Computers and Electronics in Agriculture 213 (2023) 108239

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Unsupervised and supervised machine learning approach to assess user readiness levels for precision livestock farming technology adoption in the pig and poultry industries

Kevin Mallinger [a,b,*], Luiza Corpaci [c], Thomas Neubauer [a,b], Ildikó E. Tikász [d], Thomas Banhazi [e]

[a] SBA Research, Floragasse 7/5OG., 1040 Vienna, Austria
[b] TU Wien, Favoritenstraße 7-9, 1040 Vienna, Austria
[c] University of Vienna, Kolingasse 14-16, 1090 Vienna, Austria
[d] Institute of Agricultural Economics Nonprofit Kft., Zsil u. 3-5, H-1093, Budapest, Hungary
[e] National Taiwan University, Center for Biotechnology, No. 1, Sec. 4, Roosevelt Rd., Da'an Dist., Taipei City 106319, Taiwan

ARTICLE INFO

Keywords:
User attitudes
Machine learning
Survey data
Decision tree
Precision livestock farming
Clustering
K-means

ABSTRACT

This study used machine learning, particularly k-means clustering, to identify distinct clusters of users and their technological readiness to adopt various precision livestock farming (PLF) technologies based on their responses to a carefully designed questionnaire. The analysis revealed initially two as well as three distinct clusters representing different levels of technological readiness among farmers considering the adoption of various PLF technologies. In addition to the validation of the cluster results by internal metrics, a related principal component analysis, and a focus group evaluation, this paper describes the application of a Decision Tree as an explainable supervised machine learning approach to investigate the predictive power of specific survey questions. In combination, this research aims to provide valuable insights for understanding farmers' technological readiness, to enhance further survey designs, and to support the development of targeted strategies to promote the successful adoption of PLF technologies in the agricultural sector generally.

1. Introduction

Precision livestock farming (PLF) technologies have the potential to revolutionize the agricultural sector by enhancing productivity, improving animal welfare, and reducing the environmental impact of farming practices (Berckmans, 2017). As the global demand for food continues to grow, adopting PLF technologies becomes increasingly important in achieving sustainable agricultural practices and meeting the needs of a growing population (Araújo et al., 2021). However, the successful implementations of these technologies depend on the willingness and readiness of farmers to introduce them in their farms and the implementation of targeted strategies to address related barriers and concerns of users (Scown et al., 2019).

Despite the diverse benefits of PLF (Banhazi, 2009; Banhazi et al., 2022), poor technology introduction without consideration for users' needs may actually hinder sustainability efforts (Mallinger et al., 2022), which highlights the importance of assessing user attitudes. Additionally, Pathak et al. (2019) stressed the need for further investigations of the multifaceted farmer perspectives towards technology adoption and the factors that influence their technological readiness.

This study aims to address this knowledge gap by applying machine learning techniques (specifically k-means clustering and a Decision Tree classifier) to analyze survey data from 266 farms and investigate distinct clusters[1] of user attitudes towards technological readiness in PLF. Therefore, the main objectives of this study were:

- To unveil farmers' underlying characteristics and attitudes towards PLF technologies, which is essential for tailoring support and educational initiatives that aid technology uptake.
- To identify user characteristics that are particularly important to increase technological readiness.
- To demonstrate the potential of applying machine learning techniques to reveal meaningful patterns in survey data and assess the identified clusters' validity through a combination of validation techniques.
- To evaluate the predictive power of the survey questions in assessing farmers' technological readiness, allowing for the development of more targeted and efficient survey instruments for future research in this area.

* Corresponding author at: SBA Research, Floragasse 7/5OG., 1040 Vienna, Austria.
E-mail address: kmallinger@sba-research.org (K. Mallinger).

[1] *cluster* is the used terminology in the Machine Learning area; a *cluster* is known in other domains as *label*, *grouping*, *level*, *class*, etc.

https://doi.org/10.1016/j.compag.2023.108239
Received 12 June 2023; Received in revised form 11 August 2023; Accepted 7 September 2023
Available online 28 September 2023

By uncovering the different clusters of user attitudes and understanding the factors that influence farmers' technological readiness, this study aims to provide structured information for policymakers, industry stakeholders, businesses, and researchers to facilitate the successful implementation and adoption of precision livestock farming technologies.

2. State of the art

In general, the topic of user attitudes in precision livestock farming is a well-researched topic that is primarily conducted on a survey basis and statistical analysis (Schukat and Heise, 2021b; Boothby and White, 2021; Groher et al., 2020; Makinde, 2020; Abeni et al., 2019; Pathak et al., 2019; Ugochukwu and Phillips, 2018). Such studies showed the various factors that influence farmers' technology decisions, ranging from economic, socio-demographic, ethical, legal, and technological, to institutional aspects that need to be considered for widespread technology acceptance (Makinde, 2020; Drewry et al., 2019; Dhraief et al., 2018).

Although some studies focused on detecting clusters of farmer characteristics before (Schukat and Heise, 2021a), the ability of machine learning to detect individual clusters of user attitudes in precision livestock farming has not been utilized to this day. Thereby, using machine learning to analyze the factors that identify the extent of readiness for technology adoption by farmers is underutilized as well. The usability of such approaches to investigate user characteristics and attitudes has already been demonstrated in domains such as water resource management (Obringer and White, 2023), in social media studies (Kaushik and Bhatia, 2022; Alsayat and El-Sayed, 2016), or in customer segmentation studies (Tabianan et al., 2022). Furthermore, to the authors' knowledge, no studies have been conducted that investigate the effectiveness of survey questions for predicting user characteristics.

As supervised machine learning focuses on finding patterns, relationships between the data, or data separation criteria, its goal is to train a model that enables the prediction of an output when confronted with new data. In doing so, it incorporates technological approaches that are well suited to handle non-linear relationships, need to consider fewer assumptions about data structures and distributions, can compute a wider range of inputs (e.g., images, videos, sound), and do not need a prior hypothesis/theory to validate the results. In this context, the usability of machine learning for social sciences has been shown in various contexts, such as causal inference, theory development, or policy predictions (Molina and Garip, 2019). Focusing on explainable methods can also provide novel insights into the underlying data structures and dynamics and support exploratory analysis of respective research data. In this context, it provides novel opportunities and research directives in this field, thereby supporting other modeling approaches such as diffusion of innovation (Shang et al., 2021), reasoned action (Sok et al., 2021), or planned behavior (Mingolla et al., 2019).

3. Materials and methods

The process of assessing and validating the cluster qualities that represent user attitudes and farm infrastructure towards readiness of PLF adoption is shown in Fig. 1. This includes the whole data processing pipeline, starting from data aggregation and data cleaning right through the validation processes of applied clustering approaches. Respective steps of the experiment design will be introduced chronologically, as seen in Fig. 1.

More generally, this research is embedded in the LivestockSense project, with the goal of providing a functional prototype that assesses and predicts the technological readiness of farmers. The goal of the machine learning approach was to define class boundaries (clusters) that represent technological readiness and provide a model that is able to classify new user questionnaires correctly. These questionnaires are available on the project homepage.[2]

3.1. Survey and data

This investigation builds on the survey data collected by the research team of the LivestockSense project. The survey utilized a self-reported questionnaire to gather data from 266 farms in five European and one Middle Eastern countries from the pig (121 samples) and poultry (145 samples) industries. The questions were designed to capture information about existing infrastructure and attitudes towards smart devices/technologies in farming practices. Several different perspectives have been used in the survey design to associate answers with technological readiness, including the availability of infrastructure (question block 1, 2, and 6), the overall presence of expert knowledge, and market access of PLF technologies (question block 4), as well as the mental attitude of PLF technology users (as expressed in the question block 3 and 5). In total, 20 questions have been chosen that represent the technological readiness of farmers that consider adopting precision livestock farming technologies. Sub-questions were consolidated into a single feature with responses rated either on a 5-point scale, 4-point scale, or 3-point scale to indicate the level of agreement with the statement, in which 1 represented "Strongly disagree" and 5 depicted "Strongly agree" for example. The complete list of questions used for the analysis can be found in Annex.

