of all paired columns [44]. The CRITIC Method has been used in a large number of decision-making studies to determine objective weights [27, 31, 40].

3. The Data

In this section, we purpose to provide a series of data to cluster and evaluate the cities in Turkey in terms of agricultural production criteria. Based on the literature, expert opinions, and data availability, we constructed the nine criteria, as follows: (C1) The number of cattle (pcs), (C2) the value of live animals (thousand Turkish Lira [TL]), (C3) the value of animal production (thousand TL), (C4) the agricultural production value (TL per capita), (C5) the number of small cattle (pcs), (C6) the production amount of cereals and other plant products (ton), (C7) the total cultivated agricultural area (hectare), (C8) the total value of agricultural production (thousand TL), (C9) and the total value of plant production (thousand TL). For convenience, we gave codes to the criteria in the methodology. The data set for each criterion is collected from the database of TUIK.

4. Methodology

This section aims to lay out the methodology of the paper. We develop a three-step approach integrating cluster analysis and MCDM methods. The methodology applied in this study is given in Figure 1.

<table>
<thead>
<tr>
<th>Phase I: Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>The data was collected from the database of TUIK and the average data was calculated between 2007-2018. Utilizing Ward’s method, all cities were clustered according to their agricultural production features.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase II: Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRITIC method is used for determining the objective weights of agricultural production indicators.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase III: Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clusters were evaluated in terms of agricultural production performance, by applying the TOPSIS method.</td>
</tr>
</tbody>
</table>

Figure 1: Three-Step Approach

4.1. Ward’s Method

Ward’s method is a common hierarchical clustering method that composes clusters so that the variance within the clusters is minimal. It determines a cluster as a gathering of objects such that the error sum of squares amongst the members of each cluster is minimal [6].

The treatment of Ward’s method can be summarized briefly as follows. Firstly, it is assumed that there are N elements to a cluster. It starts with N clusters consisting of one entity. Then,
the similarity matrix is searched for the most similar pairs, and by combining them the number of clusters is reduced to one. These steps are applied until all clusters are obtained [15].

4.2. CRITIC Method

The steps of the CRITIC method are as follows [8]:

**Step 1: Structure of the decision matrix**

Consider a decision matrix, \( X = [x_{ij}]_{m \times n} \), where \( x_{ij} \) is the performance measurement of the \( i \)-th alternative in respect to the \( j \)-the criterion.

\( m \) : is the number of alternatives

\( n \) : is the number of criteria

**Step 2: Normalization of the decision matrix**

Normalizing the decision matrix by using (1) and (2):

\[
\begin{align*}
    r_{ij} &= \frac{x_{ij} - x_{ij}^{min}}{x_{ij}^{max} - x_{ij}^{min}} \quad \text{benefit-based} \\
    r_{ij} &= \frac{x_{ij}^{max} - x_{ij}}{x_{ij}^{max} - x_{ij}^{min}} \quad \text{cost-based}
\end{align*}
\]

\( x_{ij}^{max} = \max (x_{ij}, 1, \ldots, 6) \)

\( x_{ij}^{min} = \max (x_{ij}, 1, \ldots, 6) \)

**Step 3: Calculate the criteria weights**

In the determination of criteria weights, both the standard deviation of the criterion and its correlation between other criteria are included. The weight of the \( i \)-th criterion \( w_j \) is calculated by (3):

\[
    w_j = \frac{C_j}{n \sum_{k=1}^{n} C_i}
\]

Where \( C_j \) is the quantity of information contained in the \( j \)-th criterion obtained by (4):

\[
    C_j = \sigma_j \sum_{k=1}^{n} (1 - r_{jk})
\]

\( \sigma_j \) : the standard deviation of the \( j \)-th criterion

\( r_{jk} \) : the linear correlation coefficient between the \( j \)-th and the \( i \)-th criteria.

