sciendo

Strategic Categorization of Dairy Cow Farms in Croatia using Cluster Analysis

Maja Petrač, Krunoslav Zmaić
Faculty of Agrobiotechnical Sciences Osijek, University J.J. Strossmayer, Croatia
Jaka Žgajnar
Biotechnical Faculty, University of Ljubljana, Slovenia

Abstract

Background: The milk processing sector in the Republic of Croatia faces numerous challenges. It is a distinctly bipolar structure, with some entities resembling the largest milk producers in the EU, while many small and medium-sized dairy farms struggle to remain competitive and achieve further progress. To formulate effective policy, it is important to differentiate between these types and address their key challenges. **Objectives:** The aim is to find the most representative solution that will help us define typical dairy farms and upgrade a SiTFarm tool (Slovenian Typical Farm Model), enabling us to assess the situation in Croatia comprehensively. **Methods/Approach:** Cluster analysis was conducted using empirical data obtained from the Croatian Agency for Agriculture and Food. The analysis involved applying both hierarchical and non-hierarchical clustering techniques. **Results:** Two cluster analysis scenarios are presented, differing in the variables used. In each scenario, 16 relatively homogeneous clusters of farms were obtained. Diversity was minimized within these clusters, and they effectively explain the dairy business in Croatia. **Conclusions:** The results of this analysis thus represent an important starting point for further analysis of the dairy sector in Croatia. These findings could help policymakers identify the types of farms that would benefit most from targeted investments to enhance efficiency, economic viability, and environmental sustainability.

Keywords: cluster analysis; typical farms; dairy sector; mathematical programming; farm model

JEL classification: C38, Q13
Paper type: Research article

Received: May 12, 2024
Accepted: Jul 2, 2024

Citation: Petrač, M., Zmajić, K., & Žgajnar, J. (2024). Strategic Categorization of Dairy Cow Farms in Croatia using Cluster Analysis. *Business Systems Research, 15*(2), 140-153.
DOI: doi.org/10.2478/bsrj-2024-0021

Introduction

Over the years, the number of farms, livestock, and milk production in the Republic of Croatia has steadily declined, reflecting a typical consolidation process, as described by Gonzalez-Mejia et al. (2018). Small farms with a few animals are either closed or transferred to arable production (Mijić and Bobić, 2021), there was, however, some increase in larger farms (Očić et al., 2023). As the situation in the milk production sector in the Republic of Croatia is far from good, and the measures implemented obviously do not lead to improvement, a more detailed analysis of the sector is needed. Očić et al. (2023) see the limited number of competitive farms as the main challenge facing the Croatian dairy sector. Thus, they advocate for various agricultural and rural development measures aimed at empowering especially smaller farms and enhancing the overall competitiveness of the dairy sector. To achieve this goal, the development of appropriate models for a more detailed analysis of the sector and support, especially policy decision-makers, is a logical follow-up. So, different models are needed for decision-makers to have a better insight into what is happening on certain types of agricultural holdings and for their decisions to be based on facts (Ciaian et al., 2013).

The common agricultural policy (CAP) strategic plans (CSP) of the EU member states place increasing emphasis on the use of models that enable simulation at the farm level or the level of the selected aggregate (Lovec et al., 2020). They suggest that each member state should choose its agricultural policy priorities and, in accordance with common EU principles, determine the type, allocated funds, and scope of individual measures.

In such a manner, models become the main tool for generating scenarios or 'what-if' analyses because the effects of policies differ by farm type. It can simulate how a particular scenario, for example, a change in agricultural resource prices or agricultural or environmental policy, may affect a set of performance indicators (Ciaian et al., 2013). In the farm models, mathematical programming is most often applied, as well as models based on econometric and simulation approaches Pečnik and Žgajnar (2022). As it is impossible to carry out the analysis at the level of every agricultural farm, it is carried out at the level of typical representatives. Farms should be classified into smaller categories with common characteristics, which are called typical agricultural farms (Alvarez et al., 2018). These are generally real or hypothetical farms that best represent the situation in a certain segment of a particular sector (representative households) and allow generalization at the aggregate level. Poczta et al. (2020) identified five different types of dairy farms based on a cluster analysis of EU-wide Farm Accountancy Data Network (FADN) data. Based on average indicators, Croatian farms were classified into the first type, along with Slovenia, Austria, Poland, and Romania.

Additionally, Očić et al. (2023), using FADN data, analysed the dairy sector in Croatia, employing a different approach from the one utilized in our study. In their study, which relied on FADN data, they designed three types of farms focused only on the number of cows in the herd. Our goal, however, is to delve even deeper into researching the sector and categorize it into several types, enabling us to provide additional support to different stakeholders. The potential and structure of dairy herds within the sector vary significantly, and tailored solutions are needed to enhance efficiency.

