Agronomy Research **23**(X), xxx–ccc, 2025
https://doi.org/10.15159/AR.25.074

Using machine learning techniques to assess the technology adoption readiness levels of livestock producers

K. Mallinger[1,2,*], L. Corpaci[2], G. Goldenits[2], T. Neubauer[3], I.E. Tikász[4] and T. Banhazi[5,6]

[1]SBA Research, Complexity and Resilience Research Group, Floragasse 7/5, AT1040 Vienna, Austria
[2]University of Vienna, Kolingasse 14–16, AT1090 Vienna, Austria
[3]TU Wien, Data Science Unit, Floragasse 9-11, AT1040 Vieanna, Austria
[4]Institute of Agricultural Economics, Nonprofit Kft., Zsil u. 3–5, HU1093 Budapest, Hungary
[5]AgHiTech Kft, Kisbacom utca 1, Budapest, Hungary, 110, Budapest, Hungary
[6]Wroclaw University of Environmental and Life Sciences, Norwida Str. 25, PL50-375 Wroclaw, Poland
[*]Correspondence: kmallinger@sba-research.org

Received: March 23rd, 2025; Accepted: June 19th, 2025; Published: August 5th, 2025

Abstract. Technology adoption in agriculture, particularly in precision livestock farming (PLF), is often hindered by a range of barriers such as high investment costs, limited infrastructure, and uncertainty regarding the reliability and integration of new systems. Understanding these barriers is crucial for promoting the uptake of innovations that enhance sustainability and productivity. This study investigates technology adoption barriers in precision livestock farming to support sustainable agricultural development. A survey of 266 farms across several European countries and Israel was conducted to assess existing infrastructure and farmers' attitudes toward smart farming technologies. Using machine learning techniques, farmers were grouped into two clusters representing different levels of technological readiness. The study identified the most prominent factors influencing technology adoption, including the presence of smart technologies on-site, market accessibility, cost efficiency, and the ability to manage labor shortages. A Logistic Regression model further demonstrated high predictive accuracy for farmers' technological readiness based on these characteristics. These findings provide valuable insights into the main drivers and barriers of PLF adoption and highlight the relevance of data-driven approaches for requirement analysis and targeted policy interventions. By uncovering critical user traits and adoption barriers, this study offers structured guidance for policymakers, industry stakeholders, and researchers to foster the broader adoption of precision livestock technologies.

Key words: cluster analysis, machine learning, precision livestock farming, survey design, technological barriers.

INTRODUCTION

Precision livestock farming (PLF) refers to the use of advanced technologies - such as sensors, automated monitoring systems, and data analytics - to continuously observe and manage livestock production at the individual animal level. By providing real-time insights into animal health, welfare, and performance, PLF enables more efficient resource use, early detection of health issues, and targeted interventions. These capabilities support agricultural activities by enhancing productivity, improving animal welfare, and reducing environmental impacts (Wathes et al., 2008; Banhazi & Black, 2009; Banhazi et al., 2012; Banhazi et al., 2022). As global food demand grows, adopting PLF becomes vital for ensuring sustainable practices and meeting population needs (Araújo et al., 2021). However, successful implementation depends heavily on farmers' willingness to adopt these technologies and on the development of strategies to address existing barriers and concerns (Scown et al., 2019).

While PLF technologies offer significant advantages (Banhazi & Black 2009; Banhazi et al., 2022), their potential can be undermined if technology development and integration fail to align with the specific needs of farmers (Mallinger et al., 2022; Mallinger & Baeza-Yates, 2024). Understanding which farm characteristics and user traits distinguish technologically ready farmers from those less prepared is therefore essential for conducting targeted requirement analyses and designing sustainable, user-centered technologies. Hereby, technological readiness refers to the willingness and ability of farmers to adopt and effectively implement new technologies within their farming operations. However, existing research predominantly relies on traditional statistical (e.g., distributions) or qualitative (e.g., interviews) approaches to characterize PLF user groups and identify adoption drivers (Mallinger et al., 2023). These methods often oversimplify complex relationships between user traits and technology adoption behavior, limiting the ability to uncover distinct adoption patterns and hidden barriers. Furthermore, these approaches are not able to either systematically group farmers based on shared characteristics or investigate how the significance of adoption barriers varies among different types of users.

This study addresses this knowledge gap by applying machine learning techniques, specifically K-means clustering and a Logistic Regression classifier, to analyze survey data from 266 farms and investigate user attitudes toward technological readiness in PLF. The main objectives are to:

- Identify farm characteristics and user traits that differentiate farmers by their technological readiness for PLF adoption.
- Uncover adoption barriers specific to different user groups to support targeted interventions.
- Apply machine learning techniques to reveal latent patterns in survey data and validate user groupings.
- Assess the predictive value of survey variables to identify key barriers to technology adoption.

By uncovering user attitude clusters and understanding factors influencing technological readiness, this study provides structured information for policymakers, industry stakeholders, businesses, and researchers to support the adoption of precision livestock farming technologies.

STATE OF THE ART

The research field of user attitudes toward precision livestock farming is well-established. It is primarily relying on surveys and statistical analysis (Ugochukwu & Phillips, 2018; Abeni et al., 2019; Pathak et al., 2019; Groher et al., 2020; Makinde, 2020; Boothby & White, 2021; Schukat & Heise, 2021b; Akinyemi et al., 2025). Research in this field has identified a range of factors influencing farmers' technology adoption, including economic, socio-demographic, ethical, legal, technological, and institutional aspects, all of which play a role in ensuring broad acceptance (Dhraief et al., 2018; Drewry et al., 2019; Makinde, 2020).

Currently, machine learning has yet to be fully leveraged to detect distinct clusters of user attitudes in precision livestock farming. Some studies (Schukat & Heise, 2021a; Mallinger et al., 2023) have previously explored clusters of farmer characteristics. Schukat & Heise, 2021a adopted a hierarchical cluster analysis to understand the attitudes of German livestock farmers towards smart products. However, the study did not provide any validation of the cluster results and also did not analyze the complex relationship between the questions associated with a cluster. Similar approaches have been used in analyzing user characteristics and behaviors in water resource management (Obringer & White, 2023), social media studies (Kaushik & Bhatia, 2022), and customer segmentation (Tabianan et al., 2022).

This research builds on a prior publication that used clustering for the evaluation of technological adoption barriers (Mallinger et al., 2023). The experiment design for this study is kept the same for comparison. However, the prior study used different distance metrics for the cluster creation of the k-mean algorithm as well as different algorithms to assess the importance of farm characteristics for technological readiness (logistic regression instead of decision tree). By following and extending the conceptual design, the current study uses refined distance metrics to create improved cluster affiliations (up to 97% accuracy) while providing statistical evidence for the significance of the farm characteristics on technological readiness identified by the supervised machine learning algorithm. Thereby, this study provides novel results while contrasting existing literature.

MATERIALS AND METHODS

Fig. 1 illustrates the evaluation and validation of cluster characteristics reflecting user attitudes and farm infrastructure about PLF adoption readiness. This encompasses the entire data processing pipeline, from data aggregation and cleaning to validating applied clustering methods. As depicted in Fig. 1, the experimental design steps will be presented sequentially.

This study is part of the LivestockSense project[1], aiming to develop a functional prototype to assess and predict farmers' technological readiness. The machine learning approach focuses on defining cluster boundaries representing this readiness and creating a model capable of accurately classifying new user questionnaires, as was primarily done in the form of a web-based prototype to identify the technological readiness of precision livestock farmers (Mallinger et a., 2023).