4.3. TOPSIS Method

The TOPSIS method is utilized to overcome multiple decision-making problems [7]. The ideal solution minimizes the cost criteria and maximizes the benefit criteria, while the negative-ideal solution maximizes the cost criteria and minimizes the benefit criteria [32]. The steps of the TOPSIS method are as follows [21]:

- **Step 1: Structure of the decision matrix**
- **Step 2: Normalization of the decision matrix**
- **Step 3: Calculate the criteria weights**
- **Step 4: Construct the positive-ideal solution and the negative-ideal solution**
- **Step 5: Calculate the relative closeness coefficients**
- **Step 6: Rank the alternatives**
Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production

**Step 1: Structure of the decision matrix**

Evaluating the decision matrix which includes $m$ alternatives in association with $n$ criteria.

**Step 2: Normalization of the decision matrix**

It is required to normalize the original matrix to ensure that all the criteria are equivalent and in the same form. The normalized decision matrix is $R = [r_{ij}]_{m \times n}$, which is obtained by (5).

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{m} x_{ij}^2}} \quad i = 1, \ldots, 6; \ j = 1, \ldots, 9$$

**Step 3: Obtain the weighted normalized decision matrix**

The weighted decision matrix is calculated by the normalized decision matrix multiplied with the weights of the indexes as illustrated by (6).

$$v_{ij} = w_i r_{ij} \quad i = 1, \ldots, 6; \ j = 1, \ldots, 9$$

**Step 4: Determine the ideal and negative-ideal solutions**

The ideal solution is composed of the optimal value of each criterion from the weighted decision matrix as illustrated by (7). The negative ideal solution is composed of the worst value of each criterion from the weighted decision matrix as illustrated by (8).

$$A^* = \left\{ \left( \max_i v_{ij} \mid j \in J \right), \left( \min_i v_{ij} \mid j \in J' \right) \right\}$$

$$A^* = \{v_1^*, v_2^*, v_3^*, \ldots, v_n^*\}$$

$$A^- = \left\{ \left( \min_i v_{ij} \mid j \in J \right), \left( \max_i v_{ij} \mid j \in J' \right) \right\}$$

$$A^- = \{v_1^-, v_2^-, v_3^-, \ldots, v_n^-\}$$

**Step 5: Calculate the separation measure**

The distance of every feasible solution from the ideal solution and the negative ideal solution is calculated in turn as illustrated by (9) and (10).

$$S_i^* = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_{ij}^*)^2}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_{ij}^-)^2}$$

**Step 6: Determine the relative closeness to the ideal solution**
The relative degree of approximation is calculated by (11).

\[ C_i = \frac{S_i^-}{S_i^- + S_i^+} \quad (0 \leq C_i \leq 1; \; i = 1, 2, \ldots, 6) \] (11)

The alternative is sorted in accordance with the value of the relative degree of approximation. The bigger the value is, the better the alternative is.

5. Empirical Application

The purpose of this section is to give information about the implementation of the methodology and the results. The implementation has two phases in parallel with the scope.

