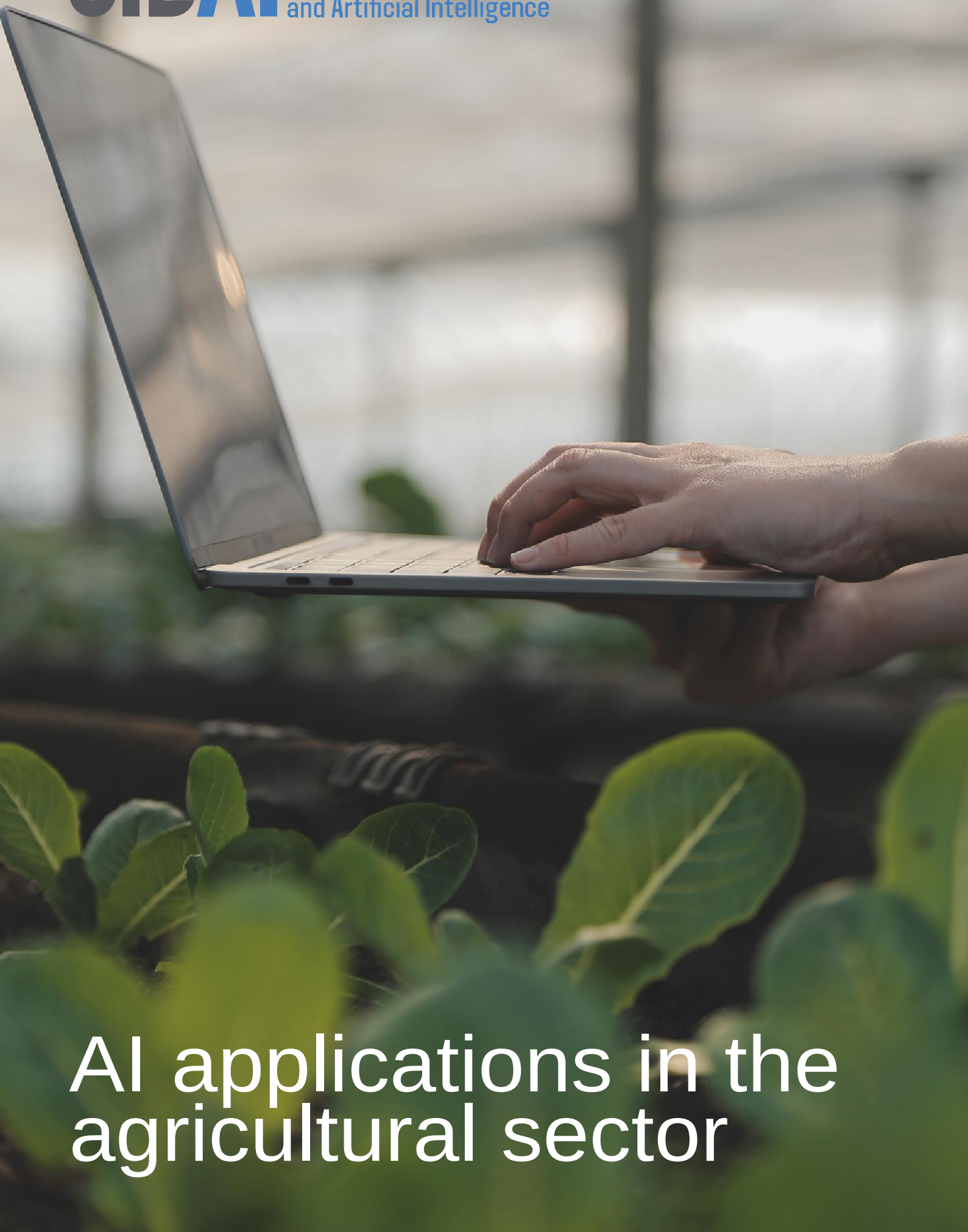




AI applications in the agricultural sector



[CIDAI-PAI 02-2022-D07](#)

This work is available under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) licence.

Under the terms of this licence, you may copy, redistribute, and adapt the work for non-commercial purposes, provided the work is properly cited as detailed below.

No use of this work should suggest that CIDAI endorses any specific organization, product or service. Use of the CIDAI logo is not permitted.

If you adapt the work, you must license your work under the same Creative Commons licence or an equivalent. If you create a translation of this work, you must add the following disclaimer along with the suggested citation: "This translation was not created by the Centre for Innovation for Data Technology and Artificial Intelligence (CIDAI). CIDAI is not responsible for the content or accuracy of this translation. The original edition in English will be considered authentic and binding."

Any mediation related to disputes arising from the licence will be conducted in accordance with World Intellectual Property Organization mediation rules.

Recommended citation: CIDAI-High-Impact Project: AI applications in the agricultural sector. Final report Ref.: CIDAI-PAI 02-2022-D07.

Authors and collaborators

The following authors participated in the preparation of this report:

Name	Organization	Participation
Pol Torres	EURECAT	Coordination, drafting and review
Joan Mas	CIDAI	Review
Fernando Vilariño, Coen Artens	CVC	Drafting of the second use case (AI-LAB and computer vision) along with the document review
Karina Gibert, Dante Conti	IDEAI-UPC	Drafting of the first use case
Mili Voltes, Ruth Gay, Alba Jiménez	Government of Catalonia (Service of Animal Feeding and Livestock Production Security)	Review of the first use case
Iris Pallarol, David Avaro	BSC	Drafting of the second use case (governance and sustainability model)
Roberto García	UdL	Drafting of the second use case (data space)

The presented results were achieved through the collaboration of the following experts:

Name	Organization	Participation
Pol Torres	EURECAT	Definition of use cases and project coordination
Fernando Vilariño, Coen Artens, Edgar García	CVC	Definition of AI-LAB and development of the second use case
Karina Gibert, Dante Conti, Claudia Agüero	IDEAI-UPC	Implementation of the first use case
Mili Voltes, Ruth Gay, Alba Jiménez	Government of Catalonia (Service of Animal Feeding and Livestock Production Security)	Definition and expert knowledge of the first use case
Iris Pallarol, David Avaro	BSC	Development of the governance and sustainability model
Roberto García, Rosa Gil, Lluís Miquel Pla, Pol Llagostera	UdL	Development of the data space and support in the second use case
Daniel Babot	CEP	Logistics support and expert knowledge in the second use case.

Contents



1. EXECUTIVE SUMMARY	5
2. INTRODUCTION	7
3. SUPPORT FOR DECISION-MAKING ON ANTIBIOTIC USE IN LIVESTOCK FARMS	9
4. COMPUTER VISION FOR INDIVIDUALIZED MONITORING OF PIGS ON FARMS	27
5. ANNEX	50
6. REFERENCES	67

Executive Summary

Content



The "AI Applications in the Agricultural Sector" project was conducted under the high impact projects initiative of CIDAI (*Centre of Innovation in Data Technologies and Artificial Intelligence*), funded by the Government of Catalonia. Contributors to the project's development on behalf of CIDAI included Eurecat, the Computer Vision Centre (CVC), the Data Science and Artificial Intelligence Research Centre at the Polytechnic University of Catalonia (IDEAI-UPC), the Government of Catalonia, and the Barcelona Supercomputing Center (BSC). The University of Lleida (UdL) and the Centre for Porcine Studies (CEP) also provided collaboration as external partners.

The project has addressed two distinct use cases within the agricultural sector. Firstly, the Service of Animal Feeding and Livestock Production Security of the Government of Catalonia's desire to enhance knowledge about antibiotic use in livestock farms gave rise to the development of decision support tools for antibiotic application. Secondly, discussions with sector representatives from the Lleida region resulted in a partnership with the Centre for Porcine Studies to create artificial intelligence and computer vision algorithms that track the life cycle of pigs, aiming to enhance animal welfare. This collaboration also highlighted the necessity for a secure data sharing platform. In this context, the expertise of the UdL in this field has been recognized, particularly due to its involvement in initiatives like Gaia-X, which is a European project designed to establish an open, transparent, and secure data ecosystem in Europe. Consequently, a proof of concept for a data space in the agricultural sector was selected. Concurrently with the development of the second use case, a governance and sustainability report was drafted. This report aims to set up a methodology for such projects and pinpoint the requirements for deploying the developed technologies, as well as their actual impact on the region. The employed methodological framework incorporates the AI LAB tool, which encompasses the three outlined dimensions (ecosystem challenge identification, technological solution and data infrastructure development, and governance, sustainability, and business model formulation) to enhance scalability and open up opportunities and business prospects beyond the initial pilot.