3.2. Clustering

This study uses the k-means approach for clustering the results. This approach is categorized as an unsupervised clustering algorithm commonly used to identify groups or patterns within a dataset. The algorithm takes a set of measurements, where each observation is an n-dimensional vector, and partitions them into k clusters (where $k \leq n$, n representing the total farms contained in the survey) based on their similarity (Jain, 2010). K-means clustering aims to minimize the within-cluster sum of squares, which is the sum of the squared distance between each data point and its assigned cluster center. In other words, the algorithm aims to find the number of (k) centers that minimize the distance between each point and its assigned center, resulting in clusters that are as compact and distinct as possible. Mathematically, the objective function of k-means is:

$$arg \min_S \sum_{i=1}^{k} \sum_{x \in S_i} \|x_i - c_i\|^2 \quad (1)$$

where c_i is the centroid of the points in S_i, and x_i is the individual data point that belongs to the cluster S_i. Therefore, $\|x_i - c_i\|^2$ represents the squared Euclidean distance between a given data point x_i and the centroid c_i of the cluster S_i that x_i belongs to.

3.3. Validation

The validation process consisted of a multi-step mixed-method approach including quantitative (Principal Component Analysis, Internal Validation Metrics, Decision Trees) and qualitative evaluation phases (Focus Group). This approach was chosen to ensure that the cluster quality is not only measured based on a mathematical basis that usually includes distance-based metrics but also represents qualities of user attitudes that are not directly measurable. Thereby, the validation process was iteratively conducted to ensure that the survey questions and cluster results convey user attitudes and technological readiness characteristics.

To assess cluster quality, several **internal validation metrics** have been used:

[2] https://livestocksense.eu/

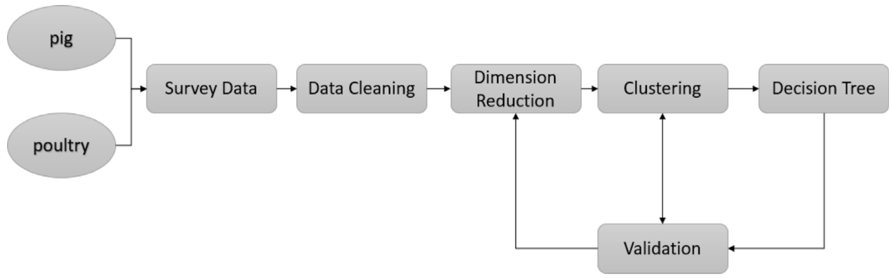


Fig. 1. Systemic organigram of the research experiment design used in this study. The validation step was an iterative procedure and incorporated qualitative and quantitative methods.

– **Davies–Bouldin Index**: measures the average similarity between clusters, in which similarity is the ratio of within-cluster distances to between-cluster distances. Lower values represent a higher clustering quality (Davies and Bouldin, 1979).
– **Calinski–Harabasz Index**: calculates the ratio of the sum of between-cluster dispersion to the sum of within-cluster dispersion. The index is the ratio of the sum of between-clusters distribution and of within-cluster dispersion for all clusters (whereas dispersion is defined as the squared sum of distances) (Caliński and Harabasz, 1974). A higher value indicates a model with better-defined clusters.
– **Silhouette Score**: measures how well each data point fits within its cluster by comparing the average distance between the data point and all other points in its cluster with the average distance between the data point and all other points in the nearest neighboring cluster. The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample (Rousseeuw, 1987). The best value is 1, and the worst value is −1. Values around 0 indicate overlapping clusters, while negative values indicate that a sample has been assigned to the wrong cluster.

The validation process also includes the **Principal Component Analysis** (PCA) to confirm the k-means clustering results and visualize the separation of identified clusters. PCA is a widely-used dimensionality reduction technique that aims to transform the original high-dimensional feature space into a lower-dimensional space while preserving as much of the data's variance as possible. The transformed data is represented by linear combinations of the original features, called principal components (PCs) (Abdi and Williams, 2010). This method was used to reduce the dimensionality of the survey data, enabling visualization of the identified clusters in a two-dimensional space. By plotting the data points along the first two principal components, the separation between the clusters and the relationships between the different user attitudes towards technological readiness in adopting PLF technologies will be assessed.

Finally, a **Focus Group** has been tasked to evaluate the results alongside the entire experiment pipeline continuously. The focus group consisted of six experts from four different countries in the field of livestock farming, PLF technology development, and survey design. This process was conducted iteratively after every significant processing step, ensuring the logical coherence of chosen features and clustering results.

3.4. Decision Tree Classifier

This study employs a Decision Tree Classifier to (1.) investigate the predictive power of the survey questions in assessing the technological readiness of the farmers, as represented by the cluster labels obtained from the k-means model, (2.) provide a model that enables the prediction of farmer attitudes, and (3.) to provide an explainable model such as the decision process can be tracked through the tree, and the prediction can be explained based on the contributions added at each decision node (Maimon and Rokach, 2014). Functionally, a Decision Tree recursively splits the features (survey questions) into subsets with the goal of minimizing the impurity of the resulting subsets, thereby, finding features that separate the clusters best. The impurity measure used in this study is entropy, which is defined as:

$$Entropy(t) = -\sum_{i=1}^{c} p_i \log_2 p_i \tag{2}$$

where t represents an arbitrary node in the tree, c is the number of labels, and p_i is the proportion of samples in node t that belong to class i.

By applying the Decision Tree to predict the identified clusters of different technological readiness, our goal is twofold:

– Investigate if a supervised modeling approach can sufficiently map the relationship between the clusters (target variable) and the input variables (survey questions).
– Understand which survey questions provide the most information (decrease the entropy the most) to identify and separate cluster affiliation.

The Decision Tree was set with a train/test split of 85:15, with a continuous random state of 42 to ensure reproducibility. No hyperparameter search was conducted and all calculations have been subject to 5-fold cross-validation. The reported importance of the features has been calculated by the feature importance function of scikit-learn, which reports the total reduction of entropy by a feature over the entire tree. All the predictions will be evaluated based on the Accuracy metric, which is formulated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

whereas TP represents the True Positive, TN the True Negative, FP the False Positive, and FN the False Negative result.

4. Results and validation

When analyzing the k-means clustering results based on the internal validation metrics, splitting the data into two clusters provides the best scores for all three metrics (as seen in Fig. 2). Hereby, the Davies–Bouldin score is around 1.58 for two clusters and 1.9 for three clusters. In contrast, the Calinski–Harabasz and the Silhouette scores indicate better separation of clusters when the scores are higher. Thereby, Fig. 2 displays that the Silhouette Score reaches a maximum value at $k = 2$ (0.22). In the case of the Silhouette Score in Fig. 3, values close to 1 show good clustering, values close to 0 represent a fuzzy classification, and negative values are representative of misclassification. In all four analyzed cases, some elements are badly classified. The figure indicates that the smallest fluctuations in the size of the clusters appear when choosing $k = 2$ clusters. Overall, all three metrics indicate that two clusters provide a good separation of the data points, with three clusters being the next best split. Based on the Silhouette score in Fig. 3, the limit of valid clusters is set with 5, as the cluster sizes get too small at this point.

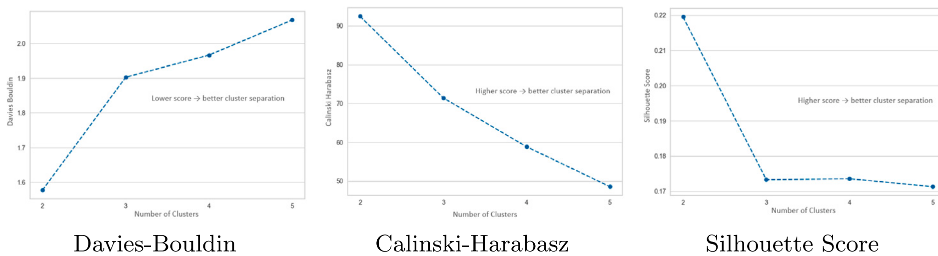


Fig. 2. Internal validation scores for different cluster sizes (k values on the x-axis). Lower scores for the Davies–Bouldin score indicate better cluster separation, whereas higher scores for the Calinski–Harabasz and Silhouette Score indicate more favorable separation boundaries.