5.1. Clustering Cities According to Agricultural Production Criteria

Firstly, we calculated the arithmetic mean of the data to obtain the average data set. Thus, we obtained a decision matrix containing all cities. Then, the summary statistics table of the decision matrix was obtained (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>81</td>
<td>9527.66</td>
<td>62999.83</td>
<td>166042.97</td>
<td>130004.11</td>
<td>1690106757</td>
</tr>
<tr>
<td>C2</td>
<td>81</td>
<td>47495.91</td>
<td>343847.50</td>
<td>818874.85</td>
<td>601144.83</td>
<td>3686248474</td>
</tr>
<tr>
<td>C3</td>
<td>81</td>
<td>28085.33</td>
<td>352832.83</td>
<td>327116.71</td>
<td>267602.58</td>
<td>1663590925</td>
</tr>
<tr>
<td>C4</td>
<td>81</td>
<td>70.83</td>
<td>10995.75</td>
<td>3766567901</td>
<td>1914374171</td>
<td>36648285</td>
</tr>
<tr>
<td>C5</td>
<td>81</td>
<td>5367.85</td>
<td>250049.33</td>
<td>452003.67</td>
<td>4425922962</td>
<td>1957082263</td>
</tr>
<tr>
<td>C6</td>
<td>81</td>
<td>8956.66</td>
<td>10849041</td>
<td>1189113.44</td>
<td>1385399697</td>
<td>1919082950</td>
</tr>
<tr>
<td>C7</td>
<td>81</td>
<td>7977.09</td>
<td>256596.80</td>
<td>1890689.64</td>
<td>3574702932</td>
<td>1682280759</td>
</tr>
<tr>
<td>C8</td>
<td>81</td>
<td>259319.58</td>
<td>9722508.66</td>
<td>2352225.22</td>
<td>2352225.22</td>
<td>8164930381</td>
</tr>
<tr>
<td>C9</td>
<td>81</td>
<td>31873.53</td>
<td>7367297.16</td>
<td>1296553.66</td>
<td>1296553.66</td>
<td>10822872055</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Descriptive Statistics

Cluster analysis has the advantage of not making any assumptions about the number of clusters or cluster structure [17]. In addition, it is another advantage that assumptions such as normality, linearity, and homoscedasticity, which are important in other multivariate statistical analysis techniques, are not taken into account in cluster analysis [12].

Ward’s method was chosen due to its several advantages over other algorithms. These are as follows: (1) There is no need to predetermine the number of clusters. The methods decides how many clusters there should be. (2) The procedure can be applied to any number of clusters between 1 to N, where N is the total number of elements in the original set. (3) It is one of the few techniques that actually reduces the total square error. (4) Ward’s method allows cluster analysis to be performed even with a limited number of observations [18]. By utilizing Ward’s method, a dendrogram was created to determine the clusters. With the help of a dendrogram, clusters were listed from top to bottom. The dendrogram is scaled from left to right as 0-25 units. This scale serves to show the distance between the clusters. In terms of the data used in the dendrogram, the cities that are most similar to each other form a cluster at a distance of one unit, while the cities that are least similar to each other come together at a distance of 25 units. (Figure 2). According to these results, six clusters were produced at a distance of one unit and all cities were divided into six clusters (Table 2).
Figure 2: Structural model: standardized solution.
The fact that the clusters contain a different number of cities is due to their variance in agricultural production criteria such as the number of cattle, the value of live animals, the value of animal production, agricultural production value, the number of small cattle, the production amount of cereals and other plant products, the total cultivated agricultural area, the total value of agricultural production, and the total value of plant production. In addition to criteria related to the amount and value of animal/plant production, criteria related to suitable lands
Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production

for agriculture is also effective in clustering. The city of Konya was included in Cluster 6 alone because this city is unique among all other cities in terms of agricultural production potential and production criteria. It is also in the leading position.

The Kruskal Wallis test was applied to determine whether there is a significant difference between the clusters in terms of each criterion (Table 3).

According to the Kruskal Wallis test results, it was observed that there were statistically significant differences between the clusters in terms of other criteria, except for the agricultural production value per capita (C4) criterion.

5.2. Ranking Clusters According to Agricultural Production Criteria

The purpose of this section is to rank clusters produced in the first phase from best to the worst in terms of agricultural production values. While the TOPSIS method was applied in the evaluation of the clusters, the CRITIC method was used in determining the objective criteria weights required for the application of the TOPSIS method.

The CRITIC method was chosen due to its several advantages over other methods. (1) It allows a more objective evaluation by considering the correlation coefficients between the variables. Thus, the weights of the criteria are determined objectively by avoiding the negative effects that may be caused by the subjective interpretations of the decision maker. (2) CRITIC is a simpler technique that requires less computational effort [8]. (3) It does not distinguish between beneficial and non-beneficial criteria, instead attempting to assess the severity of the contrast in the decision making problem’s structure [1].