The initial application of decision support tools for antibiotic use in livestock farms employs a *Knowledge Discovery in Databases (KDD)* approach. This encompasses data collection, preprocessing, and data mining to assess and interpret antibiotic usage data in Catalan livestock farms. The conducted analysis concludes that the zootechnical category significantly influences antibiotic consumption. However, the foundational indicator, mgUR (mg Unit of Reference), does not correspond with the broader national reference indicator, IR (Reference Index), in fattening farms.

The second use case involves monitoring the pig life cycle to enhance animal welfare, employing computer vision tools and neural networks to distinguish and track individual pigs within a pen over time. Metrics for movement/rest and pen location have been gathered, providing a foundation for the creation of new metrics aimed at deriving indicators pertinent to animal welfare. Additionally, the data space proof of concept has enabled us to investigate a reliable and secure method for exchanging data and algorithms, as well as to examine potential monetization avenues. Finally, the governance and sustainability report, derived from *background and foreground research*, stakeholder meetings, and direct interactions with sectoral agents, offers an initial assessment of the potential value enhancement of these technologies.

All project participants have entered into a Confidentiality Agreement to ensure secrecy throughout the project's duration and to safeguard data exchange. The Service of Animal Feeding and Livestock Production Security of the Department of Climate Action, Food and Rural Agenda of the Government of Catalonia has supplied anonymized data concerning antibiotic use in Catalan livestock farms. Additionally, the CEP has contributed recordings from a pig farm situated on its experimental grounds.

Introduction

Content



This document compiles the outcomes of the high-impact project in the agro-livestock sector (PAI-AGRO) carried out in 2022-2023, with contributions from CIDAI Eurecat, CVC, IDEAI-UPC, BSC, and the Government of Catalonia. Collaborative efforts also included the University of Lleida and the Centre for Porcine Studies.

PAI-AGRO is a multidisciplinary initiative centred on agriculture and livestock in Catalonia, proposing two use cases that address significant challenges in the sector: decision support tools for pattern identification in antibiotic application on livestock farms, and life cycle analysis of pigs to enhance animal welfare. The project aims to create intelligent tools to tackle these challenges and to define a roadmap for the ongoing development of the technological solutions devised.

The initial use case emerged from a specific requirement of the Service of Animal Feeding and Livestock Production Security (Department of Climate Action, Food and Rural Agenda of the Government of Catalonia) to detect patterns and minimize the use of antibiotics on livestock farms across the nation.

To improve engagement with the region, a second use case involved the creation of an AI LAB centred on the Lleida territories. This AI LAB addresses enduring challenges within the framework of the RIS3CAT 2030 shared agendas (Catalunya, n.d.).

The AI LAB offers a dynamic framework to tackle enduring challenges, embracing the shared agendas and test-before-invest principles. Challenges and solutions are clearly delineated around TRL 6, which, together with a business model and sustainability, permits investment decisions (invest) post-validation of the test (*test*). This test should be viewed not merely as technological validation but also as an assessment of the ecosystem, enabling the identification of market factors, structural, and regulatory barriers to establish a viable business model.

It is crucial that CIDAI's resolved challenges correspond to the actual challenges faced by the region. Co-creation with interested communities is vital, and selecting challenges that align with the medium to long-term objectives extends beyond the project's scope, ensuring the integration of the nation's industrial fabric into these challenges in the future. Conversely, the participation of CIDAI partners in co-creation and governance definition processes facilitates the integration of AI technology into market-relevant challenges by utilizing cutting-edge AI-based technological solutions, thereby offering a competitive edge to Catalan business sectors. Accordingly, a sustainability proposal has been formulated to stimulate investment in these solutions, predicated on transparent intellectual property management models (both background and foreground) that support the establishment of a viable business model for the mid-term.

Support for decision-making on antibiotic use in livestock farms

Content



3.1 Introduction

Antimicrobials play a crucial role in preventing and combating diseases in humans and animals. However, recent years have seen a decline in their effectiveness against organisms and diseases that were once quickly and effectively neutralized. This rise in antimicrobial resistance poses a grave threat to both public and animal health. Infections that were once considered mild can now become prolonged, worsen, and may even result in death, while treatments that were effective in the past have lost their efficacy.

Antimicrobial resistance generally describes the ability of various microorganisms, including worms, parasites, bacteria, viruses, and fungi, to withstand the drugs designed to treat the infections they cause. Multidrug-resistant microorganisms are those that have developed resistance to a wide range of antimicrobials.

In the case of bacteria, they have developed resistance to antibiotics, the drugs designed to fight them.

This antimicrobial resistance stems from bacteria's natural capacity to mutate and become resistant. The rise in antimicrobial resistance is due to multiple factors, including the excessive and improper use in livestock production and human healthcare.

Addressing this issue requires a reduction in the use of antimicrobials and the necessity for their use. As a global challenge, it necessitates collective action (One Health approach). Coordinated efforts at various levels are underway to devise a solution, engaging experts from all impacted sectors.

A crucial aspect of combating resistance is understanding antibiotic consumption in livestock farming.

The Ministry of Climate Action, Food and Rural Agenda conducts a quarterly assessment of antibiotic consumption. This is to ascertain the actual usage of these substances on farms based on veterinarian prescriptions, alongside data on farm capacity, livestock census, and movements.

The consumption metric is derived from the quantity of antibiotics prescribed (in milligrams) and the livestock count considered on the farm (using UR: units of reference).

Challenge

The study aims to identify the factors influencing the varying levels of antibiotic use across different farm contexts and to pinpoint deviations in behaviour among similar farms.

The findings should enable the establishment of correlations between various study parameters and antibiotic usage, thereby understanding how different elements relate to antibiotic consumption and focusing on the significant ones.

Goals

The primary goal is to ascertain the connection between various contextual factors (such as farm features and management) associated with livestock farming and the consumption of antibiotics (both in type and frequency).

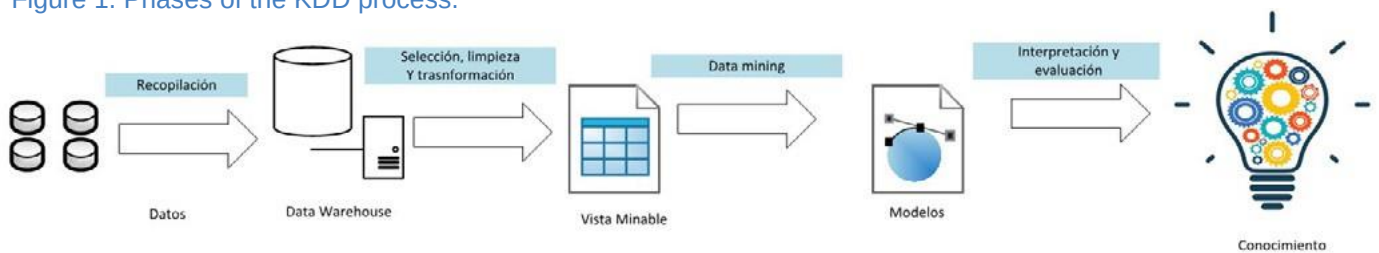
To accomplish this, it is essential to first establish strata that group farms with analogous characteristics and then gather the requisite data for each domain.

Although the research in this instance has concentrated on the pig industry, the approach and smart platform developed can be applied to other animal types, provided that comparable data is accessible.

3.2 General description of the KDD methodology

Data processing was conducted using the KDD (Knowledge Discovery in Databases) methodology. KDD can be described as "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad, 1996). The process encompasses several stages, including data preparation, pattern identification, knowledge assessment, and profiling. It's not merely a straightforward calculation but an intricate procedure of exploration and deduction. A pattern qualifies as knowledge when it surpasses a user-defined interest threshold, though this doesn't equate to a philosophical or extensive definition of knowledge. Here, knowledge is specific to the user and domain, shaped by the criteria and thresholds established by the expert.