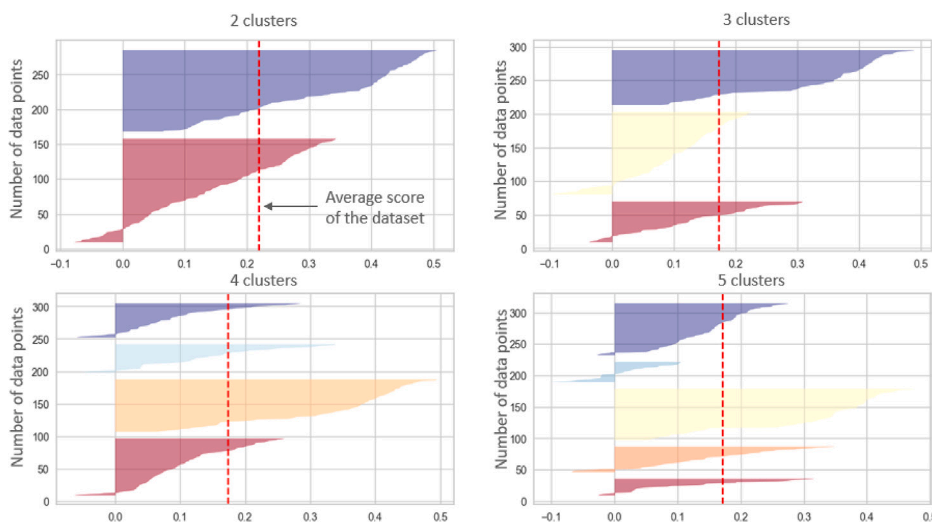


Fig. 3. Silhouette Score (x-axis) for different cluster sizes (k values on y-axis). The red line indicates the average silhouette score, representing a critical threshold for cluster validity.

Although two clusters provide the best separation based on the analyzed metrics, our research is particularly interested in the user characteristics of farmers that are close to technology adoption but may need additional support in certain categories. For this, a third group of farmers was identified with moderate attitudes towards technology adoption, which was further emphasized based on the expert group evaluation, the PCA results as seen in Fig. 8, and the Decision Tree in Table 3. Table 2 and Fig. 8 highlight that the data points for this third cluster were consistently situated between the two main clusters, indicating a moderate level of agreement with both positive and negative attitudes towards technology adoption, thereby, highlighting user characteristics that are relevant for the perception of PLF technology.

4.1. Farmer characteristics of 2 clusters

The analysis of the survey data revealed distinct patterns and variations in farmers' attitudes towards smart devices/technologies and their technological readiness. Based on the internal validation metrics, 2 clusters were identified that can be interpreted in terms of hierarchical attitudes towards technological adoption: "Ready" and "Not Ready". The density plots in Fig. 5 visually represent the agreement levels among these clusters, highlighting the differences in farmers' perceptions. The "Ready" group exhibited a higher overall level of available infrastructure (question blocks 1, 2, and 6), expert knowledge and market availability towards PLF technologies (question block 4), as well as the mental attitude to use such technologies (as expressed in the question block 3 and 5). The "Not Ready" category displayed lower agreement rates in all six categories. These findings represent a distinct separation of farmers' attitudes towards technological adoption based on two categories (Ready/Not Ready). Principal Component Analysis (PCA) was further used to validate the accuracy of the clustering method. Fig. 4 shows the separation of the data in an underlying dimensional space. By projecting the data onto the principal components, all plots with principal component 1 express the visual separation of two clusters horizontally and vertically, indicating a clear distance of data points between the various dimension (questions). This method provided an additional level of confidence in the accuracy of the applied clustering approach.

A total of 149 farms out of the initial sample of 266 were categorized as "Ready" based on the adopted clustering approach (see Table 1). These farms exhibited a positive attitude towards adopting smart devices and technologies. Specifically, the respondents from the "Ready" farms expressed a high level of agreement to utilize smart products as a means to address the challenges associated with labor shortage ($Q3_a$, $\mu = 4.213$, $\sigma = 1.032$), for enhancing day-to-day decision-making in the livestock buildings ($Q3_b$, $\mu = 4.333$, $\sigma = 0.919$), as well as for helping with the environmental pollution reduction obligations ($Q3_d$, $\mu = 4.136$, $\sigma = 1.299$).

Furthermore, the farmers labeled as "Ready" displayed strong agreement regarding the effectiveness of smart devices/technologies in improving production efficiency ($Q3_e$, $\mu = 4.700$, $\sigma = 1.495$). They also agreed that these devices provide real-time information ($Q3_g$, $\mu = 4.700$, $\sigma = 0.672$) and that the provided information is reliable ($Q3_f$, $\mu = 4.470$, $\sigma = 0.676$).

Regarding the accessibility of smart devices/technologies, farmers from the "Ready" group showed stronger agreement in terms of easy access to the market ($Q4_a$, $\mu = 4.547$, $\sigma = 0.835$), easy access to information ($Q4_c$, $\mu = 4.427$, $\sigma = 0.823$), and availability of technical assistance ($Q4_d$, $\mu = 4.094$, $\sigma = 0.798$). However, they expressed a relatively low agreement regarding the availability of proper training for using smart technologies ($Q4_e$, $\mu = 3.965$, $\sigma = 1.306$).

Table 1
Statistical overview of cluster results based on the survey answers (on a scale from 0 to 5) for $k = 2$ clusters. Higher values indicate a stronger agreement with the question, whereas lower values are associated with disagreement.

Feature	Question	"Ready"		"Not Ready"	
		Mean (μ)	Std (σ)	Mean (μ)	Std (σ)
Q1. Average availability of internet access at your farm (Scale: 1–4)		3.085	0.942	2.697	0.875
Q2. Average level of automatization at your production farm (Scale: 0–2)		1.957	0.241	1.879	0.384
Q3. SMART DEVICES/TECHNOLOGY ... (Scale: 1–5)					
Q3_a	... help/support to cope with labor shortage.	4.213	1.032	2.268	1.378
Q3_b	... help/support day-to-day decision making in the livestock buildings.	4.333	0.919	3.335	1.426
Q3_c	... help/support enterprise, marketing and investment decisions.	3.931	1.165	2.906	1.430
Q3_d	... help/support to meet environmental pollution reduction obligations.	4.136	1.299	2.785	1.421
Q3_e	... enable to increase the effectiveness of production.	4.700	0.495	3.791	1.128
Q3_f	... provide reliable information.	4.470	0.676	3.691	1.077
Q3_g	... provide information in a real-time manner.	4.700	0.672	4.013	1.208
Q4. Indicate how much you agree with the following statements: (Scale: 1–5)					
Q4_a	It is easy to access smart technologies on the market.	4.547	0.835	2.852	1.362
Q4_b	Smart technologies can be purchased at an affordable price.	3.367	0.934	1.973	1.114
Q4_c	It is easy to get information on smart technologies and distributors.	4.427	0.823	2.449	1.259
Q4_d	It is easy to get technical assistance to smart technologies.	4.094	0.798	1.979	1.159
Q4_e	Proper education is available for using smart technologies.	3.965	1.306	2.013	1.102
Q5. SMART DEVICES/TECHNOLOGY ... (Scale: 1–5)					
Q5_a	... can be maintained at a reasonable cost.	3.709	1.017	2.322	1.306
Q5_b	... are easy to operate.	4.350	0.812	2.832	1.215
Q5_c	... can be connected well with other equipment/software.	2.905	1.766	2.563	1.406
Q5_d	... operate in a reliable manner.	4.273	0.988	2.738	1.342
Q5_e	... are secure in terms of data management.	3.683	1.164	2.248	1.589
Q6. Do you use smart devices at the farm you represent? (Scale: 1–3)		2.769	0.480	2.328	0.817

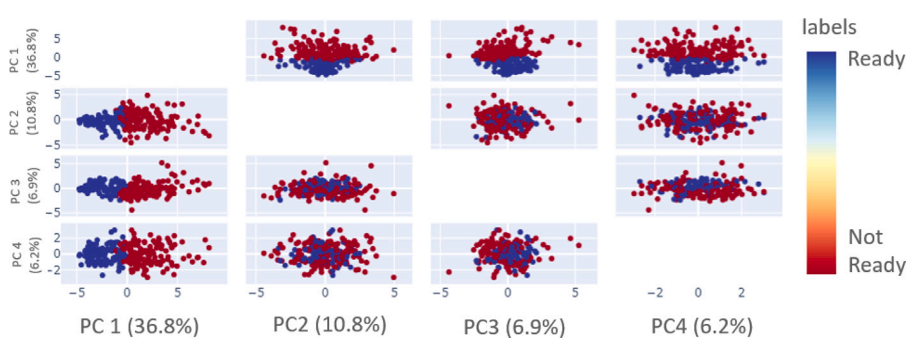


Fig. 4. PCA using k-means labeling with $k = 2$ clusters. Each plot compares the principal components' lower-dimensional space, showing that the clusters clearly separate the groups over the totality of answers. The most important comparison is based on PC1, as this is the principal component that explains the highest amount of variance of the dataset.