[1] https://livestocksense.eu/}.

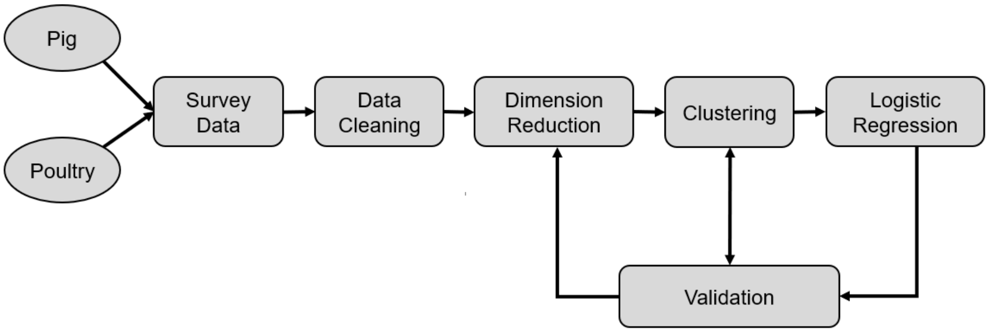

Figure 1. Experiment design and processing pipeline chosen for this study (see also Mallinger et al., 2023).

Survey and Data

This study builds on survey data collected by the LivestockSense research team. Based on self-reported questionnaires, the survey gathered responses from 266 farms across five European countries and one Middle Eastern country, covering the pig (121 samples) and poultry (145 samples) sectors. The questionnaire was designed to capture insights into existing infrastructure and farmers' attitudes toward smart farming technologies.

To assess technological readiness, the survey incorporated multiple perspectives. It examined infrastructure availability (question blocks 1, 2, and 6), the presence of expert knowledge and market access to PLF technologies (question block 4), and farmers' mental attitudes toward PLF adoption (question blocks 3 and 5). In total, 20 key questions were selected to represent farmers' readiness for precision livestock farming. Sub-questions were consolidated into single features, with responses rated on 5-point, 4-point, or 3-point scales to indicate levels of agreement, where 1 represented ‘Strongly disagree’ and 5 indicated ‘Strongly agree,’ for example. The complete list of survey questions used in the analysis is in the Appendix.

Clustering

This study uses the k-means approach to cluster the results. 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, the sum of the squared distance between each data point and its assigned cluster centre. In other words, the algorithm aims to find the number of (k) centres that minimize the distance between each point and its assigned centre, resulting in clusters that are as compact and distinct as possible. Mathematically, the objective function of k-means is:

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

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.

This study used the Gower's distance to determine the pairwise similarities between observations and create equidistant categories when applying the k-means algorithm. Gower's distance is defined as:

$$\mathrm{sijk} = |\mathrm{xi} - \mathrm{xj}|/\mathrm{Rk} \quad (2)$$

where xi and xj are two observations, and Rk is the range of the k-th variable. This means that for two answer patterns, the absolute difference will be computed for each question and then divided by the range of possible answer categories. This is particularly useful in our case, as it normalizes the differences between each pair of observations and, therefore, between different scales.

Validation

The validation process consisted of a multi-step mixed-method approach including quantitative (UMAP, Internal Validation Metrics, Supervised Machine Learning) 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. The validation process was iteratively conducted to ensure that the survey questions and cluster results convey user attitudes and technological readiness characteristics.

Firstly, Cluster quality was assessed using three internal validation metrics:

– **Davies-Bouldin Index:** Evaluates cluster similarity by comparing within-cluster and between-cluster distances. Lower values indicate better clustering (Davies & Bouldin, 1979).

– **Calinski-Harabasz Index:** Measures the ratio of the sum of between-cluster to the sum of within-cluster dispersion, with higher values indicating well-defined clusters (Caliński & Harabasz, 1974).

– **Silhouette Score:** This score assesses how well data points fit within their clusters by comparing intra-cluster and nearest-cluster distances. Scores range from -1 (misclassified) to 1 (well-clustered), with values near 0 indicating overlap (Rousseeuw, 1987).

The validation process also includes the UMAP embedding (McInnes et al., 2018) of the data into a two-dimensional space to assess the clustering results visually. UMAP is a non-linear dimensionality reduction technique based on the widely used t-SNE method that embeds the datapoint distribution from a high dimensional space into lower dimensions. We chose UMAP, particularly for its ability to preserve structures, like clusters, in high-dimension and lower-dimension representations. Another advantage of UMAP is that it works with custom distance metrics such as Gower's distance and does not assume continuous variables or normal data distribution. In dimensionality reduction, each explanatory variable represents a dimension, and each observation is a point in this multi-dimensional space. For example, if height and weight are collected for individuals, they can be plotted on a two-dimensional scatterplot with one variable on each axis. However, with more variables - such as age, education, or family size - he data exist in a higher-dimensional space that cannot be easily visualized. Dimensionality reduction techniques like UMAP project this complex, high-dimensional structure into a lower-dimensional space (typically two dimensions) while preserving essential relationships, such as clustering patterns.

Finally, a Focus Group was assigned to assess outcomes throughout the experiment pipeline continuously. This group included five specialists from four different countries, covering areas such as livestock farming, PLF technology development, and survey design. The evaluation process was carried out iteratively after each significant processing stage, guaranteeing the logical consistency of the selected features and clustering outcomes.

Logistic Regression

This study utilizes a Logistic Regression Model to (1) assess how well the survey questions predict farmers' technological readiness, indicated by the cluster labels derived from the k-means model, (2) develop a model for predicting farmer attitudes, and (3) create an explainable framework that allows for assessing feature importance according to the odds ratio, providing insight into predictions based on the contributions from each feature. Logistic Regression models the probability that a given input belongs to a particular class by applying the logistic function to a weighted sum of the features (survey responses). This probability is expressed as:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}} \tag{3}$$

where $\beta_0$is the intercept, β_i are the feature weights, and x_i are the input features. The model is trained by finding β values that maximize the likelihood of the observed labels. In most software solutions, the optimization process is done by gradient descent or iterative methods like the Newton-Raphson Method.

Regarding output interpretation, each β_i represents the effect of a one-unit change in the corresponding feature on the log odds of the outcome occurring. Because interpreting log odds directly can be unintuitive, using the odds ratio, given by e^{β_i}, is common. The odds ratio indicates how much the odds of the outcome change for a one-unit increase in the feature. In the case of predicting only two groups, odds ratios greater than 1 suggest a positive association with group 1, while values less than 1 suggest an association with group 0.

Using a Logistic Regression Model to predict clusters with varying technological readiness, we have two main goals:

1. To see if a supervised modelling method can summarise the correlation between clusters (target variable) and input variables (survey questions).
2. To identify which survey questions provide the most significant insights for distinguishing cluster associations.

For calculating all evaluation metrics, coefficients, p-values, and odds ratios, we took the mean values of a 10-fold stratified cross-validation with a random state set to 42. We assessed feature significance by calculating the p-value for each feature and ranking the significant features according to their significance in descending order.

Predictions were evaluated using the following metrics:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{P}} \tag{6}$$

$$\text{F1 Score} = 2 * \frac{\text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \tag{7}$$

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

RESULTS AND VALIDATION

As a first step, we deploy the K-means clustering described in Section 3.2. We use three internal validation metrics: the Davies Bouldin Score, the Calinski Harabasz Score, and the Silhouette Score, to estimate a suitable number of clusters.