Before the actual creation and application of the farm model in the dairy sector in the Republic of Croatia, it is necessary to define typical farms focusing on dairy production. These farms should be categorized according to main common characteristics and production endowments. Building on these insights, we will establish the initial parameters for production plans per each farm type. With the assistance of experts and stakeholders, we will refine and calibrate these plans, and all needed technological parameters to reflect the realities of the sector accurately. Therefore, the main purpose of this work is to obtain typical farms by applying cluster analysis based on available data following the example of similar studies. It does not require any assumptions about the number and structure of the categories into which the data will be distributed, but the categorization is done based on similarity between the data of farms. The aim is to find the most representative solution that will help us define typical dairy farms and, in a further step, upgrade a SiTFarm (Slovenian Typical Farm Model) tool, enabling us to assess the milk production situation in Croatia comprehensively. Namely, SiTFarm is a microsimulation tool based on mathematical programming and serves as an example of a bioeconomic farm model (BEFM). It enables various analyses at the level of the agricultural production plan, with results that can also be aggregated at the sector level (Žgajnar and Kavčič, 2024). The primary purpose of SiTFarm is to facilitate analyses from the perspective of income sustainability for typical farms that are representative of a certain number of real farms. This model does not require FADN data, allowing for more detailed analyses of smaller farms that would otherwise not be included in the sample. The model calculates various economic indicators and accommodates different CAP interventions at multiple levels and under various conditions, considering the socio-economic context of the analysis.

Dairy farms are integral to the milk sector, and their development directly influences milk production. As emphasized by Parzonko et al. (2024), there is a wide diversity of dairy farms across EU countries concerning the scale and technology of milk production, and this applies to Croatia as well. Cluster analysis will be conducted for both family farms and farms owned by legal entities, covering the entire milk sector in the Republic of Croatia. In the future, this will also enable a comprehensive analysis of all three aspects of sustainability, where significant differences exist between individual types of farms, as well as regions and production conditions.

In the continuation of the paper, we first present the database used for the analysis, including brief descriptive statistics. A description of the methodology employed follows this. In the results chapter, we analyse the possibilities and the influence of various variables on the design of typical farms. We conclude with key findings and guidelines for future work.

Methodology

Data

Empirical data was obtained from the Croatian Agency for Agriculture and Food (Agency). This data is collected from all agricultural holdings in Croatia that deliver milk. Farmers report in a standardized format, including dates and other relevant details. This data is at the farm level. The database contains a list of all registered farms in the Republic of Croatia engaged in milk production.

The database consisted of 4198 dairy cow farms that supply milk in the Republic of Croatia. There was no comprehensive farm database in the Republic of Croatia that contained the information we needed (e.g., number of cows, amount of milk delivered, breed structure, area under cultivation, main crops, land structure, number of employees, location). Unfortunately, not all the needed data were available, so we worked with the data that were accessible to us. Thus, the resulting database was combined from the available date of the Agency dated 2022, with issues such as multiple identifiers and the removal of duplicates and some mistakes, such as farms with zero cows. After arranging the resulting database (connecting data from different farm databases, removing duplicates and inactive farms, etc.), 3393 farms remained for analysis. Of the total number of farms, 3331 farms are family farms, while 62 farms have the status of a legal entity and constitute a special category. These are large, highly specialized dairy farms playing a key role in milk production. Among the family farms, there are also a few large-scale milk producers.

As can be seen from Table 1b, it is a very diverse group of dairy farms. We have chosen four variables that clearly show the differences between the two categories of farms (Table 1a). We tracked the number of cows (NOC) in the herd, representing the average annual stock in the year (2022). Milk production (ADOM) was monitored on the farm level, including the annual production of milk supplied to dairies. Additionally, we monitored agricultural production on the arable land, which is essential for fodder and cash crop cultivation. These crops often supplement or, in some cases, even replace milk production. This trend is expected to continue, especially as further structural changes occur. As highlighted in the introduction, many small and medium dairy farms often shift to arable production and abandon milk production (Mijić and Bobić, 2021). Further, we included two additional variables: the number of plant cultures (NOPC) and the area under cultivation (AUC). These variables shed light on the farm's land use, feeding strategies, and development potential, providing a more comprehensive characterization of the farms.

Table 1a
List of variables at the farm level

Variable Name	Variable
NOC	Number of cows
ADOM	Annual delivery of milk (kg)
NOPC	Number of plant cultures
AUC	The area under culture (ha)

Source: Author's work

Table 1b
Descriptive statistics for dairy farms

Variable Name	Mean	SD	Min	Max	Q1	Q2	Q3
Family farms (n = 3331)							
NOC	14.78	20.47	1	456	5	10	18
ADOM	73257	146015	21	2799071	14438	34840	76882
NOPC	6.29	2.27	1	18	5	6	8
AUC	23.68	31.68	0.15	469.05	7.64	14.1	27.4
Legal entity dairy farms (n = 62)							
NOC	337.74	682172	1	4051	28	88.5	369.75
ADOM	2853309	5996202	9938	34601195	124963.25	473117	3111024
NOPC	6.1	3.28	1	21	4	6	7
AUC	2076.19	12946.85	2.38	101427.88	31.06	104.84	306.78

Source: Author's work

The variables and their descriptive statistics are presented in Table 1b. The first category of farms includes 3331 family farms that submitted milk and were included in the resulting register as such (Table 1b). As can be seen from Table 1b, the first category includes farms that raise an average of 14.78 dairy cows, with the median being 10 cows. On 75% of these farms, the number of dairy cows is less than 18, while only 25% of the farms have more than 18 animals. High variability is indicated by a high standard deviation (SD) value. The same applies to milk production. Here, the differences are even greater because, in addition to herd size, lactation milk yield, which reflects the intensity of production, also plays an important role. This indicator further underscores the high variability of family dairy farms. The median is 34840 kg of milk per average farm, with 75% of farms producing less than 76882 kg of milk per year. We can observe that on family farms engaged in milk production, the number of agricultural activities averages 6.3. The variability here is significantly lower, indicating that, in most cases, it involves home-produced fodder. However, in 25% of cases, we can also find examples of more diversified production plans where the number of plant cultures exceeds 5. The last indicator, the area under cultivation, again shows greater variability. On average, dairy farms cultivate 23.68 ha, with a median of 14.1 ha, and the largest 25% of farms more than 27.4 ha.