TOPSIS was chosen due to its several advantages over other MCDM methods. The TOPSIS Method has: (1) a sound logic that embodies the rationale of human choice, (2) a scalar value that considers both the best and worst options simultaneously, and (3) a simple computation procedure that can be easily programmed in a spreadsheet [19].

Firstly, by utilizing the decision matrix, the average values of the criteria were calculated for each cluster. Thus, a new decision matrix was obtained for both the CRITIC and TOPSIS methods (Table 4).

By utilizing the CRITIC method, the weights of criteria were obtained (Table 5). According to the criterion weights, the criterion of agricultural production value per capita (C4) was determined to be the most important criterion with a value of 0.264 and, the number of small cattle (C5) is the least important criterion with a value of 0.066 (Table 5).

To obtain a ranking of clusters, the TOPSIS method was used. Firstly, Equation (5) was utilized to normalize the decision matrix. Then, the weighted decision matrix was created by multiplying the normalized values with the weights for each criterion. Criterion weights obtained by the CRITIC method were used to obtain the weighted decision matrix. The ideal ($A^+$) and negative-ideal ($A^-$) solution values for each criterion were calculated by using Equation (7) and Equation (8), respectively. The positive ideal solution is the one that maximizes the benefit criterion while minimizing the cost criterion. The negative ideal solution, on the other hand, is considered to be the solution that maximizes the cost criterion while minimizing the benefit criterion. By utilizing the ideal and negative-ideal solution values, the distance values to the ideal and negative-ideal points ($S_i^+$ and $S_i^-$) were calculated using Equation (9) and Equation (10), respectively. In the last step, the relative closeness to the ideal solution ($C_i^*$) was calculated by utilizing Equation (11) (Table 6). Clusters were listed from greatest to smallest according to their closeness to the ideal solution $C_i^*$ (Table 7).
6. Conclusion

In this paper, we purposed to cluster Turkish cities according to the criteria for agricultural production, and to evaluate the clusters according to their performance. For this purpose, we utilized a three-phase approach to address research questions. In the first phase, we utilized Ward’s method for clustering all cities. In the second phase, the CRITIC was utilized for determining the objective weights of agricultural production indicators, and in the third phase, by using the TOPSIS method, we evaluated these clusters in terms of their agricultural efficiency.

This study makes two main contributions to the literature. The first is to identify cities with similar characteristics in terms of agricultural production. Thus, the units in the same cluster...
are provided with information about alternatives with which they can cooperate more effectively in future decision-making and implementation processes. The second contribution of the study is to show that some cities can produce more individually than tens of cities in combination. Thus, it is possible to have information about the regions that perform low production despite having sufficient opportunities.

In this sense, we consider that a strategic and sustainable mechanism needs to be established to ensure agricultural growth, including successful central strategic planning, well-disciplined local practices and performance assessments. Central strategic planning, which takes regional situations into account, can lead to the level-headed distribution of agricultural production across the country. In this way, in compliance with particular local characteristics or circumstances, plant and animal products can be supplied. On the other hand, well-disciplined practices can help to enhance distinctive local production. In terms of ensuring sustainable growth, increasing welfare, and reducing external dependency, all of these initiatives primarily serve the country. It is therefore possible to achieve sustainable agricultural production.

In this analysis, we needed to find quality data for all the parameters because it was empirical. However, it was difficult to find data available for all criteria and all towns because the database is limited. To eliminate such limitations, the database should be developed in terms of data diversity and regularity. The following agricultural variables should be established in this context: fishing production; greenhouse production; pasture fields; geographic, climatic, and irrigation conditions; and soil quality.

Finally, we emphasize the importance of creating a Turkey-specific index that will enable the comparative evaluation of local agricultural performance. This will allows for a simple, comparative and logical evaluation of regional or local contributions to the overall output. For policy-makers and practitioners, a further suggestion would be to constantly track local and regional agricultural output in order to recognize problem areas and develop effective strategies and solutions.

References

Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production

10.1016/j.neunet.2007.01.002