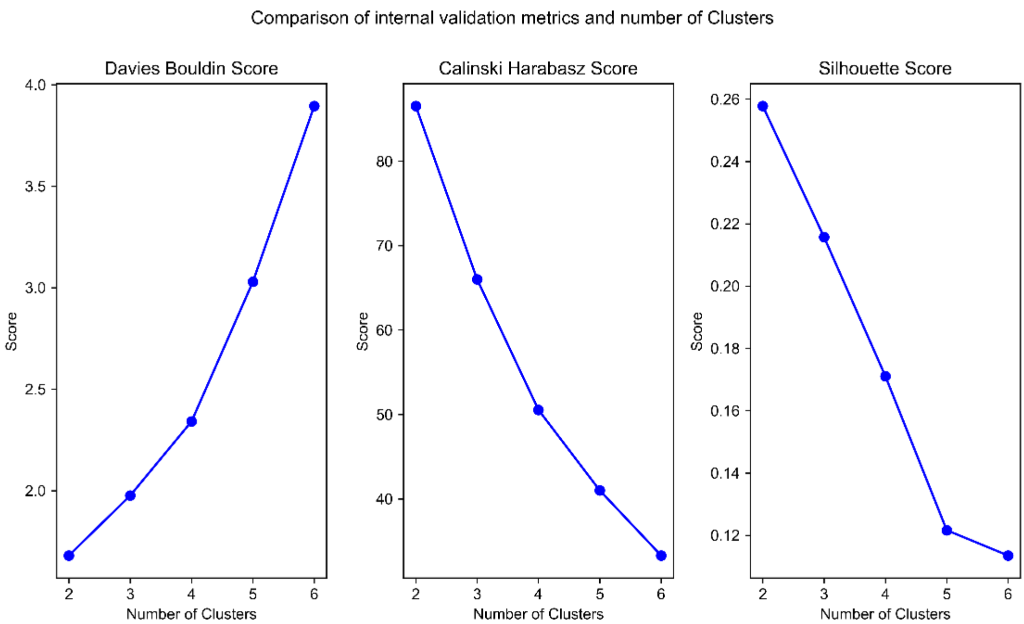

Figure 2. Davies-Bouldin Score on the left (lower is better), Calinski Harabasz Score in the middle (higher is better), and the Silhouette Score on the right (higher is better).

The score development for each number of clusters and each metric is shown in Fig. 2, which depicts the Davies-Bouldin Score in the left diagram, the Calinski Harabasz Score in the centre, and the Silhouette Score to the right. All internal validation metrics indicate that the best separation can be achieved for choosing two clusters, as the Davies Bouldin Score is the lowest at 1.68, and the other two scores are the highest at 86.51 and 0.258, respectively, for all compared clusters.

Of the 266 survey results, 118 belong to Cluster 0 (later defined as technologically not ready), and 148 to Cluster 1 (later defined as technologically ready). To evaluate if the two clusters are distinct from each other and form logical opposites, we used a UMAP embedding. This embedding represents the 20-dimensional farm characteristics in two-dimensional space, enabling a visual inspection of the data distribution. Fig. 3 shows this representation, with the clusters each lying in different regions of the plot, highlighting the differences between the two clusters. Some overlap exists between the clusters, which is to be expected, as some farmers' answers indicate behaviour belonging to the other cluster compared to the one they were assigned to. However, these clusters do not bear any semantic meanings yet, as the separation is currently based on mathematical evaluations. To overcome this, we analyse the distribution of given answers for each cluster.

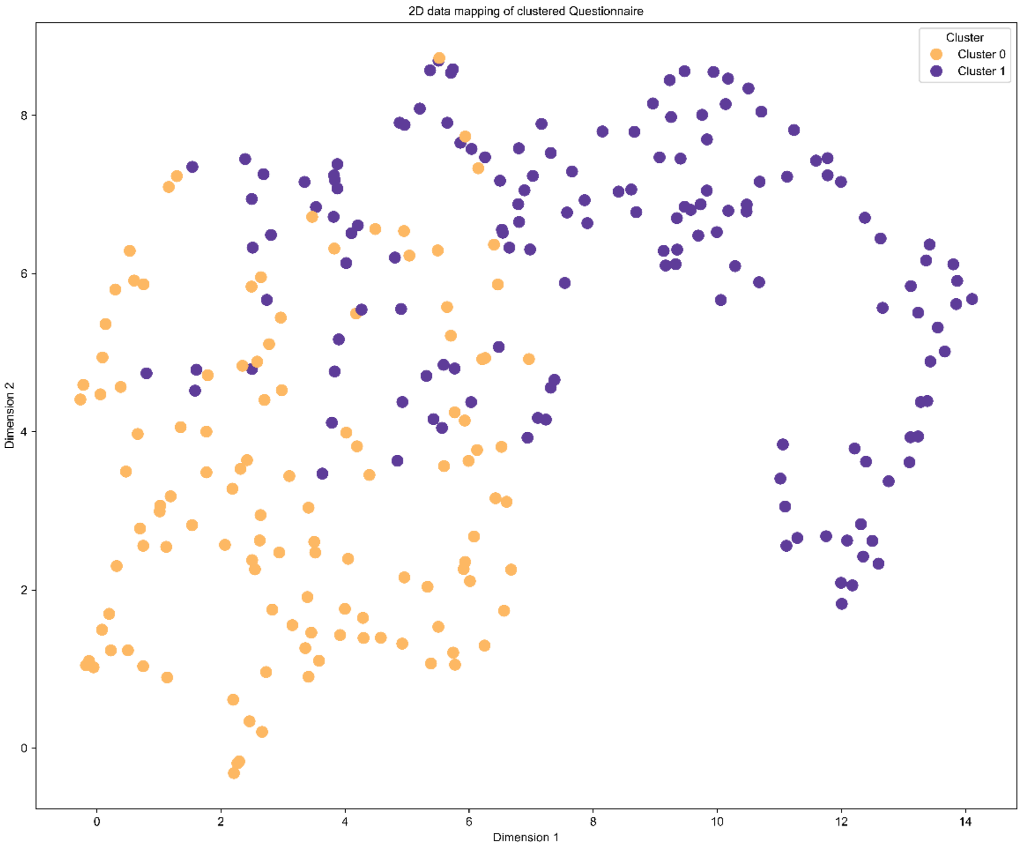

Figure 3. UMAP representation of the two clusters for the visual evaluation of cluster distriubtion. On the left (orange) are the data points later considered as technologically not ready, where as on the right (purple) the datapoints defined as technologically ready. The x- and y-axes correspond to the two reduced dimensions obtained through UMAP, capturing the main structure of the high-dimensional data.

Farmer characteristics per cluster

We first analyse the distribution of cluster and survey question characteristics using grouped bar charts, such as those depicted in Fig. 4. We chose these figures in particular because we have six questions, some of which have subquestions. For Questions 3, 4, and 5, each containing multiple subquestions, the response patterns within each cluster are somewhat comparable across the subquestions. Consequently, we selected one subquestion from each as a representative. The plots for the remaining questions can be found in the appendix section.