The second category is 62 farms that have the status of a legal entity (Table 1b). In this example, we observe that the average herd size is notably higher, with 337.74 dairy cows per herd. Additionally, among the largest farms, 25% have a herd size exceeding 369.75, and 75% exceed 28 dairy cows. These are particularly large farms that play a pivotal role in the dairy sector. Annual milk production correlates closely with herd size, indicating significant variability among farms in this aspect. This variability suggests the presence of both intensive forage farms and those employing economies of scale with extensive forage from grasslands. Furthermore, these farms cultivate an average of 6 crops, although there is slightly higher variability compared to family farms. As indicated in the Table 1b, these farms are notably large, averaging 2076 ha of cultivated land.

Cluster analysis and scenarios

The cluster analysis was first performed in relation to all the mentioned variables. First, hierarchical (agglomerative) grouping was performed using Ward's method. Agglomerative grouping starts from a single object, in our case, from a dairy farm. In the first step, each dairy farm forms one cluster. The two most similar dairy farms are grouped into one cluster. Then, a new dairy farm is added to that cluster, or the other two dairy farms are grouped into a new cluster. The merging continues according to mutual similarities until all subgroups are merged into one cluster. The matrix of similarity (distance) between all objects (dairy farms) is the basis for this method. It is a symmetrical matrix with dimensions $n \times n$.

One of the methods used to determine the similarity between clusters is Ward's method or the minimum variance method. This method minimizes the sum of squares between any two clusters that could be formed. Agglomerative grouping is more often used in practice because it is more implemented in computer programs than, for example, divisive grouping. The graphical presentation of the results of agglomerative and divisive grouping can be graphically presented in the form of a two-dimensional hierarchical diagram, known as a dendrogram.

After hierarchical clustering, non-hierarchical clustering was performed, i.e., the k-means algorithm (Scitovski and Sabo, 2020). The k-means algorithm belongs to the optimization algorithms of non-hierarchical clustering that enables the redistribution of objects. It organizes objects (dairy farms) into a predetermined number of clusters, k, and then iteratively reassigns objects to clusters until a specified numerical criterion is met. Achieving the criteria is closely related to achieving the goal of cluster analysis, which consists of finding as compact and better-separated clusters as possible. Therefore, the goal is to minimize the distance between objects within a cluster and maximize the distance between the clusters themselves.

Given that all variables are numeric, the squared Euclidean distance $d(x, y)$ is defined as

$$d(x,y) = \sum_{i=1}^{n} (x_i - y_i)^2 \quad (1)$$

where $x, y \in S$, S is the data set, $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$, $n \in \mathbb{N}$, and x_i, y_i, $\forall i \in \{1,2,\dots,n\}$, are numeric variables.

All algorithms were applied to standardized z-score so that all variables had equal weight in the cluster analysis. Let $x^{(1)}, x^{(2)}, \dots, x^{(n)} \in S$ be any objects $1 \leq k \leq n$. The formula calculates the Z-score.

$$z_k = \frac{x_k{}^{(j)} - \mu_k}{\sigma_k} \quad , \qquad 1 \leq j \leq n, \quad (2)$$

where μ_k is arithmetic mean and σ_k is the standard deviation. Among the many solutions, two final solutions were chosen for each category. Cluster analysis was conducted for each category using the IBM SPSS Statistics V22.0 software package.

Cluster analysis solutions are not unique and depend on various elements of the analytical procedure, such as the choice between hierarchical or non-hierarchical methods and different algorithms within each method. The selection of variables used to measure similarity also influences the solution. Therefore, it is important to consider the impact of each decision carefully when choosing variables and conducting cluster analysis.

In this paper, two cluster scenarios are analysed for each category. The first scenario describes the clusters in which all the mentioned variables (Table 1a) are included separately for family farms (Table 2) and legal entities (Table 3). In the second scenario, only two variables, NOC and ADOM, were included for both categories (family farms Table 4 and legal entities Table 5).

The utilization of two scenarios serves the purpose of analysing the disparities and the typology of farms that emerge as prototypical representatives within their respective groups. Simultaneously, we will monitor which types of agricultural holdings are more rational from a professional standpoint and will serve as the foundation for designing typical farms for further modelling with the SiTFarm tool.