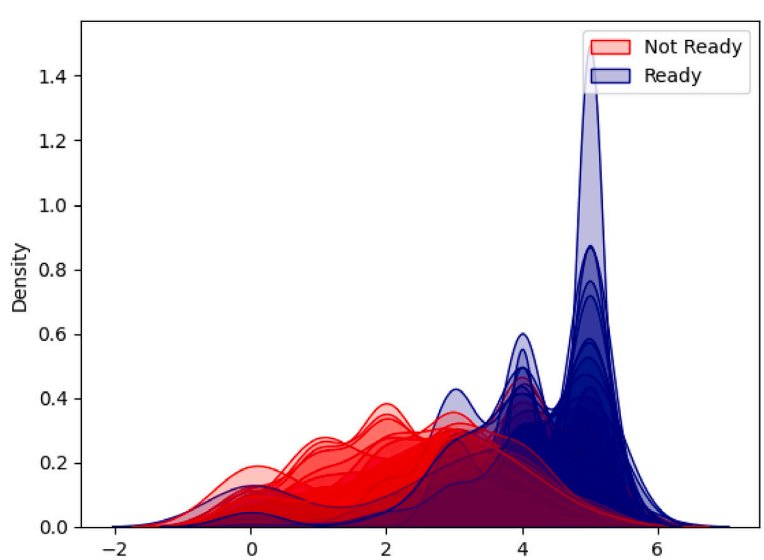


Fig. 5. Density plots for $k = 2$ clusters showing the distribution of all individual answers of the respective groups (blue = ready, red = not ready) within the survey.

Additionally, farmers perceived smart devices/technologies as highly user-friendly (μ = 4.350, σ = 0.812), indicating that they find them easy to operate. Furthermore, these farmers expressed confidence in the reliability of smart devices/technologies ($Q5_d$, μ = 4.273, σ =

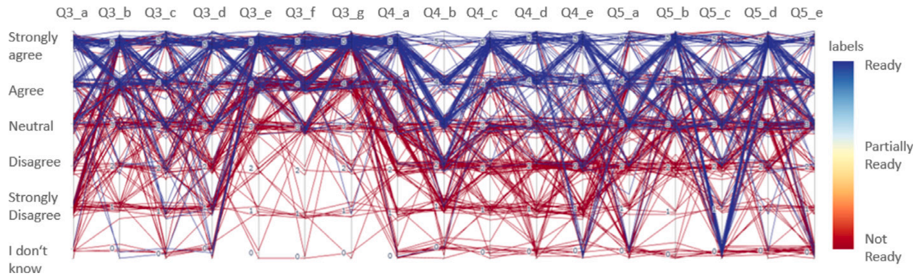


Fig. 6. Parallel coordinates for k-means labeling with $k = 2$ clusters for all answers with a scale from 1–5 (Strongly Disagree... Strongly Agree). The value 0 (I don't know) indicates that the question could not be answered. The colors represent the group affiliation (ready, not ready). Every line represents the answers given by one survey participant, thereby visualizing the connections between responses given by individual participants and the distribution of answers within and between clusters.

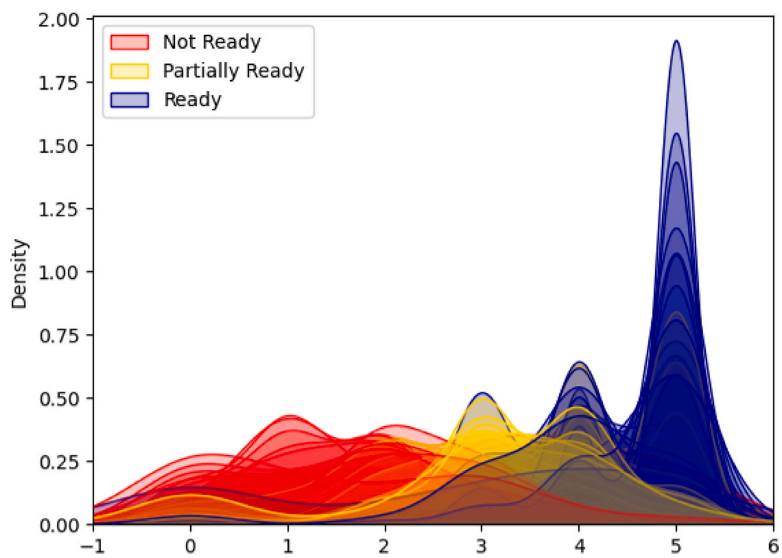


Fig. 7. In this graph, the density plots for $k = 3$ clusters show the distribution of all individual answers of the respective groups (blue = ready, yellow = partially ready, red = not ready) within the survey.

0.988). However, they encountered challenges in maintaining smart devices/technologies at a low cost ($Q5_a$, $\mu = 3.709$, $\sigma = 1.017$).

The analysis reveals that the connectivity of smart devices/ technologies with other equipment and software received the lowest agreement levels among the surveyed farmers ($Q5_c$, $\mu = 2.905$, $\sigma = 1.766$). Furthermore, opinions on the costs associated with smart devices/technologies showed relatively low agreement in the responses ($Q4_b$, $\mu = 3.367$, $\sigma = 0.934$). Lastly, when examining the security aspect of data management, the results also indicate a lower level of agreement than most questions ($Q5_e$, $\mu = 3.683$, $\sigma = 1.164$).

The second cluster, labeled the "Not Ready" group, comprises 117 farms. A lack of consensus regarding the benefits of smart products predominantly characterizes this cluster. It demonstrated the lowest level of support for the affordability of smart devices/technologies ($Q4_b$, $\mu = 1.973$, $\sigma = 1.114$) as well as for the accessibility to technical assistance ($Q4_d$, $\mu = 1.979$, $\sigma = 1.159$). These findings highlight the areas where further attention and improvements are needed to enhance connectivity, address cost concerns, and ensure robust data security in the context of smart devices/technologies adoption in farming.

Fig. 6 gives additional information about the characteristics of the user groups by connecting the answers to the survey provided by every user. It further confirms the findings from Fig. 5 and shows that there is a clear tendency of the "Ready" category to express high levels of agreement towards the questions (blue lines). It also highlighted that the cluster expressed as "Ready" showed several farms that are unsure about the interoperability of smart devices (as seen in question $Q5_c$). The "Not Ready" cluster is expressed in red, indicating the broader variability of answers and the overall low levels of agreement concerning the survey questions.

4.2. Farmer characteristics of 3 clusters

In addition to the previously discussed two clusters, we focus in this section on identifying a user group with moderate attitudes towards technology adoption, potentially providing insights on targeted advice to increase PLF adoption. Figs. 7 and 9 reveal a distinct pattern of responses, with the "Ready" label (blue) displaying a stronger tendency towards positive views on technology, while the "Not Ready" label (red) tends to be associated with less support and more negative views towards these technologies. Figs. 7 and 9 also highlight that there is a third distinct cluster between the two opposing attitudes with a high and low agreement to the survey questions ("Partially Ready" in yellow). Every distribution in this plot represents a distinct category of the survey, showing that the average answer of that category is in the range between 3 and 4 (1 being the lowest and 5 being the highest). The PCA further confirmed this in Fig. 8. Here we can observe that the k-means model generates an equally clear separation for the three cluster results as seen for two clusters in Fig. 4.

The cluster categorized as the technological "Ready" group showed particularly high agreement with the statement that smart devices support enterprise, marketing, and investment decisions ($Q3_c$, $\mu = 4.698$, $\sigma = 0.489$), that smart devices increase the effectiveness of production ($Q3_e$, $\mu = 4.738$, $\sigma = 0.669$), provide reliable information ($Q3_f$, $\mu = 4.852$, $\sigma = 0.450$), and are easily accessible on the market ($Q4_a$, $\mu = 4.763$, $\sigma = 0.484$). All four questions expressed low standard deviations, further emphasizing the homogeneity of this group.