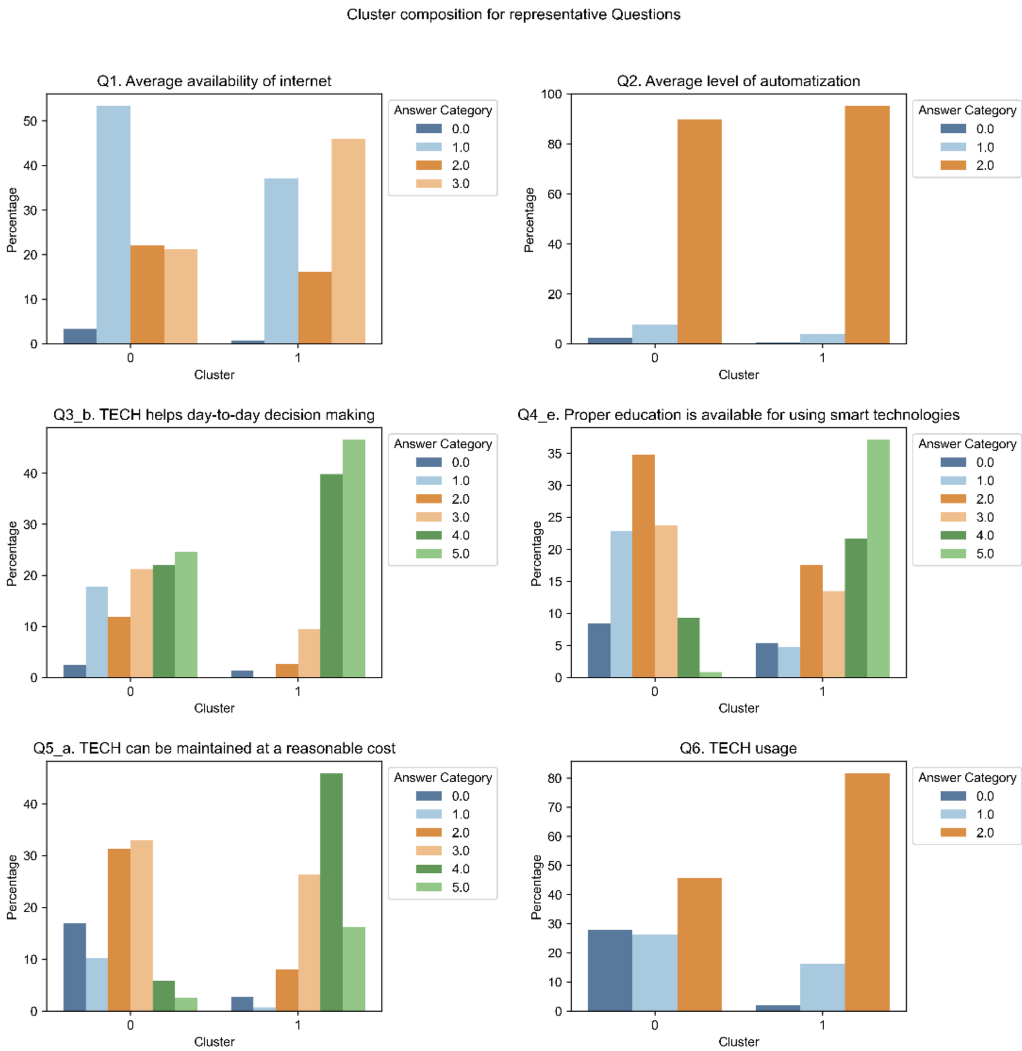

Figure 4. Distribution of answer results as a comparison for Cluster 0 (Technologically not ready) vs. Cluster 1 (Technologically Ready). The number of bars represents the amount of possible answer options. The height of the bars represents the share of an answer category compared to the other categories. In every question, the sum for each cluster, therefore, results in 100.

In each diagram, the answers to one of the survey questions are shown so that the clusters are drawn on the x-axis, and for each cluster, there are as many bars as possible answer categories. As the clusters vary in size, we cannot just represent the absolute values per answer category per cluster to ensure comparability, as it would capture too much cluster size information. Instead, the bar heights represent the fraction of each answer category per cluster.

Fig. 4 compares responses to six questions related to technological readiness across two clusters (Cluster 0 and Cluster 1). Notably, five of these questions (Q1, Q3, Q4, Q5, and Q6) show pronounced differences in how each cluster responded, whereas Q2 (average level of automation) appears more similar between the two groups.

For Q1 (average availability of internet), Cluster 1 shows a substantially higher proportion of respondents in the upper rating categories (e.g., 4 and 5), while Cluster 0 leans toward lower or moderate ratings. A similar pattern emerges in Q3 (technology helps day-to-day decision-making), where Cluster 1 overwhelmingly endorses strong agreement, contrasting with the more moderate or disagreeing responses in Cluster 0.

Likewise, Q4 (proper education is available for using smart technologies) and Q5 (technology can be maintained at a reasonable cost) reveal that Cluster 1 is more likely to perceive educational support and affordability favourably, whereas Cluster 0 exhibits relatively lower levels of agreement. Finally, Q6 (overall technology usage) follows this trend: Cluster 1 reports higher usage, whereas Cluster 0 indicates less frequent or more constrained use. Similar patterns can be found in the other questions presented in the Annex.

Based on the intended focus of the questions that highlight certain barriers to technology adoption, these findings suggest that Cluster 1 is characterized by a high degree of technological readiness - reflected in stronger internet availability, more positive attitudes toward technology's utility, better access to education, perceived affordability, and higher overall usage. In contrast, Cluster 0 consistently reports lower levels of agreement or usage across these dimensions, indicating limited technological readiness. As a result, we define users in Cluster 0 as ‘not technologically ready,’ while Cluster 1 represents farm characteristics who are generally perceived as ‘technologically ready’ to deploy PLF technologies at their farms.

Feature Importance Analysis

As an additional means of result validation, we used a Logistic Regression model to test whether the survey questions are a reliable predictor for each respondent’s cluster membership. Table 1 presents the average accuracy, precision, recall, and F1 score value for a stratified five-fold cross-validation and demonstrates that predictions using the questions allow for high scores in each computed metric. With a total of 97% accuracy and F1-Score, the model provides almost no false negatives or false positives and can, therefore, be used as a robust model to distinguish technological-ready from less technology-ready farmers.

Table 1. This table depicts the result of the logistic regression model, which was used to predict the technological ready and not ready cluster based on the survey questions

Classifier	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.97	0.96	0.98	0.97

In addition to the results presented in Section 4.1, we chose Logistic Regression as an interpretable means to analyse the farm characteristics that separate the technologically ready from less-ready farmers. If the model accurately classifies respondents, it confirms that the survey items capture meaningful differences in technology adoption. Furthermore, we can pinpoint the key factors that distinguish the clusters by computing the p-value and the log odds for each question's estimator. The odds ratio thereby shows the difference in the odds of the outcome occurring for a one-unit increase in the predictor variable, holding all other variables constant. We deemed features corresponding to estimators that are significant on a 0.05 significance level and have a high odds ratio as the most influential questions for predicting cluster association.

Table 2 shows the features, corresponding estimator, p-value, and odds ratio. The non-significant entries in the table are ranked according to their p-value, and the remaining five variables (and the intercept term) are ranked according to their odds ratio to uncover their importance in separating respondents into 'technologically ready' and 'less ready' clusters. Q6, 'TECH usage', has the highest odds ratio, followed by Q4_a, 'It is easy to access TECH on the market', Q5_a, 'TECH can be maintained at a reasonable cost', Q3_a, 'TECH helps labour shortage', and Q5_e, 'TECH is secure in terms of data management'. These questions highlight economic considerations and affinity towards technical solutions as the most influential factors for predicting technological readiness. Interestingly, data security is a strong predictor for technological readiness, which indicates that technologically ready farmers value privacy more than non-ready groups.