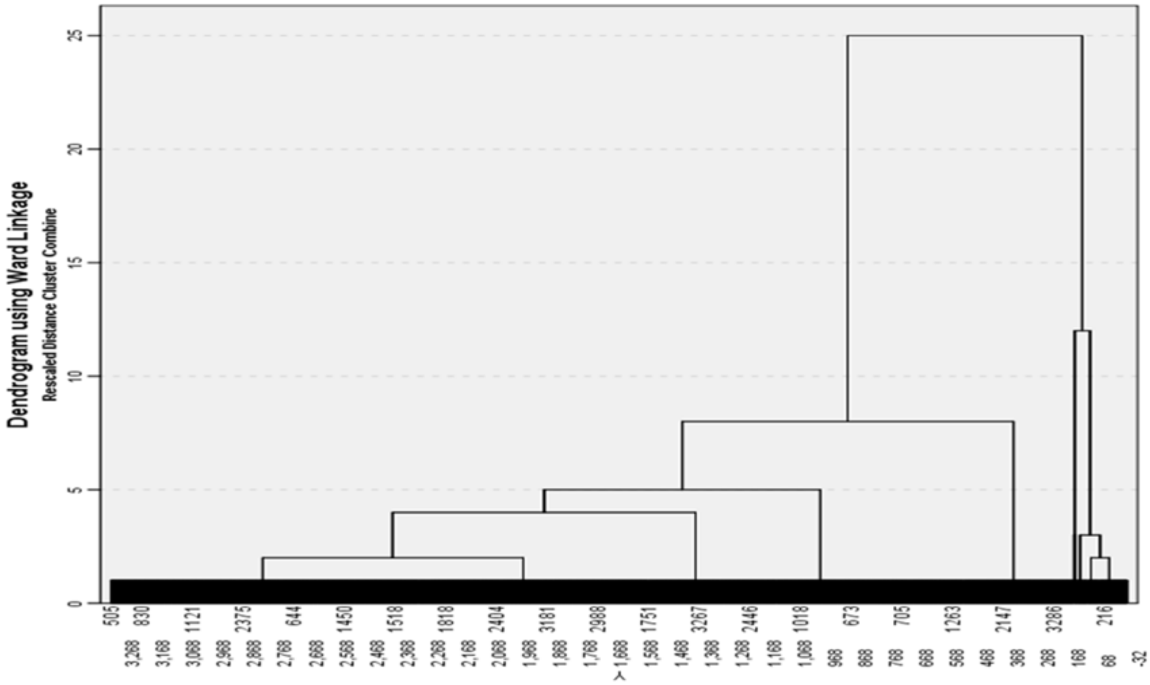
Results

Scenario 1: Four variables

Family dairy farms

The subsequent step in the analysis is to determine the optimal number of clusters using the dendrogram. While the dendrogram assists in this determination, researchers and experts generally make the final estimation. As depicted in the dendrogram, the analysis suggests the presence of 10 clusters (Figure 1). The data structure was analysed, revealing distinct categories of farms with shared characteristics. As the number of clusters increases, the heterogeneity within clusters decreases. Following the hierarchical analysis (dendrogram), a non-hierarchical algorithm (k-means) was applied to the data, with a maximum of 10 clusters imposed. Table 2 presents the composition of all clusters post k-means implementation, illustrating the distribution of farms across the 10 clusters. It delineates which farms belong to each cluster, their respective characteristics, the number of farms within each cluster, and the average number of cows per cluster.

Figure 1
Dendrogram – Scenario 1 – Family dairy farms



Source: Author's work

In Scenario 1, with four variables, 45% of farms in the category of family farms belong to Cluster 1. The average number of cows in this cluster (7.70) is smaller than the average number of cows in the Republic of Croatia (14.78), the average annual milk delivery (30107 kg) is lower than the national average of 73275 kg, the average area of land per farm (10.67 ha) is also smaller than the national average. This implies that cluster 1 consists of very small dairy farms. In some cases, these are also semi-subsistence farms. Cluster 2 is also relatively similar to Cluster 1, but farms in Cluster 2 have much more land and cash crops than farms in Cluster 1. Namely, there are many farms with a few cows and much arable land, and these are usually not farming whose primary activity is milk production. We observe significant variability in the number of cows (NOC) across individual clusters, with relatively common minimum values, particularly within the first five clusters and partially in the sixth cluster.

However, the average values vary considerably, indicating distinct farming technologies and production intensity. Notably, clusters four through six comprise farms involved in milk production on a semi-professional level, despite the relatively low average production per dairy cow, which limits the scope for major investments needed for further growth and development of these farms. From cluster 6 onward, the milk yields of the dairy cows are sufficient to indicate production development potential. This is also noted by Žgajnar and Kavčič (2024) in their study on Slovenian dairy farms. Additionally, livestock density tends to be higher in these clusters. The cultivated areas stand out in certain clusters (2, 3, and 5), reflecting a significantly larger amount of cultivation and, consequently, a significantly lower livestock density.

This suggests that within these clusters, there are also farms where milk production is not the primary economic activity. However, on a certain share of farms, the rationale may also stem from extensive fodder production and low livestock density. To substantiate the latter, additional analyses will be necessary, wherein we will endeavour to assess the impact of the region. Furthermore, based on available data from other sources and expert assessment, we aim to determine the production potential of the land. For the stated reason, the variables AUC and NOPC are excluded from the analysis in Scenario 2 with two variables. The two largest farms are in special clusters (clusters 9 and 10). In both cases, the number of cows is very high. The essential difference between them is that one represents intensive management, while the other, on the contrary, exemplifies extensive cultivation with a significantly lower milk yield. There are two types of dairy farms, as mentioned by Gonzalez-Mejia et al. (2018) in their study, which differ in intensity, and each strives for profitability in milk production in its way.