It could also be shown in Table 2 that there are particularly big gaps between the groups for the questions of market accessibility ($Q4_a$) and

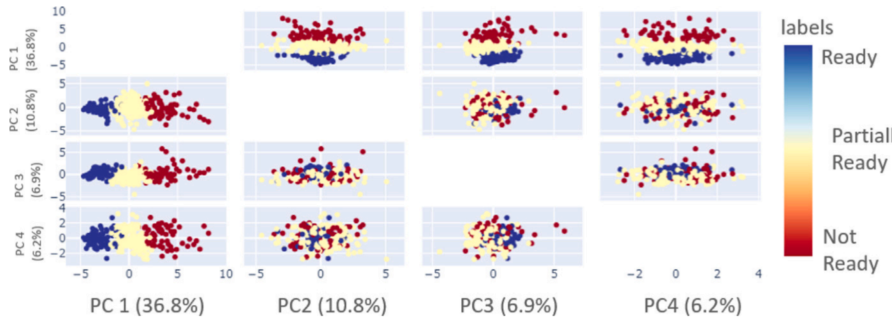


Fig. 8. PCA using k-means labeling with k = 3 clusters. Each plot compares the principal components' lower-dimensional space, showing that the clusters clearly separate the groups over the totality of answers. The most important comparison is based on PC1, as this is the principal component that explains the highest amount of variance of the dataset.

Table 2
Statistical overview of cluster results based on the survey answers for k = 3 clusters. Higher values indicate a stronger agreement with the question, whereas lower values are associated with disagreement.

Feature	Question	Ready		Part. Ready		Not Ready	
		Mean	Std	Mean	Std	Mean	Std
Q1. Average availability of internet access at your farm (Scale: 1–4)		3.225	0.968	2.752	0.900	2.618	0.821
Q2. Average level of automatization at your production farm (Scale: 0–2)		1.960	0.190	1.899	0.367	1.824	0.474
Q3. SMART DEVICES/TECHNOLOGY ... (Scale: 1–5)							
Q3_a	... help/support to cope with labor shortage.	4.065	1.036	3.153	1.469	2.745	1.405
Q3_b	... help/support day-to-day decision making in the livestock buildings.	4.513	0.930	3.148	1.459	2.335	1.354
Q3_c	... help/support enterprise, marketing and investment decisions.	4.698	0.489	4.321	0.842	3.251	1.271
Q3_d	... help/support to meet environmental pollution reduction obligations.	4.587	0.584	4.095	0.779	3.152	1.229
Q3_e	... enable to increase the effectiveness of production.	4.738	0.669	4.479	0.751	3.415	1.473
Q3_f	... provide reliable information.	4.852	0.450	3.505	1.128	2.118	1.380
Q3_g	... provide information in a real-time manner.	3.673	0.733	2.409	1.107	1.507	0.963
Q4. Indicate how much you agree with the following statements (Scale: 1–5)							
Q4_a	It is easy to access smart technologies on the market.	4.763	0.484	3.059	1.120	1.888	1.245
Q4_b	Smart technologies can be purchased at an affordable price.	4.304	0.692	2.746	1.186	1.393	0.923
Q4_c	It is easy to get information on smart technologies and distributors.	4.366	1.075	2.346	1.199	1.872	1.154
Q4_d	It is easy to get technical assistance to smart technologies.	3.819	0.933	2.989	1.238	1.614	1.129
Q4_e	Proper education is available for using smart technologies.	4.631	0.662	3.433	0.946	2.121	1.151
Q5. SMART DEVICES/TECHNOLOGY ... (Scale: 1–5)							
Q5_a	... can be maintained at a reasonable cost.	3.825	0.921	2.983	1.225	1.618	1.131
Q5_b	... are easy to operate.	4.554	0.828	3.444	0.992	1.825	1.272
Q5_c	... can be connected well with other equipment/software.	3.824	1.089	2.952	1.455	1.504	1.459
Q5_d	... operate in a reliable manner.	4.546	0.830	3.446	1.004	1.806	1.246
Q5_e	... are secure in terms of data management.	3.796	1.075	2.942	1.433	1.470	1.246
Q6. Do you use smart devices at the farm you represent? (Scale: 1–3)		2.756	0.494	2.568	0.705	2.110	0.884

affordability ($Q4_b$). For question $Q4_a$, the gap between the ready and not ready category is around 1.7 points, and for question $Q4_b$ around 1.5 points. This indicates that these categories could be potential factors that need to be addressed when fostering the readiness of the intermediary cluster group. The biggest gap between the "Ready" and "Partially Ready" groups found in this analysis was for ($Q4_c$) with around 2 points difference, targeting the accessibility of information about smart devices/technologies. This can potentially be attributed to the fact that more technologically advanced farmers are also more familiar with the sources of information available to them. On the other hand, the "Partially Ready" and "Not Ready" categories might encounter difficulties in identifying reliable sources of information, which could hinder their ability to adopt smart devices/technologies. This highlights the need for accessible and reliable sources of information that can help educate and support farmers in their technological adoption journey.

We can identify only a narrow gap between the "Ready" and "Partially Ready" clusters when asking for potential effectiveness gains ($Q3_e$) while using smart devices. Both groups show overall positive tendencies in this aspect, indicating that this factor may represent a less critical element for the intermediary group. This was also observed for the question about the potential support of smart devices for enterprise, marketing, and investment decisions ($Q3_c$ with a gap of around 0.4 points) as well as for the benefits regarding environmental pollution reduction obligations ($Q3_d$ with a gap of about 0.5). The same pattern could be extracted for the average level of existing automatization on the farms ($Q2$) and the prior use of smart devices ($Q6$). In this regard, questions $Q3_c$, $Q3_d$, $Q3_e$, and $Q6$ show considerable overlap between the distributions of the answers, indicating that these factors individually are not as clear of an indicator for technology adoption as other questions.

Question block four focused on the availability of smart technologies in terms of market access, pricing, and information availability. Table 2 demonstrates that the respondents' answers displayed the biggest spread in this overall category, indicating that this topic separates the individual group opinions well. Hereby, we can observe that the average score of the "Ready" cluster is around 4.38, whereas the average scores for the "Partially Ready" and "Not Ready" clusters are 2.91 and 1.78, respectively. Compared to blocks 4 and 5, the agreement in block three about the expectations of smart devices is generally higher on average. The lowest agreement for the "Not Ready" group can be found for block five, targeting topics like operability, security, and reliability with an average score of 1.63.

When analyzing the "Partially ready" category, we can observe particularly high agreements that PLF technology increases the effectiveness of production ($Q3_e$, $\mu = 4.479$, $\sigma = 0.751$), support enterprise, market, and investment decisions ($Q3_c$, $\mu = 4.321$, $\sigma = 0.842$), and helps to meet environmental pollution reduction obligations ($Q3_d$, $\mu = 4.095$, $\sigma = 0.779$). These three answers also showed low standard deviations (ranging from 0.751 to 0.842), indicating a uniform consensus on the topics.

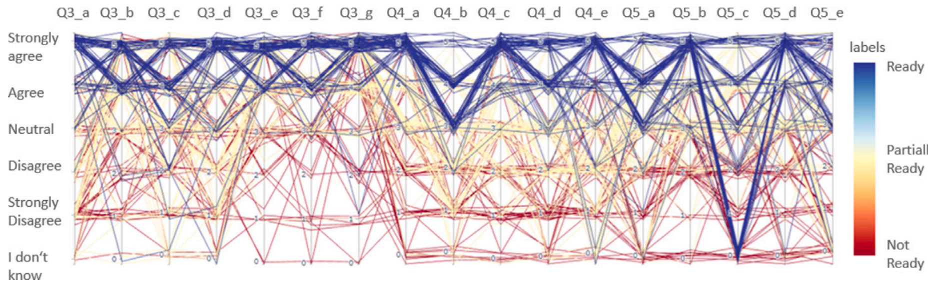


Fig. 9. Parallel coordinates for k-means labeling with $k = 3$ clusters for all answers with a scale from 1–5 (Strongly Disagree... Strongly Agree) on the y-axis. The value 0 indicates that the question could not be answered and the x-axis represents the individual questions. The colors represent the group affiliation (ready, partially ready, not ready). Every line represents the answers given by one survey participant, thereby visualizing the connections between responses given by individual participants and the distribution of answers within and between clusters.

On the other side, the "Partially Ready" group expressed skeptical views on the security of smart devices ($Q5_e$, μ = 2.942, σ = 1.433) as well as the interoperability with other devices ($Q5_c$, μ = 2.952, σ = 1.455). Low statistical values could also be observed when asked about the accessibility of information about PLF technologies ($Q4_c$, μ = 2.346, σ = 1.199), as well as the ability to get technical assistance to smart technologies ($Q4_d$, μ = 2.989, σ = 1.238). All of those questions inhibited higher standard deviations, indicating that the attitude of that group is less homogeneous compared to the positive attitudes towards PLF.