Figure 1. Phases of the KDD process.



Source: (<https://bit.ly/3CHqG1o>)

Specifically, the phases of the KDD process are defined as follows:

- Data collection
- Preprocessing: which includes data selection, cleaning, and transformation
- Data mining
- Evaluation and interpretation

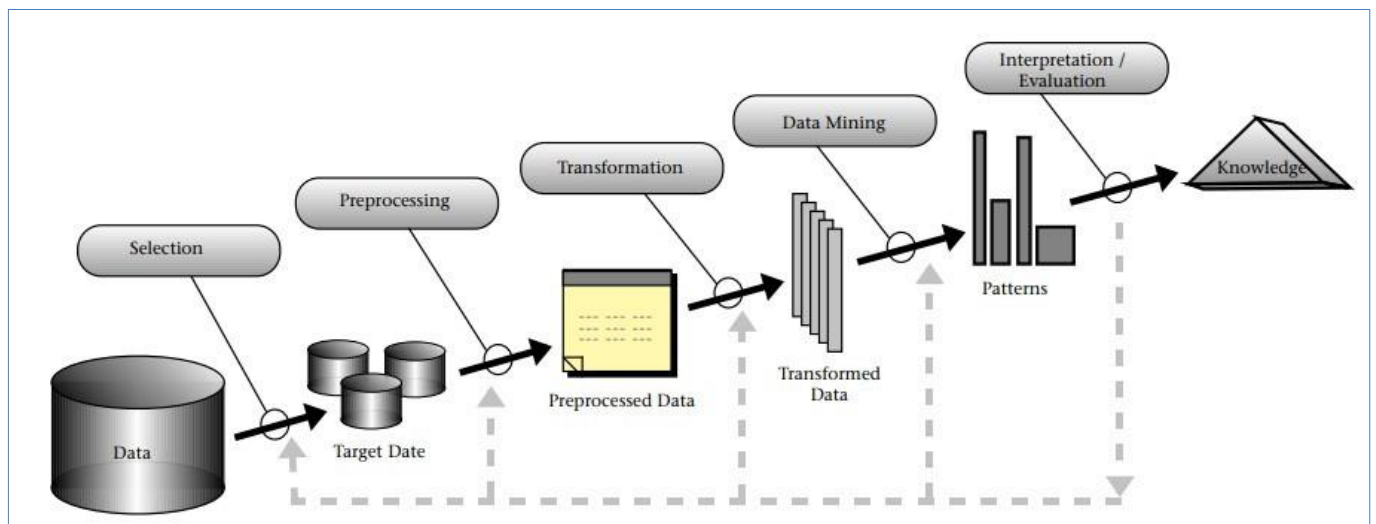


Figure 2. Overview of the KDD methodology steps (Fayyad, 1996).

Additional details of the experimental procedure are available in the ANNEX (A1, A2).

33. Analysis of experimental data

This chapter outlines the conclusive experiments conducted in accordance with expert recommendations, centring on the analysis per kilogram of animal, with a focus on the utilization of mgUR variables.

Final master table

This section displays four entries of the variables incorporated into table F1, including:

- Disaggregated consumption variables and their corresponding percentages, to ensure they align with the variables listed in F1
- It also introduces four variables deemed crucial by experts for the analysis: is_premixed, NOT_premixed, ADS NAME and staggered
- Additionally, it includes variables from auxiliary tables to enhance the analysis.

Figure 3. Master table variables

Anio_Trimestre	mgConsumo_B	mgConsumo_C	mgConsumo_D	pct_mgConsumo_B	pct_mgConsumo_C	pct_mgConsumo_D
2021_2	250000	6300100	22311000	0.87	21.83	77.3
2021_2	44250	7020000	10112564	0.26	40.87	58.87
2021_2	1505000	19895150	38330000	2.52	33.31	64.17
2021_2	0	0	5566000	0	0	100

es_premezcla	NO_premezcla	CODI POSTAL EXPLOTACIÓ	SERVEI TERRITORIAL EXPLOTACIÓ	PROVINCIA EXPLOTACIÓ	COMARCA EXPLOTACIÓ	MUNICIPI EXPLOTACIÓ
NO	0.818181818181	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca
NO	0.757575757575	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca
NO	0.891304347826	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca
SI	0.4	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca

NOM ADS	total_animals	prom_bloc1	prom_bloc2	prom_bloc3	prom_punt_total	esglaonada
1 NA		0	0	0	0	1
1 NA		0	0	0	0	0
1 NA		0	0	0	0	1
1 NA		0	0	0	0	0

- Is_premixed: mode of prescriptions discriminating between yes/no premix (binary)
- Not_premixed: fraction of no premix reported in the data
- ADS NAME: 0 is no description (assuming no ADS) and 1 otherwise (binary)
- staggered: if the farm is staggered (binary)

The final master table is designed for internal analysis and, importantly, for presenting graphical analyses on the Power-BI dashboard. This facilitates intelligent and user-friendly decision-making by experts in the field.



34. Part 1. Multivariate analysis of variables

Correlation analysis with all numerical variables

The following figures show the correlation between the variables that have been analysed.

Figure 4. Corrplot of all numerical variables in table F1

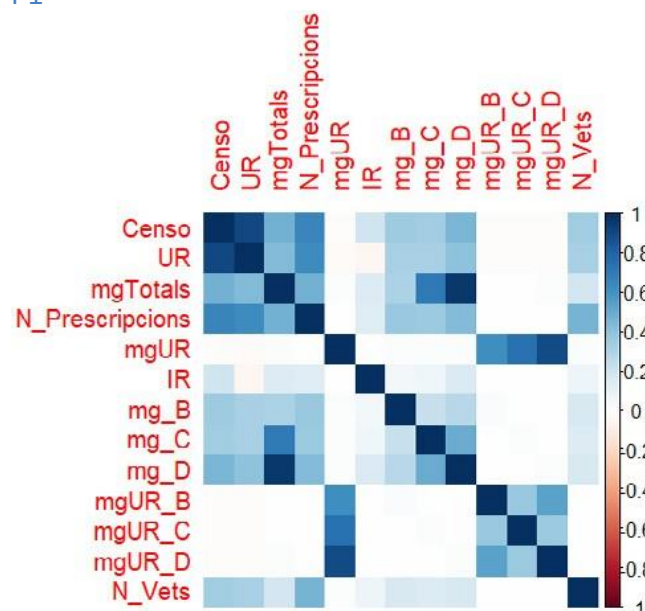


Figure 5. Biplot of all numerical variables in table F1

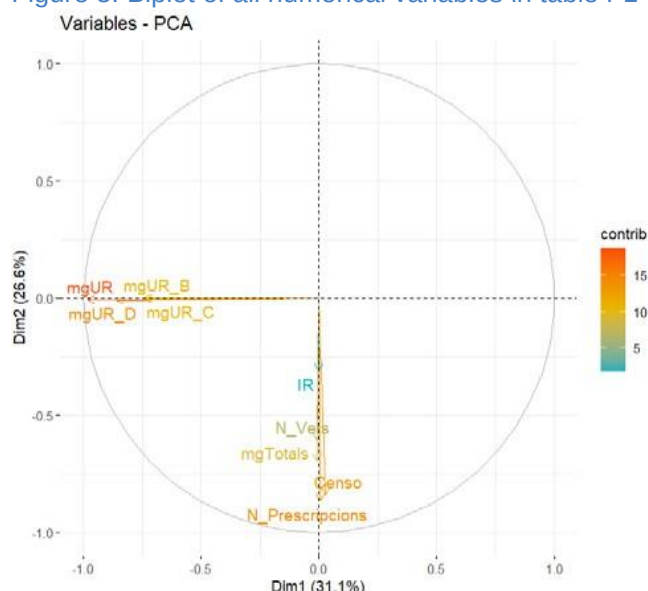
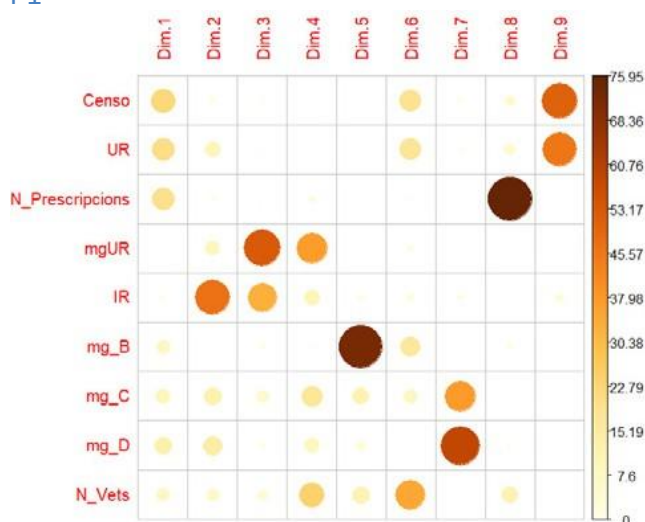


Figure 6. Heat map of all numerical variables in table F1



The data ranked the three most frequent zootechnical categories. The third column displays their percentage representation.