Table 2
Cluster structure – Scenario 1 - Family dairy farms

Clusters	Variables								
	Number of farms	Average NOC	Min NOC	Max NOC	Average ADOM	Average NOPC	Average AUC	Yield per cow	Land area per cow
1	1496	7.70	1	40	30107.85	4.56	10.67	3973.20	2.09
2	873	8.41	1	25	33634.14	7.84	15.13	4020.36	2.70
3	310	17.85	1	53	77299.30	10.24	44.31	4409.71	3.57
4	464	24.99	6	56	122814.55	5.96	34.39	5096.74	1.69
5	25	46.52	8	111	216429.72	8.80	220.43	4831.20	8.04
6	124	53.60	23	124	342978.66	7.10	70.73	6569.65	1.45
7	27	98.26	64	157	738 311.22	6.78	110.40	7793.64	1.19
8	10	201.50	153	238	1543495.06	6.80	192.35	7671.41	0.95
9	1	317.00	317	317	2799071.00	8.00	469.05	8829.88	1.48
10	1	456.00	456	456	2560156.00	6.00	370.27	5614.38	0.81
Σ	**3 331**	**14.78**	**1**	**456**	**73257.69**	**6.29**	**23.68**	**4329.80**	**2.34**

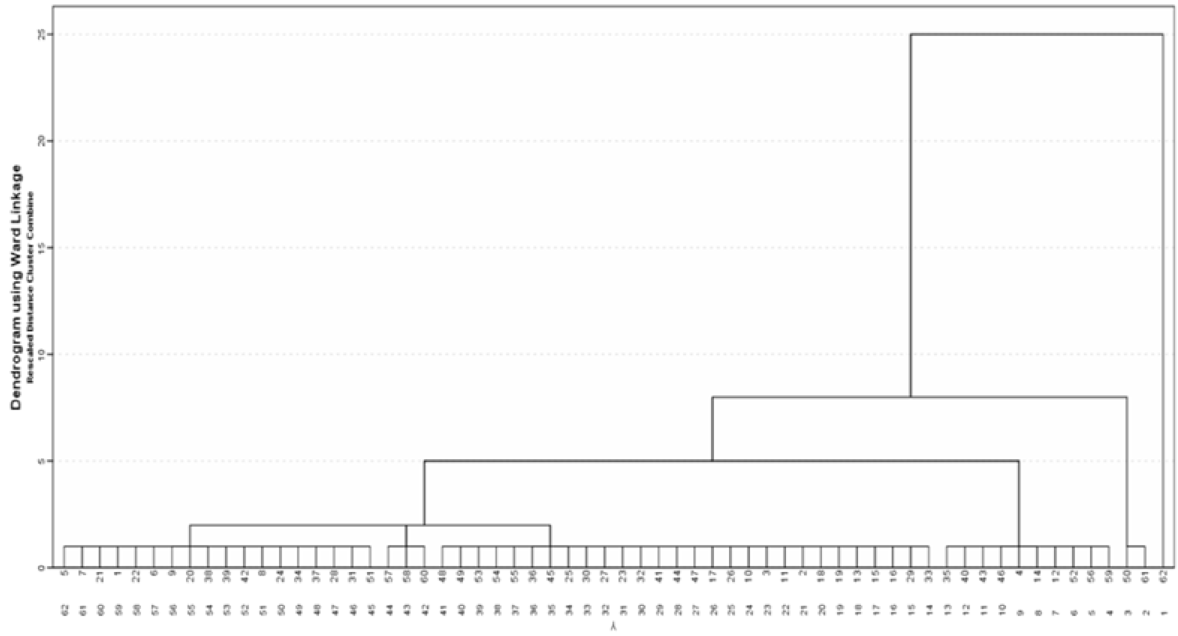
Source: Authors work

Legal entity in dairy production

In the milk processing sector, larger farms are of key importance in Croatia. However, there are also large differences between them in terms of the number of dairy cows in cultivated areas and production potential. The dendrogram for legal entity suggests 6 clusters (Figure 2). Following the hierarchical analysis (dendrogram), a non-hierarchical algorithm (k-means) was applied to the data, limited to a maximum of 6 clusters. Table 3 shows the structure of all clusters in the category of legal entity after the implementation of k-means.

In scenario 1, considering four variables, 61% of farms in the category of legal entities belong to cluster 1. The average number of cows in this cluster (66) is lower than the average number of cows of legal entities in the Republic of Croatia (338). In this cluster, the average annual milk delivery is 436356 kg. This implies that cluster 1 consists of relatively small agricultural holdings. So, it is quite like cluster 4 in family farms (Table 2), except that these farms achieve a significantly lower lactation milk yield per cow (5627 kg). On these farms, the intensity of production is also modest, which is also reflected in the lower livestock density. However, it is notable that, aside from the first cluster, there are no significant differences in the average milk yield achieved in other clusters. Cluster 2 contains farms with a higher average number of cows (246 cows) in a herd. Cluster 3 is a large dairy farm that, due to its technology, has a significantly higher milk production per cow (9073 kg). The three largest farms in the country are each in special clusters (clusters 4 and 6). These are farms with over 2000 cows, and they belong to the category of largest farms on the EU scale (Parzonko et al., 2024). The fourth cluster is interesting, as it has a distinct livestock density. Given the higher milk production, this suggests that breeding relies on purchased fodder. The largest farm has over 4000 cows and, unlike the farms in cluster 4, has much arable land (101428 ha), and a lot of different crops are cultivated on those lands (21). Hence, the last cluster represents a distinct type, engaging not only in milk production but also in crop production. This is evident from the large surface area per dairy cow, indicating a dual focus on both dairy farming and crop cultivation (Table 3).