Overall, we can observe in Table 2 and Fig. 7 that the third group (partially ready) exhibits unique characteristics between the ready and not ready group, particularly in response to supporting day-to-day decision making ($Q3_b$), the provision of reliable information in general ($Q3_f$) and in real-time ($Q3_g$), and the accessibility ($Q4_a$) as well as affordability of the technologies ($Q4_b$). These findings further confirm that the 'Partially Ready' group, with their moderate views and diverse perspectives, could be a key target for interventions designed to enhance technological readiness. Further evaluations on the distribution of the answers and the relationships between the topics can be observed in Fig. 9.

4.3. Feature importance

Decision trees allow a deeper understanding of the relationships between the survey questions and the resulting cluster labels. By highlighting the most important survey questions for cluster separation, we can identify questions that provide actionable information about technology adoption, thereby supporting the development and implementation of effective interventions and strategies for promoting innovation in the agricultural sector.

When using the identified clusters as a target to be predicted by the supervised machine learning approach (Decision Tree), we can see that the algorithm can identify the clusters with a high degree of accuracy (see Table 3). To compare against a baseline, the chance of the classifier to predict the correct group by pure chance is also given. Naturally, two clusters are easier to identify than three clusters, as the classifier has a higher chance of randomly picking the correct group. Therefore, the ability to identify three clusters with an average accuracy of 80 percent increases the confidence that the used questions provide informative markers for identifying user attitudes.

Table 4 expresses the importance of the questions for the Decision Tree when predicting the three identified clusters based on the survey questions. Higher values indicate the ability of a question to separate the three identified groups, showing apparent differences between the cluster groups. It was shown that $Q4_a$ separates the groups best, with an absolute score of 0.32. This indicates that the opinion of the three groups on the accessibility of smart technologies vary considerably from each other. This separation was partly identified in the statistical overview in Table 2. This increases the confidence that this feature is an essential indicator for technological readiness and further survey designs. The next most important features include the interoperability of smart devices ($Q5_c$ = 0.15) and the ability to cope with labor shortage ($Q3_a$ = 0.12). It was also shown that the decision tree had not used three questions, implying that these features are not a clear indicator to define technological readiness in our clustering experiments.

Table 3
Average accuracy of the Decision Tree predicting the affiliation of the survey participants to the identified clusters by applying a 5-fold cross-validation. A baseline comparison is given based on the chance of predicting the clusters correctly at random.

No. of clusters	Accuracy (mean ± std)	Baseline
2	0.90 ± 0.08	0.5
3	0.80 ± 0.04	0.33

5. Discussion

This research aims to further expand on the dynamics of farmers' attitudes and technological readiness towards Precision Livestock Farming (PLF) technologies. It is important to note that this study does not directly evaluate attitudes towards technology adoption but investigates the general attitudes of farmers about PLF technologies and their cluster characteristics. In doing so, it highlights perspectives that represent potential barriers to technology adoption, including available infrastructure, access to information, and the economic embedding. In terms of methodological contributions, this study underscores the potential of machine learning techniques in survey analysis. Specifically, k-means clustering enabled the detection of closely connected data points, thereby advancing the understanding of different user characteristics. The interpretation of splitting the groups between "Ready", "Partially Ready", and "Not Ready" towards technology adoption is conducted based on the survey design and the evaluated user characteristics within the respective clusters. Further interpretations of the inherent meaning of the clusters are possible.

The two primary clusters identified, labeled as "Ready" and "Not Ready", express distinct and uniform patterns in their attitudes regarding precision livestock technologies. The "Ready" group, comprising 149 farms, displays a high level of agreement towards the benefits of these technologies, indicating their positive attitudes and readiness for technological adoption. In contrast, the "Not Ready" group, consisting of 117 farms, displays lower agreement rates, suggesting they are less prepared for technological adoption. The principal component analysis further confirmed the distinction between this group, showing the sharp boundary between the two groups on the principal components plots. This is further validated by the supervised machine learning approach that utilizes the survey answers to classify the provided clusters correctly.

Table 4
This table shows the feature importance of the questions for the Decision Tree when predicting the three different groups of technological readiness. The importance is calculated on the overall decrease of entropy by each question. Values that are 0 imply that the Decision Tree did not use this question to predict the groups. The Description of the questions is shortened in order to give a precise overview of the questions used.

Feature	Description	Importance
Q4_a	It is easy to access smart technologies on the market	0.318004
Q5_c	Devices can be connected well with other equipment/software	0.145477
Q3_a	Devices help/support to cope with labor shortage	0.122762
Q5_b	Devices are easy to operate	0.081120
Q3_d	Devices help to meet environmental pollution reduction obligations	0.063745
Q4_b	Smart technologies can be purchased at an affordable price	0.038518
Q4_d	It is easy to get technical assistance to smart technologies	0.035099
Q5_d	Devices operate in a reliable manner	0.034218
Q4_e	Proper education is available for using smart technologies	0.030022
Q3_c	Devices help/support enterprise, marketing and investment decisions	0.027262
Q3_g	Devices provide information in a real-time manner	0.021139
Q5_a	Devices can be maintained at a reasonable cost	0.018552
Q6	Are smart devices already integrated in the farm	0.017988
Q3_e	Devices enable to increase the effectiveness of production	0.017399
Q3_b	Devices support day-to-day decision making in the livestock buildings	0.011721
Q4_c	It is easy to get information on smart technologies and distributors	0.010699
Q5_e	Devices are secure in terms of data management	0.006276
Q2	Availability of internet connection	0.000000
Q3_f	Devices provide reliable information	0.000000
Q1	Average age of existing automatization equipment	0.000000

Table 5
Overview of survey questions used in the research design.

Feature	Question
	1. Please state the average availability of internet access at your farm
	(0: I don't know, 1: No availability.... 4: High availability)
	2. Please state the average level of automatization at your production farm
	(0: I don't know, 1: Less than 10 y/o, 2: 10–20 y/o, 3: diverse, 4: Over 20 y/o)
	3. Please, indicate how much you agree with the statements on smart devices/technologies
	(sensors, cameras robots, farm management information system etc.),
	regardless of whether using them or not in the farm you represent.
	(0: I don't know, 1: Strongly disagree.... 5: Strongly agree)
	SMART DEVICES/TECHNOLOGY...
Q3_a	...help/support to cope with labor shortage.
Q3_b	...help/support day-to-day decision making in the livestock buildings.
Q3_c	...help/support enterprise, marketing and investment decisions.
Q3_d	...help/support to meet environmental pollution reduction obligations.
Q3_e	...enable to increase the effectiveness of production.
Q3_f	...provide reliable information.
Q3_g	...provide information in a real-time manner.
	4. Regarding the availability of smart technologies, please, indicate how much you agree
	with the following statements.
	(0: I don't know, 1: Strongly disagree.... 5: Strongly agree)
Q4_a	It is easy to access smart technologies on the market.
Q4_b	Smart technologies can be purchased at an affordable price.
Q4_c	It is easy to get information on smart technologies and distributors.
Q4_d	It is easy to get technical assistance to smart technologies.
Q4_e	Proper education is available for using smart technologies.
	5. Regarding the operation of smart technologies, please, indicate how much you agree
	with each of the statements.
	(0: I don't know, 1: Strongly disagree.... 5: Strongly agree)
	SMART DEVICES/TECHNOLOGY.....
Q5_a	...can be maintained at a reasonable cost.
Q5_b	...are easy to operate.
Q5_c	...can be connected well with other equipment/software.
Q5_d	...operate in a reliable manner.
Q5_e	...are secure in terms of data management.
	6. Do you use smart devices (sensors, cameras, robots etc.) at the farm you represent? (0: I don't know, 1: Yes. 2: No.)

As this article focuses on identifying user clusters for technological readiness and evaluating informative survey questions, it also highlights important attitudes relevant to technology adoption. The clustering results show that common barriers to technology adoption have been identified, such as the increased security concerns and expected interoperability issues of PLF technologies. This further supports the findings of Drewry et al. (2019), which stated in their study that privacy and security concerns (for farmers in Wisconsin) are one of the most prominent barriers that inhibit digital technology adoption. It also reconfirms the findings of Pivoto et al. (2018) and Boothby and White (2021) that highlighted the need for interoperability among PLF-relevant systems to increase usability and address existing barriers.