Table 2. The table shows the features used in the Logistic Regression model. The features are sorted based on the coefficient. Higher coefficients represent a higher degree of importance or influence on the modelling output. Statistical significance for the importance is tested and represented in the column p-value

Feature	Description	Co-efficient	*p*-value	Odds ratio
Q6	TECH usage	1.554	0.03	4.731
Q4_a	It is easy to access TECH on the market	1.258	0.024	3.520
Q5_a	TECH can be maintained at a reasonable cost	1.057	0.025	2.876
Q3_a	TECH helps labour shortage	0.952	0.011	2.592
Q5_e	TECH is secure in terms of data management	0.854	0.031	2.348
-	Intercept	-44.888	< 0.01	0
Q3_d	TECH helps meet environmental pollution reduction obligations	0.790	0.060	2.204
Q4_c	It is easy to get information on TECH and distributors	0.991	0.075	2.693
Q3_b	TECH helps day-to-day decision making	0.865	0.093	2.375
Q4_b	TECH can be purchased at an affordable price	0.805	0.105	2.236
Q5_d	TECH operates in a reliable manner	0.918	0.134	2.505
Q4_d	It is easy to get technical assistance to smart technologies	0.917	0.161	2.501
Q5_b	TECH is easy to operate	0.789	0.188	2.201
Q1	Average availability of internet	0.647	0.190	1.911
Q3_c	TECH helps enterprise, marketing and investment decisions	0.471	0.237	1.602
Q3_e	TECH enables the increase production effectiveness	0.552	0.461	1.736
Q4_e	Proper education is available for using smart technologies	0.348	0.468	1.416
Q3_g	TECH provides reliable information	0.538	0.526	1.712
Q5_c	TECH can be connected well with other equipment/software	0.236	0.610	1.266
Q3_h	TECH provides information in a real-time manner	0.469	0.637	1.598
Q2	Average level of automatization	0.104	0.911	1.110

An intercept term was added to the regression model to ensure unbiased coefficient estimators and accounts for the base case, where all other variables are 0. The intercept is also statistically significant at a p-value close to 0, but the odds ratio is almost 0, thus removing its importance in predicting cluster associations. This means it is very unlikely that a respondent answered with ‘0’ on each question.

While most features do not have statistical significance, this does not mean these questions are irrelevant. Instead, their influence on the prediction is not high enough to rule out a random connection to the prediction outcome.

DISCUSSION

This research explores farmers' attitudes and technological readiness toward PLF technologies. Rather than directly assessing technology adoption, it examines general perceptions of farmers, calculates clusters of similar characteristics and identifies critical farm characteristic and personal perceptions on technology that distinguish technologically ready from non-ready farmers. Methodologically, it highlights the value of machine learning techniques for survey and requirement analysis.

The two primary clusters labelled ‘Ready’ (technologically ready) and ‘Not Ready,’ (barriers to technology adoption) exhibit distinct and consistent attitudes toward PLF technologies. The ‘Ready’ group (148 farms) shows strong agreement with the benefits of these technologies, indicating a high level of acceptance and readiness for adoption. In contrast, the ‘Not Ready’ group (118 farms) demonstrates lower agreement, suggesting a lower preparedness for technological adoption. The distinction between these clusters was further confirmed using UMAP, which revealed a clear separation in the reduced-dimensional space. This separation was also validated by a supervised machine learning model, which could successfully predict the cluster affiliation based on survey responses.

A prior study about barriers to technological readiness (Mallinger et al., 2023) highlighted the ‘ease of access to smart technologies’, ‘the interoperability’ ‘the ability to cope with a labour shortage’, and the ‘ease of operability’ as four distinct characteristics that distinguish technologically ready from non-ready farmers. The current study partially supports these results, as the ‘access to smart technologies’ as well as the ‘ability to cope with labour shortage’ are also highlighted as significant variables. However, this study also highlighted the prior existence of smart technologies as a primary factor used to distinguish the two clusters. This is not particularly surprising and supports the findings of (Mallinger et al., 2024), in which the distribution of said characteristic was used to validate the usefulness of the cluster results. As the tree-based mechanisms in the study of Mallinger et al., 2023 use the concept of diminishing entropy-values as a factor for identifying important features, it is possible that some farm characteristics were not highlighted adequately if a high degree of predictions are already successful with a small subset of survey questions. As a logistic regression approach automatically uses all questions, the differences between the characteristics are less pronounced compared to a decision-tree algorithm. This highlights that logistic regression might provide a more suitable approach to identifying important characteristics for cluster differentiation. However, more research is necessary to better understand the usability of different machine learning classifiers and regressors in this context.

Prior studies have reported high investment costs as a major barrier (Ugochukwu & Phillips, 2018; Abeni et al., 2019; Pathak et al., 2019; Groher et al., 2020; Makinde, 2020; Boothby & White, 2021). Our study shows a clear separation of farmers’ answers between the two clusters regarding the affordability of precision livestock technology.

The trained algorithm for predicting technological readiness exhibited stronger links to maintenance costs (coefficient 1.057) than to the initial purchase price (coefficient 0.805), with the prior links also supported by stronger p-values. Robustness and reliability are also frequently cited barriers (Drewry et al., 2019; Makinde, 2020; Boothby & White, 2021). We also observe a visible divergence of opinions between the clusters on this point, as shown in Fig. 7, and record a high coefficient of 0.918, although the feature-importance score (coefficient) was not statistically significant. Persistent apprehension about data security was further analysed in this study, again supporting earlier findings (Drewry et al., 2019). Finally, accessibility to the market emerged as one of the most prominent factors separating the clusters, with a high feature-importance coefficient of 1.258 and statistical significance (p-value 0.024). As this factor is rarely highlighted in previous work, it should be considered more closely in future research on technological barriers.

More generally, the cluster analysis showed that farmers were nearly evenly divided between those classified as technologically ready and those not ready. This suggests substantial potential for intervention strategies targeting the less-ready group. Notably, the average level of farm automatization did not differentiate between the two groups, suggesting that general mechanization alone is not a sufficient indicator of readiness. In contrast, prior use of smart devices emerged as one of the strongest predictors of technological readiness. Furthermore, the analysis highlights that barriers to technology adoption in livestock farming are not singular but multivariate in nature. Multiple factors must be considered simultaneously to accurately capture the diversity of farmer needs and the complexity of the adoption process.

A limitation of this study is the potential variability of the machine learning techniques. K-means clustering involves randomness in initialization, which can lead to slight differences in cluster assignments across runs. Consequently, small distances between clusters with high standard deviations should be interpreted cautiously, as some data points may shift clusters in different iterations. Similarly, the feature importance analysis from the Logistic Regression should be interpreted carefully. Changes in the distribution of answers could potentially change the order of highlighted questions. To give a more complete picture, these results should be combined with other tools to evaluate importance, such as the principal component analysis or explainable AI techniques like Partial Dependence Plots or SHAP values. A framework to apply this can be found in (Mallinger et al., 2024).

A key limitation of this quantitative machine learning approach is its data-driven nature. Unlike theory-driven qualitative studies, machine learning identifies patterns without considering the broader context. As a result, it may highlight correlations that do not imply causation. Careful research design, including survey questions, expert analysis, focus groups, or interpretable models, is essential for meaningful interpretation. Additionally, this approach cannot capture nuanced human behaviours during data collection, such as emotions, response delays, or non-verbal communication.