Figure 2
Dendrogram – Scenario 1 – Legal entity



Source: Author's work

Table 3
Cluster structure – Scenario 1 - Legal entity

Clusters	Variables							
	Number of farms	**Average NOC**	**Min NOC**	**Max NOC**	**Average ADOM**	**Average NOPC**	**Average AUC**	**Yield per cow**	**Land area per cow**
1	38	66.42	1	227	436355.81	4.74	97.33	5626.9	1.68
2	11	245.91	11	754	1996714.73	10	634.38	6562.2	3.36
3	9	631.44	387	1332	5713704.89	5.11	238.35	9073.1	0.48
4	2	2 406.5	2057	2756	21518697	5.5	251.14	8896.6	0.11
5	1	1164	1164	1164	9297902	10	13971.76	7987.9	12
6	1	4 051	4051	4.05	34601195	21	101427.9	8541.4	25.04
Σ	**62**	**337.74**	**1**	**4051**	**2853309.96**	**6.1**	**2076.19**	**6483.7**	**2.3**

Source: Author's work

Scenario 2: Two variables

Family dairy farms

In this scenario, we present the results obtained when only two variables were considered in the analysis: the number of dairy cows (NOC) and the annual delivery of milk (ADOM). A dendrogram was constructed using the same procedure as in Scenario 1, followed by the implementation of the k-means algorithm.

Table 4 shows the structure of all clusters after the implementation of k-means for family dairy farms. Although the dendrogram suggested 7 clusters, the analysis was done with 10 clusters for the sake of comparison with Scenario 1. The obtained analysis (see Table 4) shows that the variable NOC for obtained clusters has a smaller range of disjoint intervals. The same holds for the ADOM variable since it is a correlation variable.

This analysis shows that we get a different set of types of typical farms (Table 4) because this focus is on milk production, it is expected that individual clusters follow the logic of increasing herds, where they are more homogeneous than in the scenario 1 of the first analysis (Table 2). Even in scenario 2, cluster 1 is numerically the most represented, with an average number of 5.7 cows in a herd. It is a category of dairy farms that largely belongs to self-sufficient farms with milk production of 3802 kg of milk. It is also a relatively homogeneous category, with a maximum of 15 animals in the herd. This is a special category that is expected to be abandoned in future structural changes. Appropriate solutions will need to be found to slow down this trend, as these farms are important both from the point of view of social and environmental sustainability. Interestingly, in scenario 2 with two variables, cluster 2 is quite different.

Table 4
Cluster structure – Scenario 2 - Family dairy farms

Clusters	Variables								
	Number of farms	Average NOC	Min NOC	Max NOC	Average ADOM	Average NOPC	Average AUC	Yield per cow	Land area per cow
1	1.889	5.70	1	15	20374.81	5.95	12.98	3801.96	2.96
2	994	16.77	7	32	73871.13	6.68	27.01	4640.77	1.64
3	304	33.27	16	59	168547.71	6.82	43.52	5328.34	1.31
4	101	58.07	30	111	373897.32	7.10	74.78	6719.71	1.30
5	30	96.50	56	157	699186.73	6.87	130.98	7645.72	1.42
6	6	182.17	153	204	1275357.46	7.17	136.78	7197.46	0.73
7	3	204.67	200	214	1636 69.67	7.00	247.45	7999.35	1.22
8	2	230.50	223	238	2103315.93	4.50	204.51	9123.60	0.88
9	1	317.00	317	317	2799071.00	8.00	469.05	8829.88	1.48
10	1	456.00	456	456	2560156.00	6.00	370.27	5614.38	0.81
Σ	**3331**	**14.78**	**1**	**456**	**73257.69**	**6.29**	**23.68**	**4329.80**	**2.34**

Source: Author's work

It is a category of medium-sized farms, which, like cluster 3, includes dairy farms that will face the challenge of restructuring soon. In both categories (2 and 3), we can expect that there will continue to be a cessation of production on one side and, for those with sufficient production resources, an increase in and maintenance of milk production. For the latter, greater investments and appropriate incentives will be needed so that agricultural holdings can follow the trend and achieve higher productivity. This is especially true for cluster 3, where the average production is over 5000 kg of milk. Clusters 4 and 5 combine medium-sized dairy farms with average 58 and 96 dairy cows. It is a type of family farm that is professionally engaged in farming and has development potential. This is also confirmed by the findings of Žgajnar and Kavčič (2024), who analysed a similar type of farm under Slovenian conditions. The next clusters (5 to 10) are numerically rather modestly represented. However, except for the last type of cluster, which includes only one farm, they are significant from the perspective of both the number of cows and the annual delivery of milk.

Legal entity in dairy production
Table 5 displays the composition of all clusters after the implementation of the k-means algorithm for legal entities. Although the dendrogram suggested 5 clusters, an analysis was conducted with 6 clusters for comparison with Scenario 1.