Table 1. Top 3 most frequent zootechnical categories (F1)

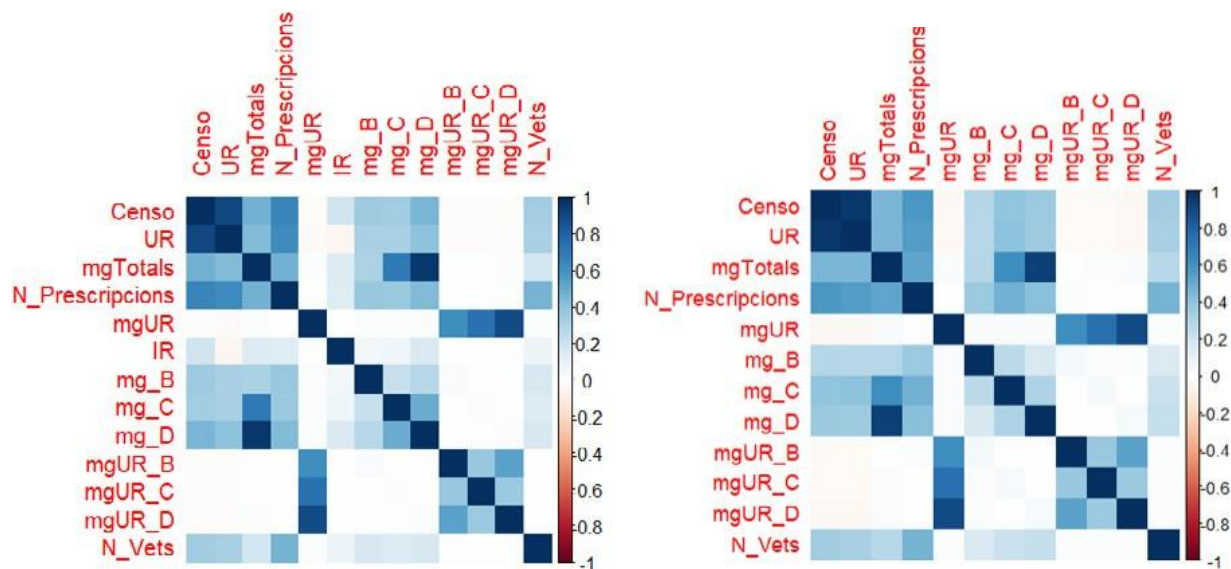
	number	%
Engreix	29871	72.505947
Producció cicle mixt	5346	12.976358
Transició de garrins	2574	6.247876

It has been noted that fattening farms constitute a significant majority (72.5%), making them the predominant category. This suggests that the analyses preceding this report may be skewed by the prevalence of fattening farms rather than by the census variable of farm size.

Consequently, a multivariate analysis will be conducted, distinguishing between these three zootechnical categories, which represent 91.6% of the total explanations, to validate or verify the experts' hypothesis.

1. Comparison of the overall results with those of fattening farms

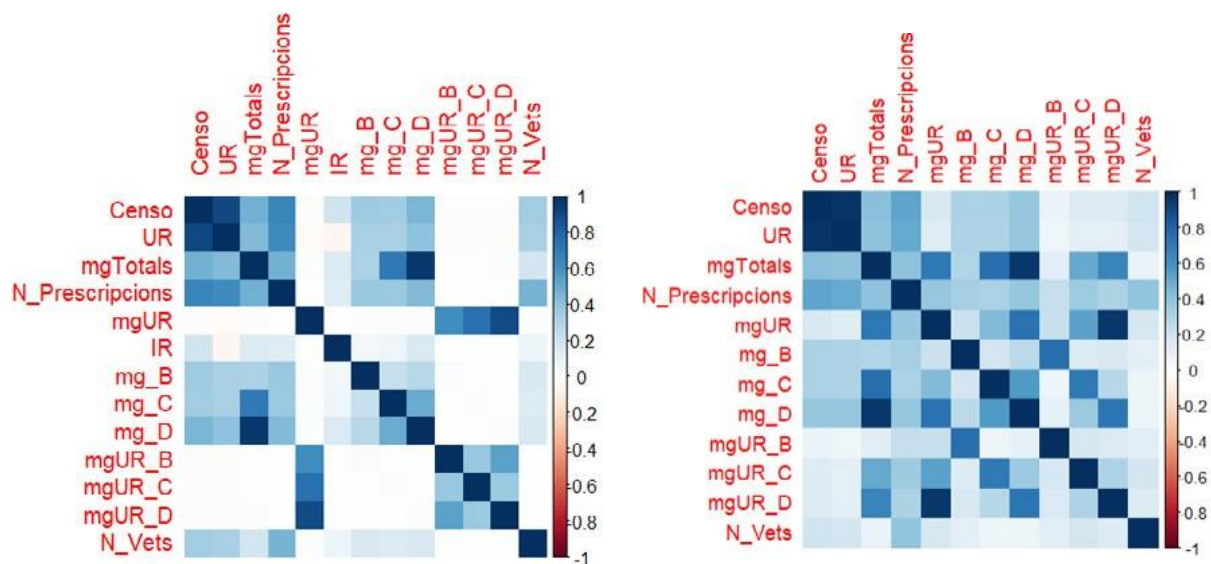
Figure 7. On the left: corplot of all records. On the right: corplot of the records of fattening farms only



The patterns observed in the analysis of overall farms are also evident in the farms within the Fattening zootechnical category.

2. Comparison of overall results with mixed-cycle production farms

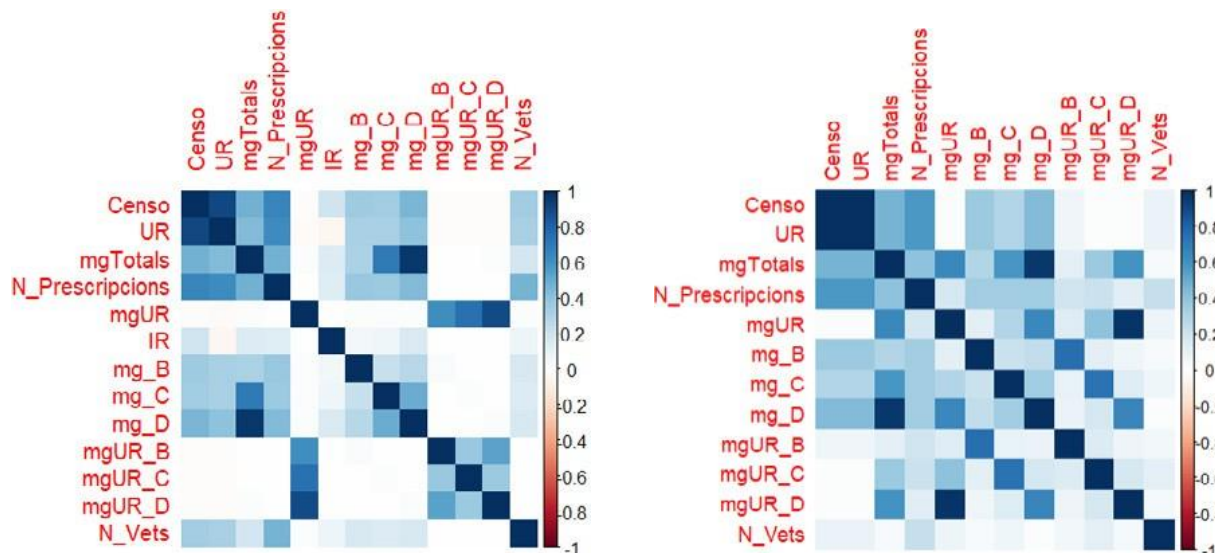
Figure 8. On the left: corplot of total records. On the right: corplot of records of mixed cycle production farms only



In this instance, the association of consumption variables alters. There appears to be a correlation between farm size (as per census), RUs, and the number of prescriptions with the consumption measured in milligrams per RU.