Finally, this research evaluated the predictive power of survey questions in determining farmers' technology readiness. Some questions demonstrated strong predictive value for cluster affiliation, making them particularly useful for identifying the readiness levels of user groups. These insights support the development of more targeted and efficient survey instruments, enhancing data collection strategies for future research. Furthermore, this process can be utilized as a data-driven approach to inform requirements and market analysis, facilitating the development of precision livestock farming technology by aligning it with targeted consumer needs. For example, companies can focus on special offers that keep maintenance and service costs at a minimum. By doing so, it enhances ease of operation and promotes the effective use of economic and environmental opportunities. As a result, the likelihood that these technologies improve production efficiency or animal well-being is significantly increased. Companies can also provide targeted market strategies based on these results. By recognizing that farmers highly value the support of precision livestock equipment in combating labour shortages, they can effectively highlight this aspect in their marketing activities.

This study aimed to demonstrate the benefits of supervised and unsupervised machine learning in this domain, particularly using interpretable models like Logistic Regression. More research is needed to explore the full potential of machine learning for technology adoption studies and to apply different algorithms to specific survey analysis tasks. A comparison to other machine learning approaches can be found here (Mallinger et al., 2023; Mallinger et al., 2024). For an in-depth statistical analysis of the survey data, we refer the interested reader to (Tikász et al., 2023a; Tikász et al., 2023b). Primary factors that distinguished these two groups inclubbded prior adoption of smart technologies, market accessibility, cost efficiency, and the ability to cope with labour shortages. The study showed that a Logistic Regression model could successfully predict 97% of farmers' technological readiness while highlighting statistically significant farm characteristics. Depending on the goal of one's work, further research could include a more detailed examination of farm characteristics (e.g., farm age, geography, farm type) and assess farmers' willingness to adopt technologies.

CONCLUSIONS

In this study, the authors identified and analysed two distinct clusters of farmers based on their attitudes toward technology adoption using unsupervised and supervised machine learning approaches. Our findings demonstrated the effectiveness of K-means clustering in revealing overarching similarities in user attitudes and technological readiness, highlighting inherent characteristics within each group and the prominent factors distinguishing them. The cluster analysis showed a balanced split between farmers classified as technologically ready and those not ready, thereby indicating substantial potential for targeted intervention strategies. The less-ready group expressed particular concern about the ability of PLF technology to address labour shortages and support day-to-day decision-making, as well as a lack of available education. Comparatively, the factors that most strongly distinguished the two groups include market accessibility, available education, cost efficiency, and the ability to cope with labour shortages. Notably, the average level of automatization did not differ between groups, suggesting that general mechanization is not a sufficient indicator of readiness. However, prior use of smart devices showed clear differences and was among the strongest predictors in the applied Logistic Regression model, along with market accessibility, maintenance costs, and data management security. The model successfully predicted 97% of farmers' technological readiness while highlighting statistically significant farm characteristics. The analysis further confirms that barriers to technology adoption are multivariate in nature and must be considered in combination to capture the diversity of farmer needs and the complexity of the adoption process.

Therefore, the integration of unsupervised and supervised machine learning methods facilitated the identification of survey questions that play a crucial role in differentiating cluster affiliations based on user attitudes. The authors highlighted the need to investigate the potential of machine learning in this field and its ability to analyse user attitudes. The results can be used to unveil targeted information for survey design, requirement analysis, and policy intervention strategies.

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

FUNDING: SBA Research (SBA-K1 NGC) is a COMET Centre within the COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMAW, and the federal state of Vienna. COMET is managed by FFG. Additionally, the financial support by the Austrian Federal Ministry of Labour and Economy, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged. Furthermore, the authors acknowledge the funding of the project 'Digitale Landwirtschaft', by the Austrian Federal Ministry of Education, Science and Research within the funding scheme 'Digitale und soziale Transformation in der Hochschulbildung'. 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 confounding 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).

Data Availability Statement: The data used is referenced in the article respectively in Section Survey and Data.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

Abeni, F., Petrera, F. & Galli, A. 2019. A survey of Italian dairy farmers' propensity for precision livestock farming tools. *Animals* **9**, 202. https://doi.org/10.3390/ani9050202

Akinyemi, B.E., Siegford, J.M., Jessiman, L., Turner, S.P., Johnson, A.K. & Akaichi, F. 2025. Precision livestock farming usage among a subset of U.S. swine producers: Insights through a structural equation modeling approach. *Smart Agricultural Technology* **10**, 100839. https://doi.org/10.1016/j.atech.2025.100839

Araújo, S.O., Peres, R.S., Barata, J., Lidon, F. & Ramalho, J.C. 2021. Characterizing the Agriculture 4.0 landscape—emerging trends, challenges and opportunities. *Agronomy* **11**, 667. https://doi.org/10.3390/agronomy11040667

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* (pp. 17–25). Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-934-3_0

Banhazi, T.M. & Black, J.L. 2009. Precision livestock farming: A suite of electronic systems to ensure the application of best-practice management on livestock farms. *Australian Journal of Multi-disciplinary Engineering* **7**(1), 1–14. doi: 10.1080/14488388.2009.11464794

Banhazi, T.M., Lehr, H., Black, J.L., Crabtree, H., Schofield, P., Tscharke, M. & Berckmans, D. 2012. Precision livestock farming: An international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering* **5**(3), 1–9. https://doi.org/10.3965/j.ijabe.20120503.001

Boothby, A.L. & White, D.R. 2021. Understanding the barriers to uptake of precision livestock farming (PLF) in the UK sheep industry. In *Proceedings of the European Conference on Agricultural Engineering (AgEng 2021)* (p. 572).

Caliński, T. & Harabasz, J. 1974. A dendrite method for cluster analysis. *Communications in Statistics – Theory and Methods* **3**(1), 1–27. https://doi.org/10.1080/03610927408827101

Davies, D.L. & Bouldin, D.W. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**(2), 224–227. doi: 10.1109/TPAMI.1979.4766909

Dhraief, M.Z., Bedhiaf-Romdhani, S., Dhehibi, B., Oueslati-Zlaoui, M., Jebali, O. & Ben Youssef, S. 2018. Factors affecting the adoption of innovative technologies by livestock farmers in arid area of Tunisia. *FARA Research Report*, **3**, 22.

Drewry, J.L., Shutske, J.M., Trechter, D., Luck, B.D. & Pitman, L. 2019. Assessment of digital technology adoption and access barriers among crop, dairy and livestock producers in Wisconsin. *Computers and Electronics in Agriculture* **165**, 104960. doi: 10.1016/j.compag.2019.104960

Groher, T., Heitkämpfer, K. & Umstätter, C. 2020. Digital technology adoption in livestock production with a special focus on ruminant farming. *Animal* **14**, 2404–2413. https://doi.org/10.1017/S1751731120001391

Jain, A.K. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* **31**, 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kaushik, N. & Bhatia, M.K. 2022. Twitter sentiment analysis using K-means and hierarchical clustering on COVID pandemic. In *Proceedings of the International Conference on Innovative Computing and Communications* **1**, pp. 757–769. Springer. doi: 10.1007/978-981-16-2594-7_61

Makinde, A. 2020. Investigating perceptions, motivations, and challenges in the adoption of precision livestock farming in the beef industry (Master's thesis). University of Guelph, Canada. https://atrium.lib.uoguelph.ca/items/602efe35-2490-4042-8516-ddd74d34ceb6