Table 5
Cluster structure – Scenario 2 - Legal entity

Clusters	Variables								
	Number of farms	Average NOC	Min NOC	Max NOC	Average ADOM	Average NOPC	Average AUC	Yield per cow	Land area per cow
1	45	71.73	1	275	467688.46	5.51	135.47	5574.9	2.01
2	12	529.33	333	824	4806974.92	7.08	556.52	9029.6	1.22
3	2	1248	1164	1332	10268474.5	6.5	7009.66	8212.8	6.02
4	1	2057	2057	2057	17657987	5	300.83	8584.3	0.15
5	1	2756	2756	2756	25379407	6	201.45	9208.8	0.07
6	1	4051	4051	4051	34601195	21	101427.9	8541.4	25.04
Σ	**62**	**337.74**	**1**	**4051**	**2853309.96**	**6.1**	**2076.19**	**6483.7**	**2.3**

Source: Author's work

In scenario 2, interesting types of legal entities emerge. Similar to the first scenario (Table 3), the most strongly represented farms are in cluster 1 (Table 5), where the average number of dairy cows in the herd does not deviate significantly from the first analysis, nor does the average milk yield per dairy cow (Table 5). However, it is a fairly variable category depending on the number of dairy cows in the herd. Similar to the first scenario, it will be necessary to establish at least two types of typical farms for this category in the future analysis when developing the starting points for typical farms in the model, as the average value does not adequately represent the variability among farms. The upper quartile, which holds greater significance for milk processing, is certainly of interest. In scenario 2, cluster 2 is similar to cluster 3 of the scenario 1(Table 3). In clusters 3 to 6, there are distinctly large agricultural holdings. There are no significant differences compared to the first analysis.

Conclusion

Summary of research

By utilizing cluster analysis on the data of dairy farms in Croatia, we aimed to identify the characteristics of basic types of dairy farms. This information will facilitate a deeper understanding of the milk production sector in Croatia. Therefore, the results of this research, as suggested by Gonzalez-Mejia et al. (2018), can be used for further research to model scenarios including economic components (e.g. gross margin per cow, gross margin per litre of milk, gross margin per ha), social aspect (e.g. labour) and environmental impacts (e.g. land use per cow, resource depletion, GHG).

The results obtained confirm the distinct polarity of farms in milk production observed in Croatia. In the initial phase, we examined the available data, which encompasses four variables. However, due to the presence of highly atypical dairy farms in this sector (such as those with a small herd and share of milk production but an extremely large area of land), we analyzed two scenarios, with a substantial variation in the variables included. It emerged that the analysis considering solely the herd size and total milk production at the annual level is more rational and yields more homogeneous groups of dairy businesses. This corresponds to scenario 2, focusing on two variables. In certain cases, however, Scenario 1 (four variables) indicates specific types of farms where there is a significant amount of land devoted to cash crops rather than milk production. The latter is less important from the perspective of analyzing the dairy sector, particularly concerning small herds. This is a special category that is expected to be abandoned in future structural changes or transformed into cash crop production. This coincides with the findings of Mijić and Bobić (2021). However, there are also cases where the volume of milking production remains significant, and these farms often have diversified production plans as one of the strategies of risk management.

The disadvantage of the database used in this study is, however, the lack of all the necessary data for a precise analysis of a single typical farm using the farm model. In future analysis, it is recommended to extend the time series and, following the example of Klímová et al. (2022), analyse from a temporal perspective.

Theoretical contributions

Based on the analysed data, to encompass the entire sector, it will be necessary to define at least 15 baseline types of dairy farms in Croatia. It is a significantly larger number than defined in previous research for Croatia, but the approach we intend to apply in the continuation of the research requires a more detailed analysis with a larger number of typical farms.

Of these 15 baseline types, two-thirds will be family farms, ranging from small self-sufficient farms to medium-sized and large family farms. These types are crucial from both social and environmental sustainability perspectives. From the perspective of structural changes, we anticipate that these types of farms will undergo the most significant transformations in the future. This will likely involve consolidation, reflected in a declining number of farms. This is also noted to a certain extent by Mijić and Bobić (2021).

A distinct category will be agricultural companies dealing with milk production, which serve as the pillar of the milk processing sector in Croatia. The analysis suggests that it would be sufficient to have six such production types in this category.

In the next step of defining farm types, given the diverse growing conditions, it would be prudent to include the influence of the region, as conditions vary significantly across regions. Additional indicators, such as the scope of grass, fodder, and cash crops, should be used to provide information on land use and feeding strategies, helping to characterize farms more accurately.

Practical implications

The results of this analysis thus represent an important starting point for further analysis of the dairy sector in Croatia. These will be typical farms that will be defined in more detail at workshops with consultants and experts in the field and will be further adjusted and upgraded with the Slovenian farm model SiTFarm (Pečnik and Žgajnar, 2022). This model will help to evaluate various economic, technological, and environmental indicators with predetermined production constraints.

Indeed, the diverse and varied category of agricultural holdings, as illustrated in this study, achieve very different levels of production efficiency, and face distinct challenges. Given the number of farms in each of the groups, our results corroborate the findings of Očić et al. (2023), who identify the limited number of competitive farms as the primary challenge facing the Croatian dairy sector.