By identifying a third group that expressed moderate characteristics of technology adoption readiness, this study highlights areas of interest for targeted interventions towards technology acceptance. Next to the security and interoperability concerns, strategic intervention points based on this analysis include the accessibility of information about PLF technologies as well as the availability of expert knowledge to assist in operating PLF infrastructures. This adheres to closely related perspectives about the complexity of technologies that need targeted

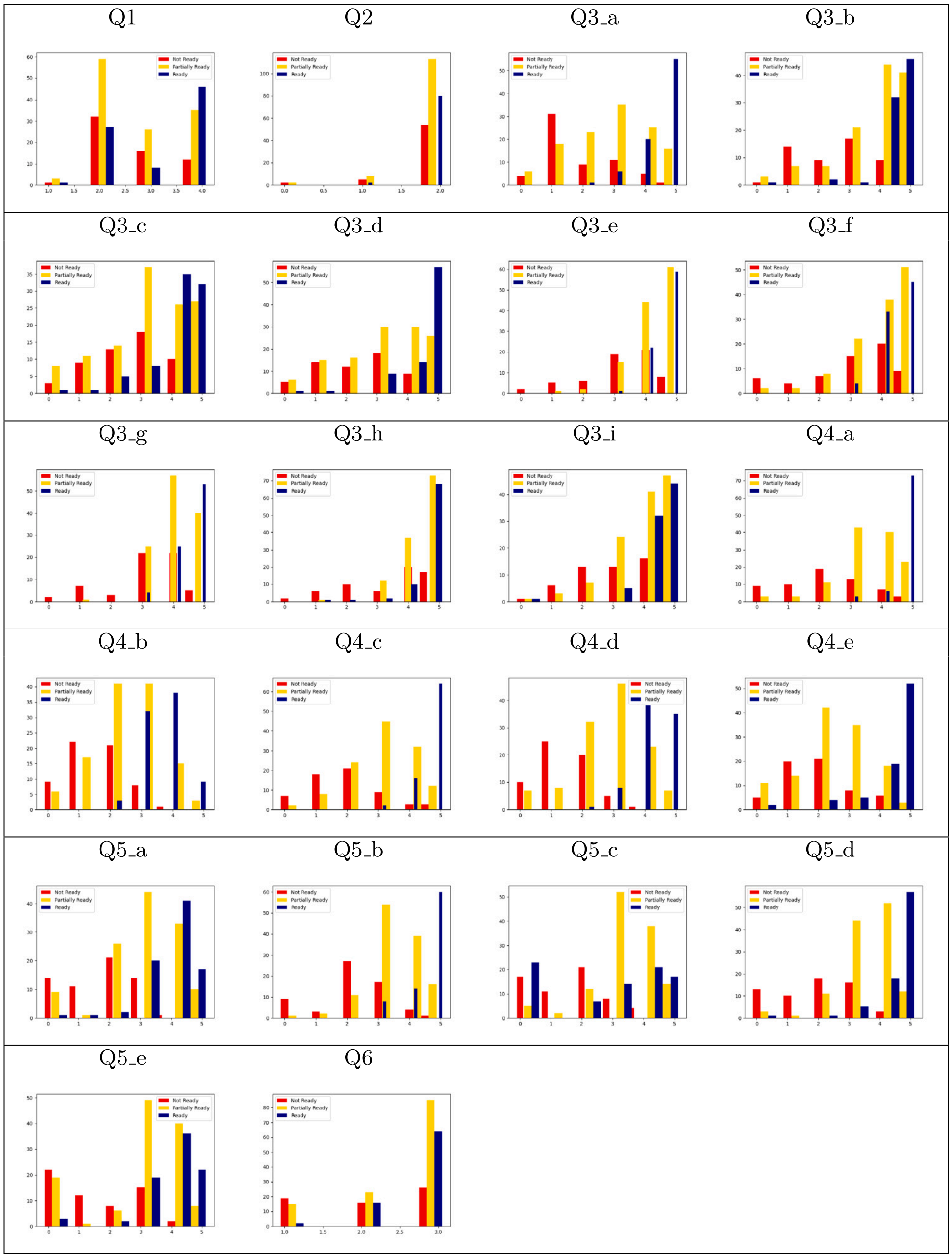
Table 6
Histograms of k-means labels, with $k = 2$ clusters (blue = ready, red = not ready) for each of the chosen questions.

education to install, interpret, and use the technology as well as the availability of and dependencies to service providers and vendors (Isabelle, 2021; Groher et al., 2020; Makinde, 2020; Abeni et al., 2019). By understanding the common concerns of this "potentially ready" cluster, this group could shift towards a more positive attitude and readiness for adopting smart devices/technologies if targeted support would be provided.

The "Not Ready" group was characterized by a lack of consensus on the benefits of smart devices/technologies. For this group, the affordability of these technologies and the accessibility to technical assistance were primary concerns. These findings highlight areas where further interventions are needed to improve farmers' attitudes towards smart devices/technologies. By enriching our understanding of these factors, such approaches can help tailor educational, marketing, and political initiatives and support systems to promote the adoption of PLF technologies.

A limitation of this machine learning approach is the potential to provide replicable results. It is important to note that the k-means clustering method has an element of randomness in its initialization process. This renders the outcome of the machine learning methods and the subsequent statistical analysis subject to a certain amount of variability. Therefore, small distances between clusters that show a high standard deviation should be treated with particular caution when interpreting the results, as a different run might change the cluster affiliation of some of the data points. It is also important to note that the importance of questions evaluated by the Decision Tree has to be viewed only in the context of the calculated clusters and not as a general judgment on their usefulness in survey designs. The relative importance score of questions is also prone to fluctuations based on the chosen hyperparameters of the Decision Tree, such as Tree Depth, Train/Test Split, Split Criterion, and many others. To ensure comparability, the standard settings of scikit learn have been used for the implemented Decision Tree.

Table 7
Histograms of k-means labels, with $k = 3$ clusters (blue = ready, yellow = partially ready, red = not ready) for each of the chosen questions.



The general limitation of this quantitative machine learning approach is its data-driven process. Compared to qualitative studies that often pursue theory-driven analysis, machine learning techniques focus on identifying underlying patterns without evaluating the context of the provided information. Therefore, a machine learning approach can identify correlated attributes that do not necessarily imply a causal relationship. Particular attention to the research design is necessary to provide the proper context and interpretation of the results (e.g., focus groups, expert analysis, interpretable models). Furthermore, the proposed quantitative approach cannot assess nuanced human behaviors during data collection, such as emotions, the time delay of answers, or non-verbal communication.

Finally, this research evaluated the survey questions' predictive power in determining farmers' readiness for technology adoption. It was shown that some questions have a considerably high predictive power to identify cluster affiliation and may be particularly suited to identify the technological readiness of user groups. This examination allows for the development of more targeted and efficient survey instruments, thereby improving data collection strategies for future research in this field. In this context, the results of this study have been integrated into an Online-Tool[3] to give targeted advice for users interested in adopting PLF technologies. This is done by asking the same questions used in this study and mapping the answers to the respective technological readiness categories with the deployed Decision Tree model. Based on the answers, individual advice will be provided to advance the respective user and farm characteristics to increase technological readiness.

[3] http://plf.farming.software/

As this study focused on exemplifying the potential benefits of supervised and unsupervised machine learning approaches for this domain, particularly when using explainable classifiers such as a Decision Tree, a comprehensive comparison of several algorithms is out of scope for this article. Therefore, further research is necessary to understand the full potential of machine learning for technology adoption studies and to utilize the diverse set of algorithms for individual tasks when analyzing survey data. Furthermore, for an in-depth statistical analysis of the presented survey data, we refer the interested reader to the publication of Tikász et al. (2023).

6. Conclusion

In this study, we identified and analyzed three distinctive clusters of farmers based on their attitudes towards technology adoption using an unsupervised and supervised machine learning approach. Our findings illustrated the capacity of k-means clustering to unveil overarching similarities in the data about user attitudes and technological readiness, highlighting inherent user characteristics of the respective groups. The most prominent separation occurred between two clusters, but a post hoc analysis revealed a third significant group with moderate attitudes towards technology adoption. This third group is of particular interest due to their balanced views on PLF, representing a potential user group for targeted interventions to facilitate technology uptake and increase farmer attitudes towards sustainable technologies. For this "partially ready" group, enhancing the accessibility of information and the availability of expert knowledge have been primary concerns. Conversely, the "not ready" group uniformly reported issues of affordability and accessibility to technical assistance.