Mallinger, K., Corpaci, L., Neubauer, T., Tikász, I.E. & Banhazi, T. 2023. Unsupervised and supervised machine-learning approach to assess user readiness levels for precision livestock farming technology adoption in the pig and poultry industries. *Computers and Electronics in Agriculture* **213**, 108239. https://doi.org/10.1016/j.compag.2023.108239

Mallinger, K. & Baeza-Yates, R. 2024. Responsible AI in farming: A multi-criteria framework for sustainable technology design. *Applied Sciences* **14**(1), 437. doi: 10.3390/app14010437

Mallinger, K., Corpaci, L., Neubauer, T., Tikász, I.E., Goldenits, G. & Banhazi, T. 2024. Breaking the barriers of technology adoption: Explainable AI for requirement analysis and technology design in smart farming. *Smart Agricultural Technology* **9**, 100658. https://doi.org/10.1016/j.atech.2024.100658

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*. https://doi.org/10.34726/4644

McInnes, L., Healy, J. & Melville, J. 2018. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv preprint*, arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426

Obringer, R. & White, D.D. 2023. Leveraging unsupervised learning to develop a typology of residential water users' attitudes towards conservation. *Water Resources Management* **37**(1), 37–53. https://doi.org/10.1007/s11269-022-03354-3

Pathak, H.S., Brown, P. & Best, T. 2019. A systematic literature review of the factors affecting the precision agriculture adoption process. *Precision Agriculture* **20**, 1292–1316. https://doi.org/10.1007/s11119-019-09653-x

Rousseeuw, P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schukat, S. & Heise, H. 2021a. Smart products in livestock farming – An empirical study on the attitudes of German farmers. *Animals* **11**(4), 1055. https://doi.org/10.3390/ani11041055

Schukat, S. & Heise, H. 2021b. Towards an understanding of the behavioral intentions and actual use of smart products among German farmers. *Sustainability* **13**(12), 6666. https://doi.org/10.3390/su13126666

Scown, M.W., Winkler, K.J. & Nicholas, K.A. 2019. Aligning research with policy and practice for sustainable agricultural land systems in Europe. *Proceedings of the National Academy of Sciences of the United States of America* **116**(11), 4911–4916. doi: 10.1073/pnas.1812100116

Tabianan, K., Velu, S. & Ravi, V. 2022. K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability* **14**(12), 7243. https://doi.org/10.3390/su14127243

Tikász, I.E., Bálint, C., Király, G., Sorensen, C.A.G., Opalinsky, S., Gunnarsson, S., Marchim, U., Kokin, E., Mallinger, K. & Banhazi, T. 2023a. A survey of pig and poultry farmers' readiness and attitudes towards smart technologies. In *Proceedings of the XL CIOSTA & CIGR Section V International Conference*, Évora, Portugal (pp. 92–100).

Tikász, I.E., Király, G., Bálint, C., Opalinsky, S., Sorensen, C.A.G., Gunnarsson, S., Marchim, U., Kokin, E., Mallinger, K. & Banhazi, T. 2023b. Conditions of applying advanced information technologies in livestock farming. In *Proceedings of the XL CIOSTA & CIGR Section V International Conference*, Évora, Portugal (pp. 101–110).

Ugochukwu, A.I. & Phillips, P.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* (pp. 361–377). Springer. https://doi.org/10.1007/978-3-319-67958-7_17

Wathes, C.M., Kristensen, H.H., Aerts, J.-M. & Berckmans, D. 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Computers and Electronics in Agriculture* **64**(1), 2–10. https://doi.org/10.1016/j.compag.2008.05.005

APPENDIX

APPENDIX A1 – Remaining Distribution of Survey Answers that haven't been displayed in the result section

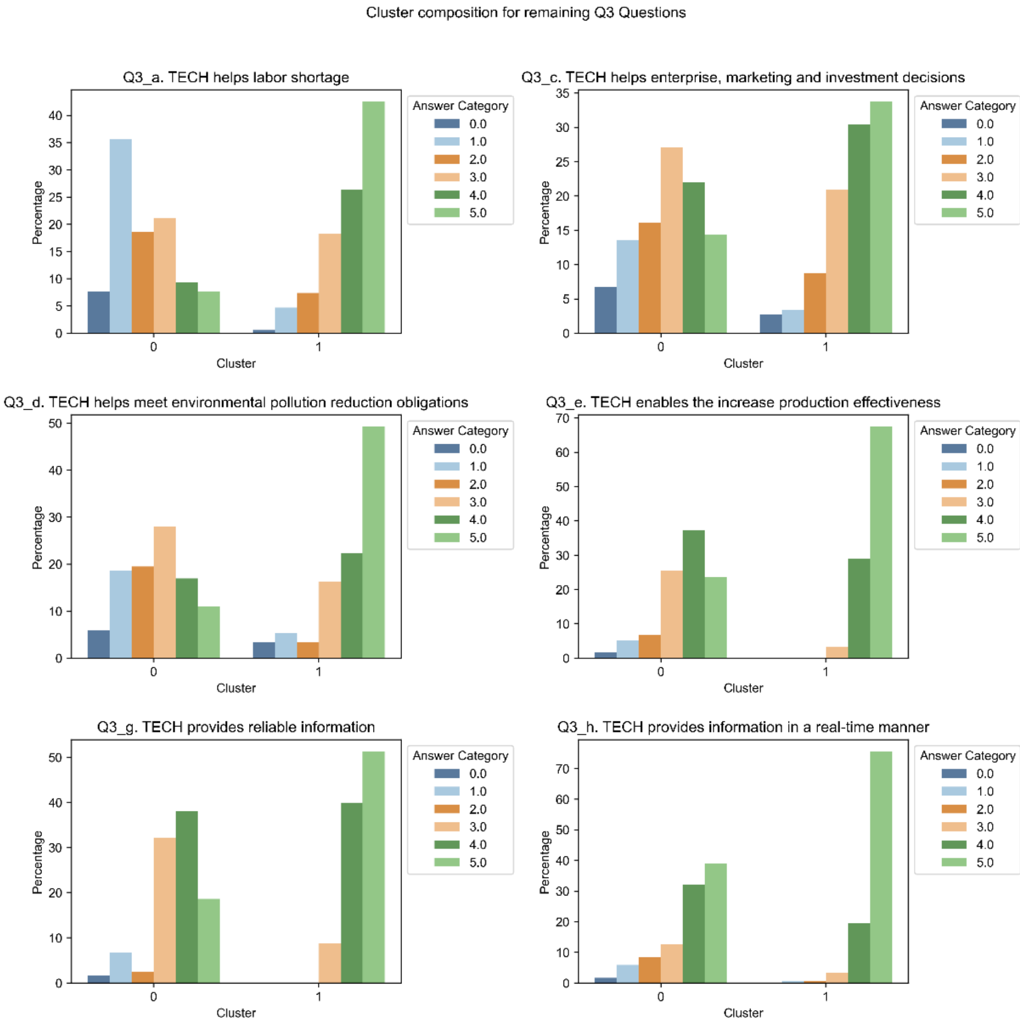

Figure 5. The figure shows the distribution of survey questions in the third category that haven't been shown in the result section. Cluster 0 represents technologically non-ready and Cluster 1 technologically ready farmers.