Consequently, addressing these challenges will require tailored agricultural policy measures suited to the specific needs of each category as well as the type of farm. In such a manner, certain policy measures that would increase the profitability of one typical farm do not mean that they would increase the profitability of another farm. So, the same measure is not equally effective for diverse categories of dairy farms. The key question here revolves around identifying which agricultural policy measures effectively address the primary challenges faced by farms, with the aim of achieving greater economic, environmental, and social sustainability. This also suggests a potential improvement in monitoring the effectiveness of individual measures at the farm level, similar to the approach taken by Klímová et al. (2022) in evaluating public support for innovation and measuring changes in the turnover of funded companies. By analysing the time series, we could determine which measures achieve the greatest effectiveness for each type of intervention. Such results could be useful to policymakers in getting information on dairy farm needs, such as which type of measure they need and how much they can adapt to the given situation. Certainly, it will be necessary to support certain types of family farms with appropriate investments to improve efficiency and enhance both economic and environmental sustainability. In the long term, such an approach will help reduce dependence on budgetary payments and increase overall sustainability.

References

1. Alvarez, S., Timler, C. J., Michalscheck, M., Paas, W., Descheemaeker, K., Tittonell, P., Andersson, J. A., & Groot, J. C. J. (2018). Capturing farm diversity with hypothesis-based typologies: An innovative methodological framework for farming system typology development. *PLOS ONE, 13*(5), e0194757. https://doi.org/10.1371/journal.pone.0194757
2. Ciaian, P., Espinosa, M., Gomez Y. Paloma, S., Heckelei, T., Langrell, S., Louhichi, K., Sckokai, P., Thomas, A., & Vard, T. (2013). *Farm level modelling of the CAP: a methodological overview*. Publications Office of the European Union.
3. Gonzalez-Mejia, A., Styles, D., Wilson, P., & Gibbons, J. (2018). Metrics and methods for characterizing dairy farm intensification using farm survey data. *PLOS ONE, 13*(5), e0195286. https://doi.org/10.1371/journal.pone.0195286
4. Klímová, V., Žítek, V., & Lelková, T. (2022). Public Support for Innovation: Changes in Turnover of Granted Companies. *Business Systems Research Journal, 13*(1), 120-137. https://doi.org/10.2478/bsrj-2022-0008
5. Lovec, M., Šumrada, T., & Erjavec, E. (2020). New CAP Delivery Model, Old Issues. *Intereconomics, 55*(2), 112-119. https://doi.org/10.1007/s10272-020-0880-6
6. Mijić, P., Bobić, T. (2021). Milk production and challenges in transition from conventional to robotic milking in Croatia. *Acta Scientiarum Polonorum Zootechnica, 20*(3), 59-64. https://doi.org/10.21005/asp.2021.20.3.08
7. Očić, V., Šakić Bobić, B., Grgić, Z. (2023). Economic analysis of specialized dairy farms in Croatia according to FADN. *Mljekarstvo, 73*(1), 50-58. https://doi.org/10.15567/mljekarstvo.2023.0106
8. Parzonko, A., Parzonko, A. J., & Bórawski, P. (2024). Economic Competitiveness of Dairy Farms from the Top Milk-Producing Countries in the EU: Assessment in 2014-2021. *Agriculture, 14*(1), 123. https://doi.org/10.3390/agriculture14010123
9. Pečnik, Ž., & Žgajnar, J. (2022). Resilience of dairy farms measured through production plan adjustments. *Journal of Central European Agriculture, 23*(1), 207-219. https://doi.org/10.5513/jcea01/23.1.3372
10. Poczta, W., Średzińska, J., & Chenczke, M. (2020). Economic Situation of Dairy Farms in Identified Clusters of European Union Countries. *Agriculture, 10*(4), 92. https://doi.org/10.3390/agriculture10040092
11. Scitovski, R., & Sabo, K. (2020). Cluster analysis and recognition of geometric objects (in Croatian). Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku
12. Žgajnar, J., & Kavčič, S. (2024). Understanding the Dairy Sector in Slovenia: A Modeling Approach for Policy Evaluation and Decision Support. *Sustainability, 16*(14), 6009. https://doi.org/10.3390/su16146009

About the authors

Maja Petrač is a university specialist in actuarial mathematics, employed at the Faculty of Agrobiotechnical Sciences Osijek University of Josip Juraj Strossmayer Osijek (Croatia) as a lecturer in the scientific field of mathematics. She is currently a student in the postgraduate doctoral study of Agroeconomics at the Faculty of Agrobiotechnical Sciences in Osijek, and her research interests are mathematical programming and farm models. The author can be contacted at **mpetrac@fazos.hr**.

Krunoslav Zmaić is a Full professor at the Faculty of Agrobiotechnical Sciences Osijek, University of Josip Juraj Strossmayer Osijek (Croatia). He teaches a course on Agroeconomics and Agricultural and Rural Policy. His areas of scientific interest are the application of economic theory in optimizing the production and distribution of food and fibre products, macro-agroeconomic issues of agricultural development, problems of sustainable development of rural areas, and the evaluation of the impact of agricultural and rural policy in the function of integrated rural development. He is actively involved in several other research projects. The author can be contacted at **kzmaic@fazos.hr.**

Jaka Žgajnar is an Associate Professor at the Biotechnical Faculty, University of Ljubljana (Slovenia). He teaches courses on farm management, animal health economics, and the utilization of mathematical programming techniques in agriculture. His research primarily focuses on various mathematical programming techniques for analyzing decision-making in agriculture. He specializes in farm-level modelling and conducting impact assessment analyses of Common Agricultural Policy measures. He is actively involved in several research projects. The author can be contacted at **jaka.zgajnar@bf.uni-lj.si**.