By combining unsupervised and supervised machine learning methods, we investigated the ability of survey questions to predict cluster affiliation. This analysis indicated that some questions play a crucial role in separating cluster affiliation of user attitudes, thereby potentially influencing future research designs in this domain. The authors highlighted the need to increase the use of machine learning in this field to facilitate its ability to investigate user attitudes and unveil targeted information about intervention strategies.

Funding

This article reports on the result related to the LivestockSense project. This research was made possible by funding from the ICT-AGRI-FOOD 2019 Joint Call. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862665. The authors acknowledge the contribution of AgHiTech Kft (HU), Institute of Agricultural Economics (HU), Galilei Research Institute Ltd. (IL), SBA Research (AT), Innvite ApS. (DK), Swedish University of Agricultural Sciences (SE), Wroclaw University of Environmental and Life Sciences (PL), Estonia University of Life Sciences (EE), Aarhus University (DK), Denmark and the co-funding of the following organizations: NRDI Funds (HU), Hungary, Israel Innovation Authority (IL), Bundesministerium, LRT Fund (AT), GUDP (DK), Ministry of Rural Affairs (EE), The National Centre for Research and Development (PL), Poland and FORMAS (SE).

CRediT authorship contribution statement

Kevin Mallinger: Conceptualization, Methodology, Writing – original draft, Formal analysis, Validation, Supervision. **Luiza Corpaci:** Data curation, Software, Visualization. **Thomas Neubauer:** Project administration. **Ildikó E. Tikász:** Writing – review & editing. **Thomas Banhazi:** Funding acquisition, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data

Annex

See Tables 5–7.

References

Abdi, Hervé, Williams, Lynne J., 2010. Principal component analysis. Wiley Interdisc. Rev.: Comput. Stat. 2 (4), 433–459.

Abeni, Fabio, Petrera, Francesca, Galli, Andrea, 2019. A survey of Italian dairy farmers' propensity for precision livestock farming tools. Animals 9 (5), 202.

Alsayat, Ahmed, El-Sayed, Hoda, 2016. Social media analysis using optimized K-means clustering. In: 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications. SERA, IEEE, pp. 61–66.

Araújo, Sara Oleiro, Peres, Ricardo Silva, Barata, José, Lidon, Fernando, Ramalho, José Cochicho, 2021. Characterising the agriculture 4.0 landscape—emerging trends, challenges and opportunities. Agronomy 11 (4), 667.

Banhazi, T., 2009. Vision for the egg industry in the 21st century: Potential use of precision livestock farming technologies. In: Agricultural Technologies in a Changing Climate: The 2009 CIGR International Symposium of the Australian Society for Engineering in Agriculture. Engineers Australia Brisbane, Queensland, pp. 498–502.

Banhazi, T., Halas, V., Maroto-Molina, F., 2022. Introduction to practical precision livestock farming. In: Practical Precision Livestock Farming: Hands-on Experiences with PLF Technologies in Commercial and R&D Settings. Wageningen Academic Publishers, pp. 213–230.

Berckmans, Daniel, 2017. General introduction to precision livestock farming. Animal Front. 7 (1), 6–11.

Boothby, Amya L., White, Davida R., 2021. Understanding the barriers to uptake of precision livestock farming (PLF) in the UK sheep industry. Agric. Eng. Ag Eng. 2021, 572.

Caliński, Tadeusz, Harabasz, Jerzy, 1974. A dendrite method for cluster analysis. Commun. Stat.-theory Methods 3 (1), 1–27.

Davies, David L., Bouldin, Donald W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. (2), 224–227.

Dhraief, MZ, Bedhiaf-Romdhania, S, Dhehibib, B, Oueslati-Zlaouia, M, Jebali, O, Ben Youssef, S, 2018. Factors affecting the adoption of innovative technologies by livestock farmers in arid area of Tunisia. FARA Res. Rep. 3 (5), 22.

Drewry, Jessica L, Shutske, John M, Trechter, David, Luck, Brian D, Pitman, Lynn, 2019. Assessment of digital technology adoption and access barriers among crop, dairy and livestock producers in Wisconsin. Comput. Electron. Agric. 165, 104960.

Groher, T., Heitkämper, K., Umstätter, C., 2020. Digital technology adoption in livestock production with a special focus on ruminant farming. Animal 14 (11), 2404–2413.

Isabelle, Adjani, 2021. The Legitimacy of Precision Livestock Farming. SLU, Dept. of Urban and Rural Development.

Jain, Anil K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 31 (8), 651–666.

Kaushik, Nainika, Bhatia, Manjot Kaur, 2022. Twitter sentiment analysis using K-means and hierarchical clustering on COVID pandemic. In: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Vol. 1. Springer, pp. 757–769.

Maimon, Oded Z., Rokach, Lior, 2014. Data Mining with Decision Trees: Theory and Applications, Vol. 81. World scientific.

Makinde, Ayoola, 2020. Investigating Perceptions, Motivations, and Challenges in the Adoption of Precision Livestock Farming in the Beef Industry (Ph.D. thesis). University of Guelph.

Mallinger, K., Purcell, W., Neubauer, T., 2022. Systemic design requirements for sustainable digital twins in precision livestock farming. In: Proceedings of the 10th European Conference on Precision Livestock Farming.

Mingolla, Carla, Hudders, Liselot, Vanwesenbeeck, Ini, Claerebout, Edwin, 2019. Towards a biased mindset: An extended theory of planned behaviour framework to predict farmers' intention to adopt a sustainable mange control approach. Prevent. Veterin. Med. 169, 104695.

Molina, Mario, Garip, Filiz, 2019. Machine learning for sociology. Annu. Rev. Sociol. 45, 27–45.

Obringer, Renee, White, Dave D., 2023. Leveraging unsupervised learning to develop a typology of residential water users' attitudes towards conservation. Water Res. Manag. 37 (1), 37–53.

Pathak, Hari Sharan, Brown, Philip, Best, Talitha, 2019. A systematic literature review of the factors affecting the precision agriculture adoption process. Precis. Agric. 20, 1292–1316.

Pivoto, Dieisson, Waquil, Paulo Dabdab, Talamini, Edson, Finocchio, Caroline Pauletto Spanhol, Dalla Corte, Vitor Francisco, de Vargas Mores, Giana, 2018. Scientific development of smart farming technologies and their application in Brazil. Inf. Process. Agricult. 5 (1), 21–32.

Rousseeuw, Peter J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

Schukat, Sirkka, Heise, Heinke, 2021a. Smart products in livestock farming—An empirical study on the attitudes of German farmers. Animals 11 (4), 1055.

Schukat, Sirkka, Heise, Heinke, 2021b. Towards an understanding of the behavioral intentions and actual use of smart products among German farmers. Sustainability 13 (12), 6666.

Scown, Murray W., Winkler, Klara J., Nicholas, Kimberly A., 2019. Aligning research with policy and practice for sustainable agricultural land systems in Europe. Proc. Natl. Acad. Sci. 116 (11), 4911–4916.

Shang, Linmei, Heckelei, Thomas, Gerullis, Maria K, Börner, Jan, Rasch, Sebastian, 2021. Adoption and diffusion of digital farming technologies-integrating farm-level evidence and system interaction. Agric. Syst. 190, 103074.

Sok, Jaap, Borges, Joao Rossi, Schmidt, Peter, Ajzen, Icek, 2021. Farmer behaviour as reasoned action: A critical review of research with the theory of planned behaviour. J. Agric. Econ. 72 (2), 388–412.

Tabianan, Kayalvily, Velu, Shubashini, Ravi, Vinayakumar, 2022. K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. Sustainability 14 (12), 7243.

Tikász, I, Bálint, C, Király, G, Banhazi, T, Gron Sorensen, C, Gunnarsson, S, Kokin, E, Mallinger, K, Marchim, U, Opalinsky, S, 2023. A survey of pig and poultry farmers' readiness and attitudes towards smart technologies, accepted for publication in proceedings of the XL CIOSTA & CIGR section v international conference in évora. In: Proceedings of the XL CIOSTA & CIGR Section V International Conference in Évora.

Ugochukwu, Albert I., Phillips, Peter W.B., 2018. Technology adoption by agricultural producers: A review of the literature. In: From Agriscience to Agribusiness: Theories, Policies and Practices in Technology Transfer and Commercialization. Springer, pp. 361–377.