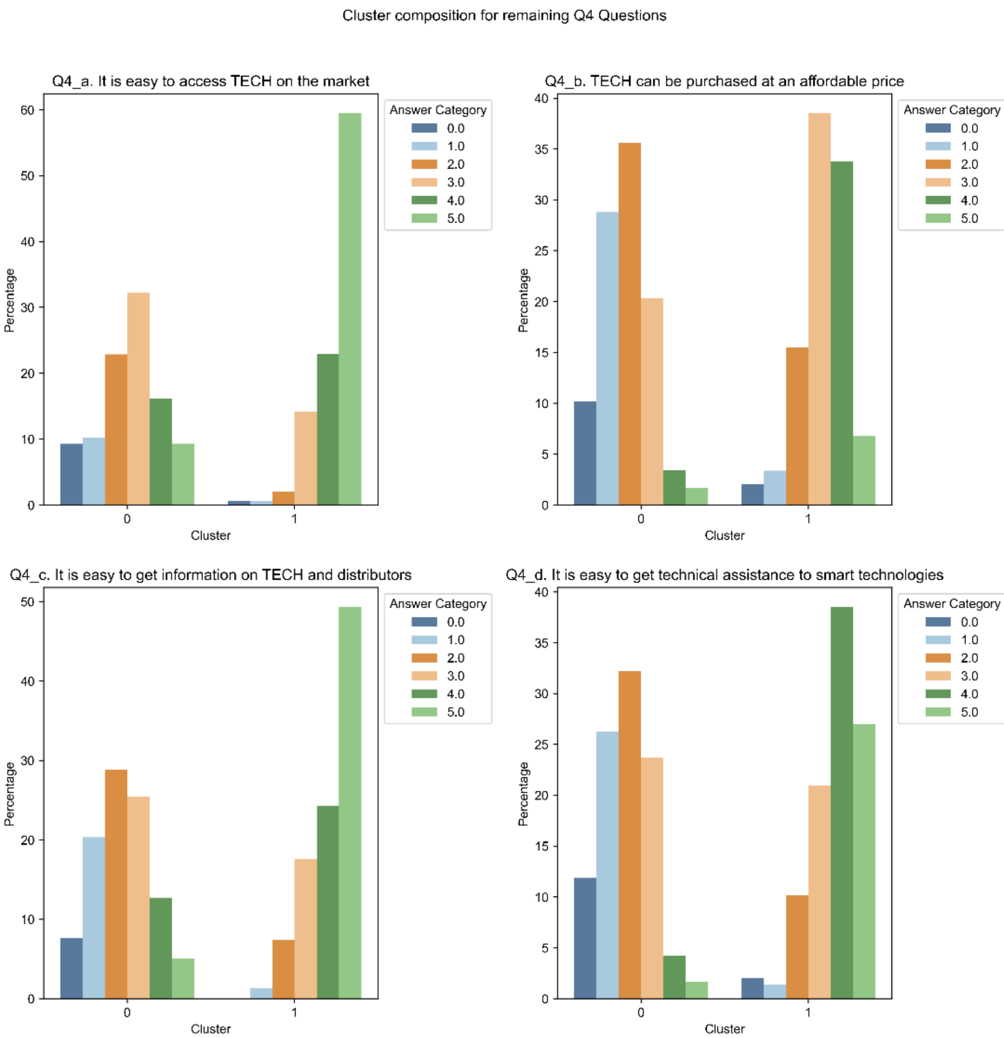

Figure 6. The figure shows the distribution of survey questions in the fourth category that haven't been shown in the result section. Cluster 0 represents technologically non-ready and Cluster 1 technologically ready farmers

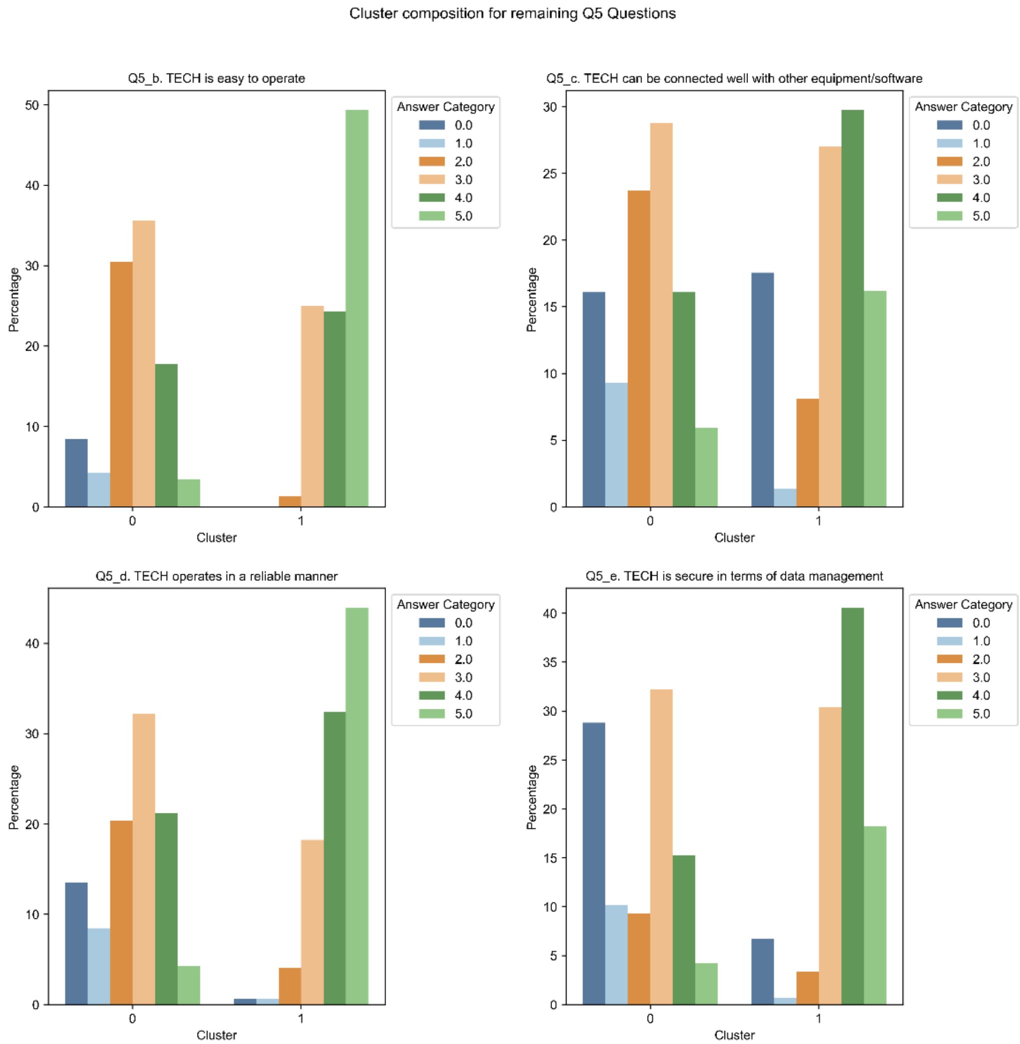

Figure 7. The figure shows the distribution of survey questions in the fifth category that haven't been shown in the result section. Cluster 0 represents technologically non-ready and Cluster 1 technologically ready farmers.

APPENDIX A2 – ***Survey questions used in this study***

Table 6. All survey questions and subquestions that have been used in this study are represented here as well as their range of answer categories

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 you use them or not on your farm. (0: I don't know, 1: Strongly disagree… 5: Strongly agree)
Q3_a	help/support to cope with labour 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 an increase in the effectiveness of production.
Q3_f	provide reliable information.
Q3_g	provide information in real-time.
	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 for 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)
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.)