3. Comparison with the overall results observed in piglet transition farms

Figure 9. On the left: corplot of F1 records. On the right: corplot of piglet transition only records



In this instance, the correlation plot is akin to that of point 2. Farms within these zootechnical categories exhibit distinct antibiotic consumption profiles compared to the aggregate of pig farms.

Given that fattening farms constitute 72.5% of all farms, the decision was made to conduct two concluding Principal Component Analyses, differentiating by zootechnical category in two separate experiments:

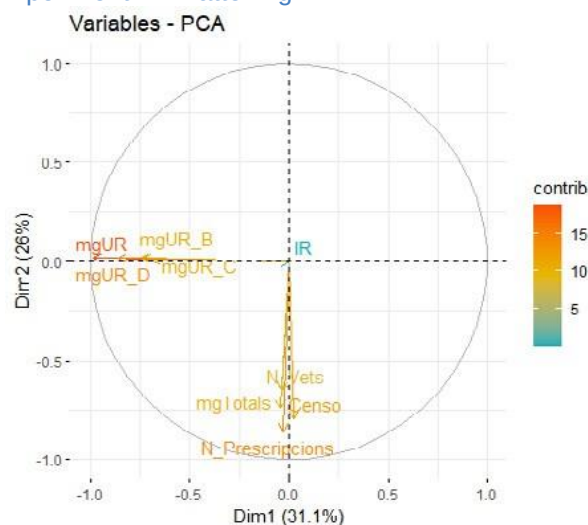
- Experiment 1: records of "Fattening" farms
- Experiment 2: farm records of all other categories except "Fattening".

Experiment 1

Principal component analysis reveals that:

- Variables pertaining to consumption correlate strongly with farm size.
- Additionally, the categories of antibiotic consumption (B, C, D) and the total mg/UR show a high degree of association among themselves.
- However, the national reference index (RI) does not show a correlation with the antibiotic consumption in fattening pig farms in Catalonia.

Figure 10. Biplot of the multivariate analysis for Experiment 1 – Fattening



The subsequent figures illustrate the correlations among the various variables. The RI is constant, thus it possesses its own dimension.

Figure 11. Heatmap of the multivariate analysis for experiment 1 – Fattening

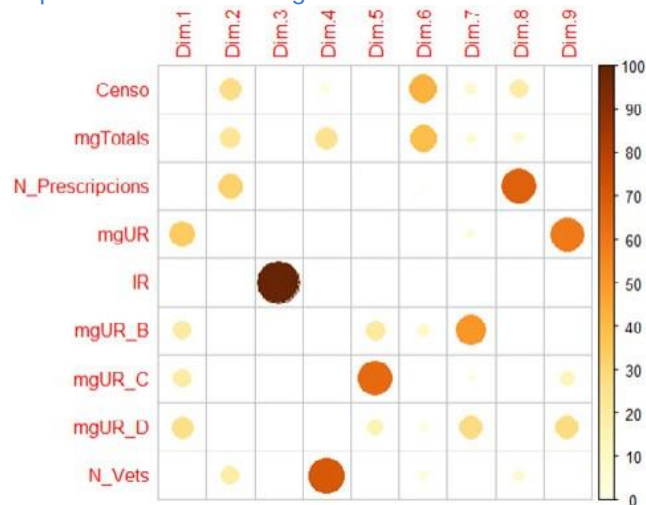
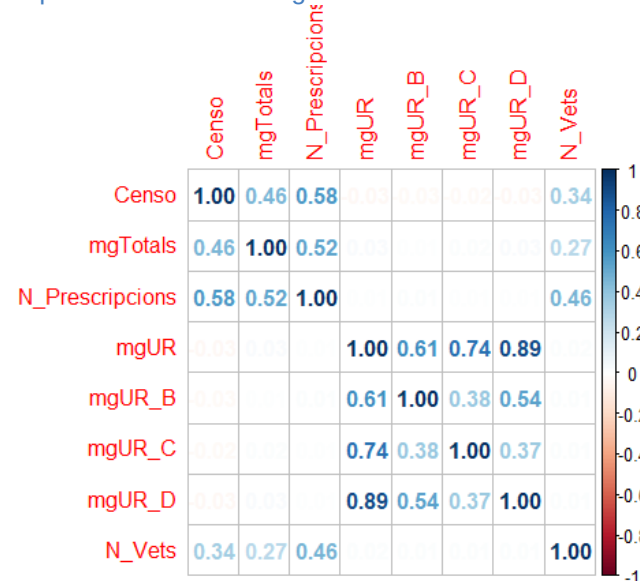


Figure 12. Corrplot of the multivariate analysis for experiment 1 – Fattening



The correlation analysis suggests that:

- Antibiotic consumption (measured in mgUR) and the reference index (RI) are independent.
- There is a negligible negative correlation between the census and antibiotic consumption. It is observed that as farm size increases, antibiotic consumption decreases, and the inverse is also true.
- There is a strong association between farm census, antibiotic consumption, and the number of veterinarians.
- Moreover, antibiotic consumption (mgUR) shows a high correlation with antibiotic D consumption (0.89), and substantial correlations with antibiotics C and B (0.74 and 0.61, respectively).

Experiment 2

Principal component analysis reveals that:

- The primary variables influencing antibiotic consumption on non-fattening farms are mgUR, followed by mgUR of type D, which are closely associated.
- The intake of antibiotics D and C shows a strong correlation.
- Conversely, the consumption of type B antibiotics is independent of these variables.
- Additionally, the variables Censo, N_Prescriptions, and N_Vets show a high degree of interrelation.

The subsequent figures illustrate the correlations among the various variables:

Figure 13. Biplot of the multivariate analysis for Experiment 2

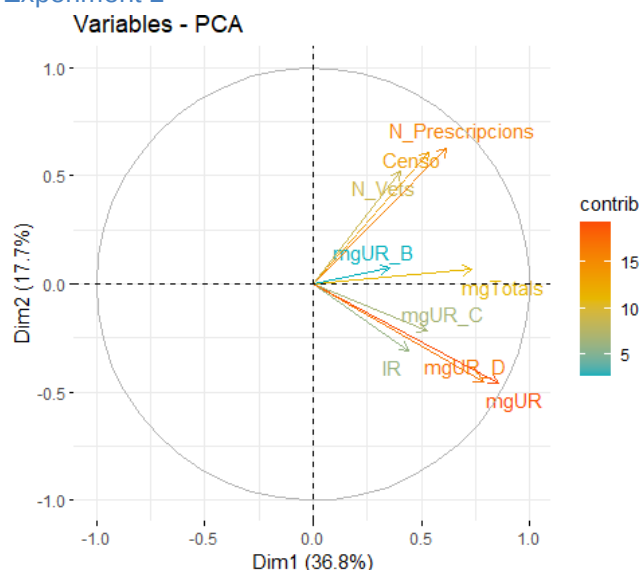


Figure 14. Heatmap of the multivariate analysis for experiment 2

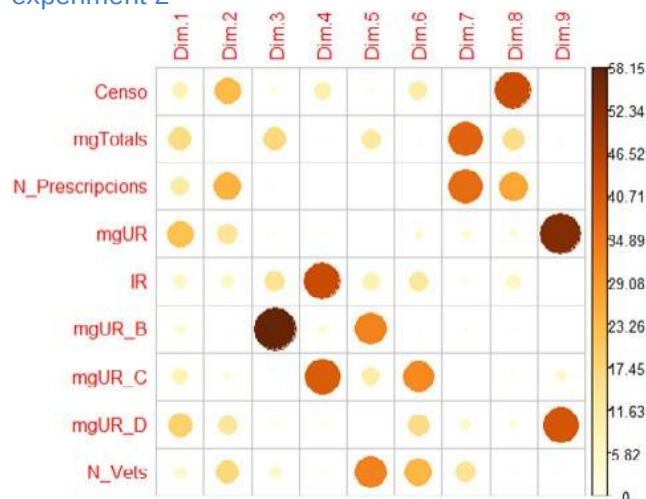
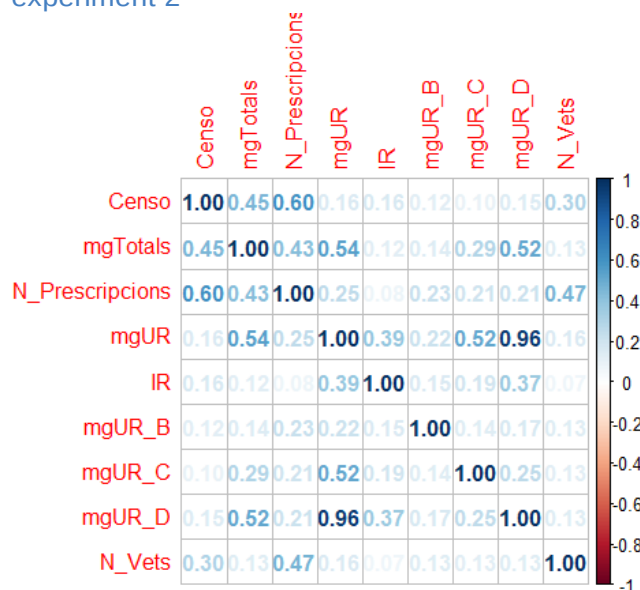


Figure 15. Corplot of the multivariate analysis for experiment 2



The correlation analysis suggests that:

- RI and the consumption of antibiotic D are highly correlated.
- The consumption of drug B remains independent from the other two categories.
- There is a strong relationship between mgUR and mgUR
- The N_Vets is associated with the farm census and the quantity of prescriptions



35. Part 2. Clustering and profiling

Hierarchical clustering seeks to identify distinct patterns that differentiate holding codes. A random sample of 15,000 records has been collected. The variables chosen for the analysis are listed below.

Table 2. Clustering, scenario 1. Selected variables

Variables
Census
IdCZ
N_Prescriptions
mgUR
mgTotal
mgUR_B
mgUR_C
mgUR_D
N_Vets

Table 3. Composition in number of classes

Classes	1	2	3	4
Number of farm IDs	10,819	1,227	1,956	998

The largest class is the first, and the smallest the fourth.

Table 4. Classes by zootechnical category

	1	2	3	4
Altres	0	319	0	0
Centres d'inseminació artificial i centres de recollida de semen	0	89	0	0
Engreix	10818	0	0	0
Producció cicle mixt	0	0	1956	0
Producció de cicle tancat	0	291	0	0
Reproductors	0	186	0	0
Transició de garrins	0	0	0	998
Transició de reproductores nul·lipares i cria de reproductors	0	342	0	0
Varies Zoo	1	0	0	0

In the classification by zootechnical category, it is noted that classes 3 and 4 correspond exclusively to farms with mixed cycle production and piglet transition, respectively. A similar situation is found in the first group, which is the largest; all entries are classified under "Fattening," with one entry under "Varies Zoo." Class 2 encompasses a diverse range of zootechnical categories, establishing a pattern of similarity.

Statistical validation

The outcomes of the aforementioned final analysis have been statistically validated. The tests for the variables mgUR and IdCZ are presented here.

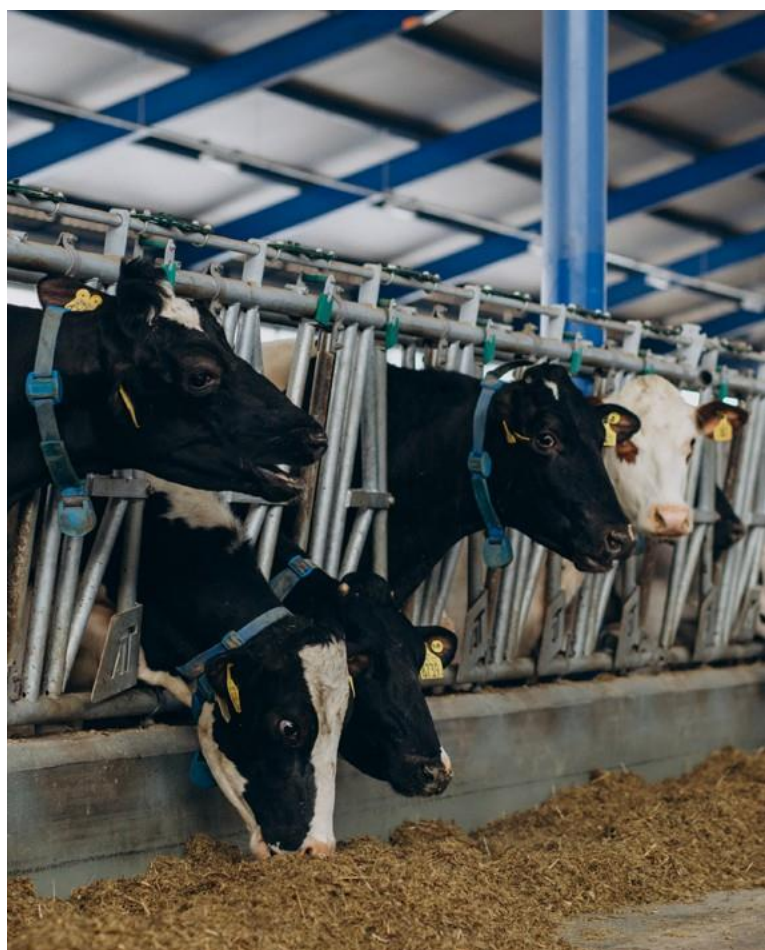
Kruskal-wallis rank sum test
data: samp\$mgUR by samp\$c4
Kruskal-wallis chi-squared = 4679.5, df = 3, p-value < 2.2e-16

Pearson's Chi-squared test
data: samp\$IdCZ and samp\$c4
X-squared = 123594, df = 24, p-value < 2.2e-16

Both tests demonstrate statistical significance. Consequently, we can deduce that:

- in at least one of the groups, the consumption of milligrams of antibiotic per UR differs from the others.
- Similarly, in at least one group, there is a difference from the others in terms of the zootechnical category.

The statistical validation for the remaining numerical and categorical variables was conducted identically. All tests yield statistically significant results.



Final profiling

In the final profiling, classes have been colour-coded as a guide or indicator based on antibiotic consumption per mgUR, following the analysts' criteria. Classes requiring end-user review are marked in red, indicating higher severity, while those with more stable consumption are marked in green.

Table 5. Final guide to the classification of farms

Class	Farm ID	Description of the classes
1	10819	Farms of medium size, production and consumption. They generally have six prescriptions and primarily use antibiotic D. On average, there are two veterinarians present. Essentially, these are fattening farms.
2	1227	Farms of medium to low size, production and consumption. Significant use of antibiotic D and typically one veterinarian on average. There is considerable variability in the zootechnical categories, with the transition of nulliparous reproducers and the breeding of reproducers being the most common.
3	1956	Farms of large size, production and consumption. They have the highest number of prescriptions, averaging 30, and predominantly use antibiotic D. On average, there are two veterinarians present. They are exclusively associated with mixed cycle operations.
4	998	Farms of the highest size, production and consumption. They have 27 prescriptions on average and a very high use of antibiotic D. On average, there are two veterinarians present. They are exclusively associated with piglet transition farms.

36. Part 3. Final results with all records

Assignment of classes by the CURE method

The CURE algorithm has been utilized for class assignment of out-of-sample records, addressing the challenge of managing large databases in R, as previously described.

Table 6. Class size

Classes	1	2	3	4
Number of farm IDs	29878	3400	5346	2574

Final profiling

This section presents the graphical representation of numerical variables as a Class Panel Graph (CPG), facilitating the comparison of classes based on these variables (Gibert K. G.-R.-S., 2008).

Figure 16. CPG for the variables Censo, UR, and N_Prescriptions.

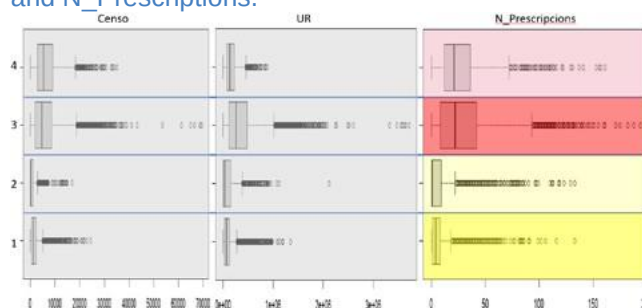


Figure 17. CPG of the variables mgTotal, mgUR, mgUR_B

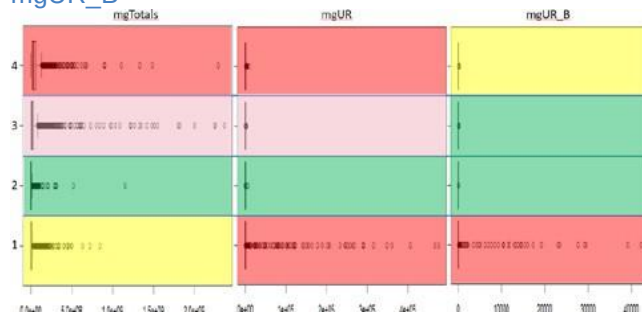
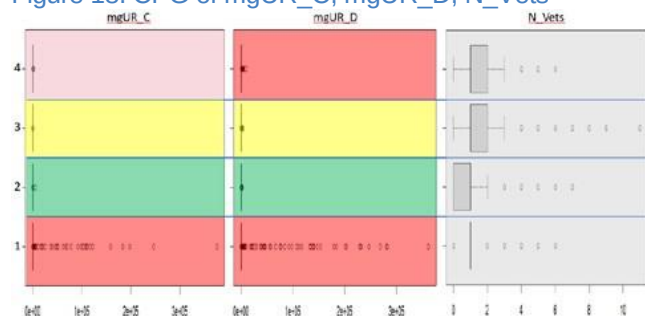


Figure 18. CPG of mgUR_C, mgUR_D, N_Vets



The Class Panel Graph (CPG) is a visual tool that enables detailed analysis of variable behaviour conditioned to each class. The figures presented illustrate this behaviour in the numerical variables utilized during the clustering process. The colour coding within the CPG serves as an additional profiling visualization tool, offering the end user a user-friendly interpretation by correlating these colours with those of a conventional traffic light.

In our context, these colours were allocated based on a comparison of the class-conditioned centroids, considering both the mean and the distribution depicted in the CPG boxplots. Consequently, green signifies positive behavioural values, while red indicates the contrary, depending on each variable's implicit measure (for instance, red mgUR denotes increased antibiotic usage). Ultimately, tallying the colours observed in each class on the traffic light sets a trend for the class's definitive colour or profile, as demonstrated in the final descriptive profiling table (Table 5).

Below are the categorical variables for each class. These complete the profiling of the classes.

Figure 19. Barplot of the DescZ variable

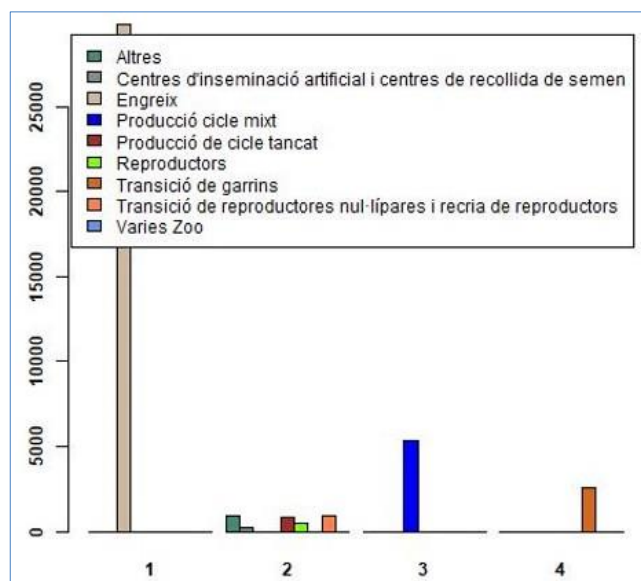
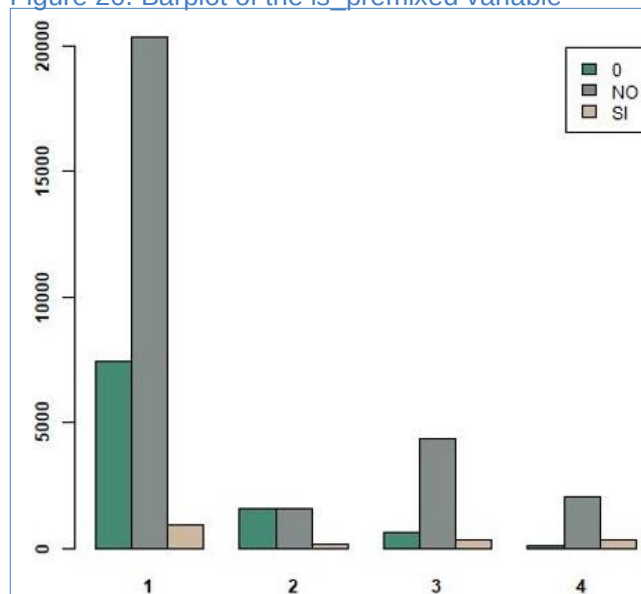


Figure 20. Barplot of the is_premixed variable



In group 2, there is an equilibrium between farms that administer antibiotics through feed (premixes) and those that use different pharmaceutical forms. For the remaining groups, there is a dominance of farms that utilize methods other than medicated feed for medication.

Figure 21. Barplot of the Territorial Service variable

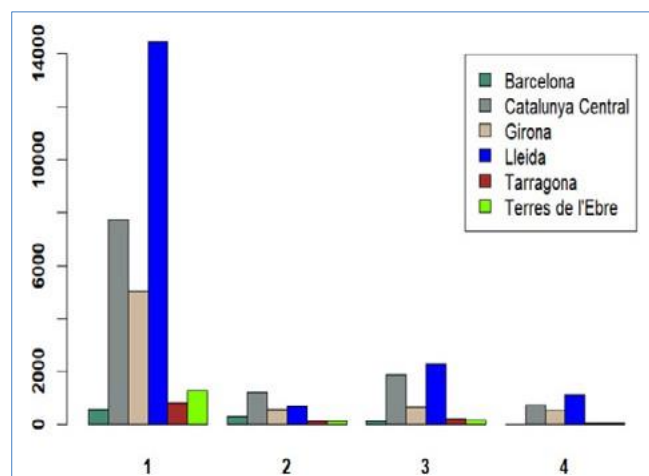


Figure 22. Barplot of the RI variable

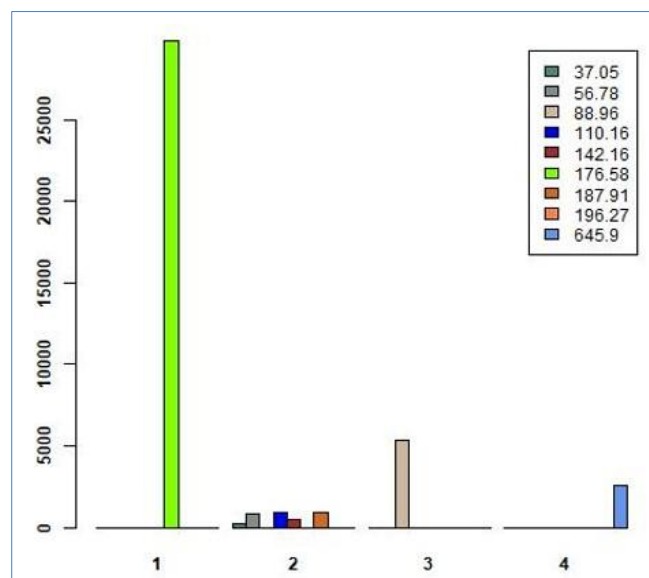


Figure 23. Barplot of the Province variable

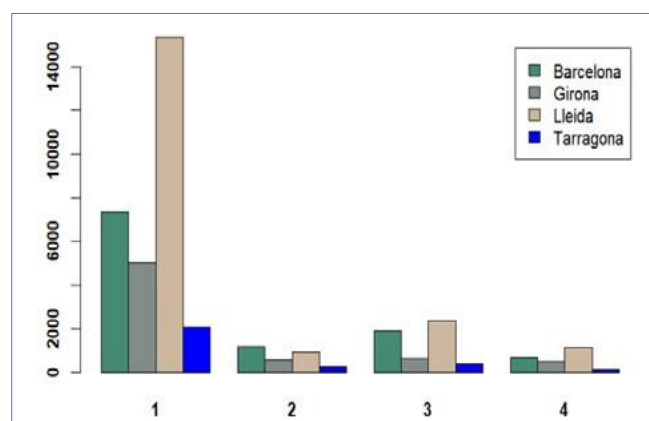


Figure 24. Barplot of the ADS Name variable

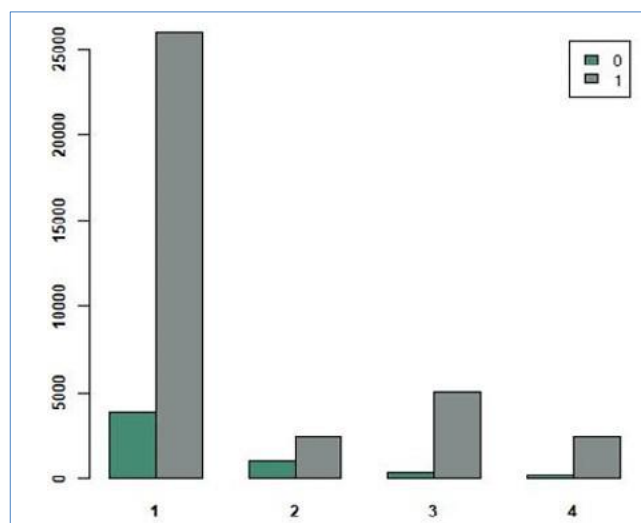
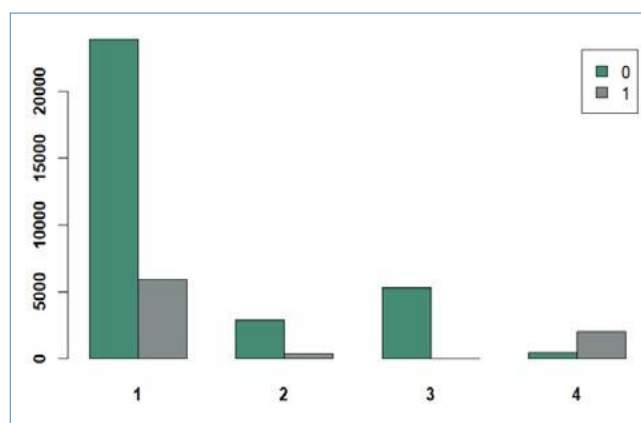


Figure 25. Barplot of the Staggered variable



3.7. Viewer – RanchingAnalytics.cat

Viewer description

Having centralized and debugged the data, we have moved on to designing and implementing a visualizer on the Microsoft Power BI platform. This visualizer is designed to enhance the information gleaned from the data analysis and previously described models, providing additional graphical support in a user-friendly manner. It presents explanatory views of antibiotic usage in Catalan farms through a narrative format, or storytelling, which initially focuses on general indicators by geographical sector, followed by a detailed examination of each individual livestock farm. This iteration of the visualizer (RanchingAnalytics.cat) was developed in collaboration with experts and is the culmination of ongoing interaction throughout the project's development.

Audience

RanchingAnalytics.cat is intended for the general public but has been tailored to include terminology and concepts unique to livestock farming, particularly concerning antibiotic consumption indicators in this field. It features a section where a glossary, links, and other pertinent information can be provided to assist users who are less familiar with the subject matter.

Dissemination medium

RanchingAnalytics.cat can be published on the Internet. This version is internally managed. The decision of its dissemination must still be considered and decided by the owners of the data (Government of Catalonia).

Main Views (features and storytelling)

Ranchinganalytics.cat is a visualizer structured into three primary sections: (1) An introductory section featuring presentation, credits, and a glossary of terms; (2) A section outlining the usage of antibiotics in pig farms across Catalonia, starting with general indicators and progressing to a more detailed view; and (3) A final section providing detailed information for each individual farm code (livestock farm), including a comparative view that allows for the evaluation and exploration of two farms simultaneously, with the capability to perform this analysis at the levels of province, municipality, farm territorial service, and county.

Figure 26. Main page of the viewer

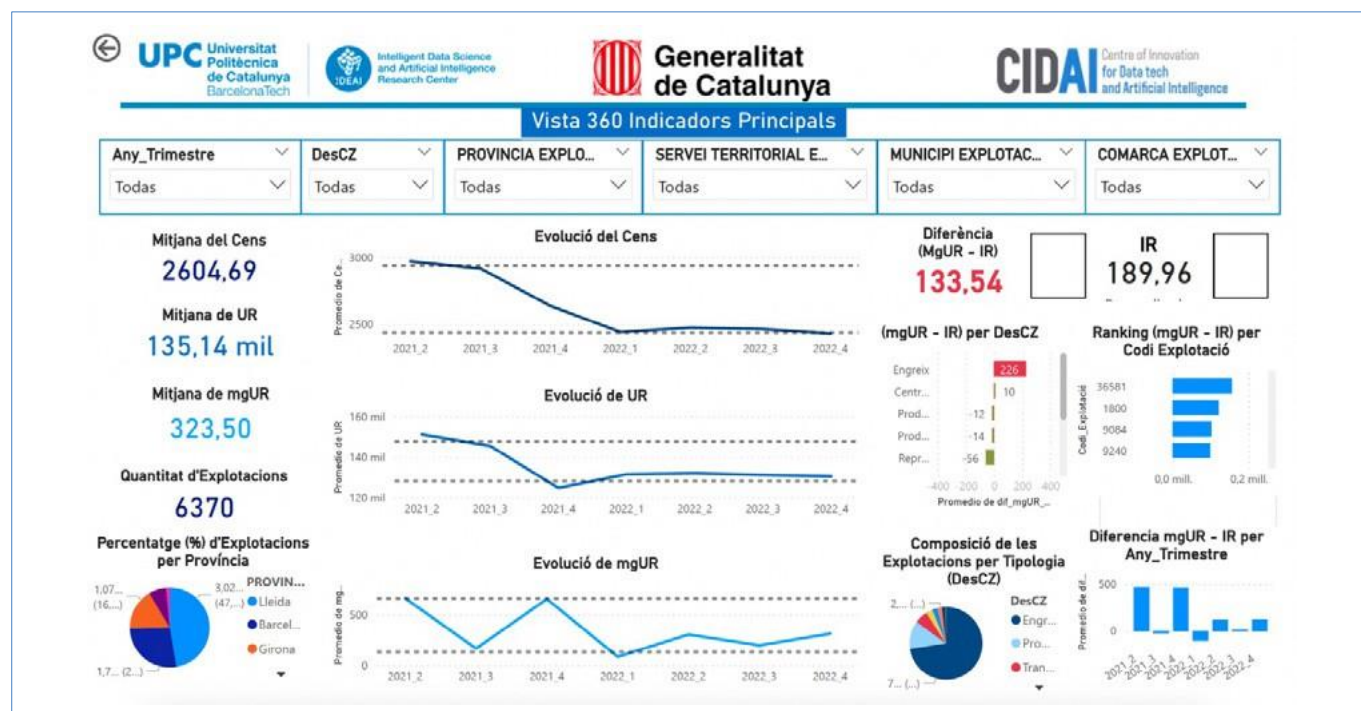


Section 1 - Use of antibiotics in pig farms in Catalonia

The 360 View of Leading Indicators

Provides a comprehensive overview of key metrics related to antibiotic usage and characteristics of the swine industry in Catalonia. It features interactive filters for different levels, such as: Quarter, Zootechnical Description, and geographical location. The platform displays trends in Census data, UR, mgUR, and the variance between Catalonia's reported mgUR and the national reference index, along with various descriptive charts. Additionally, help buttons are available to aid users in interpreting the graphs.

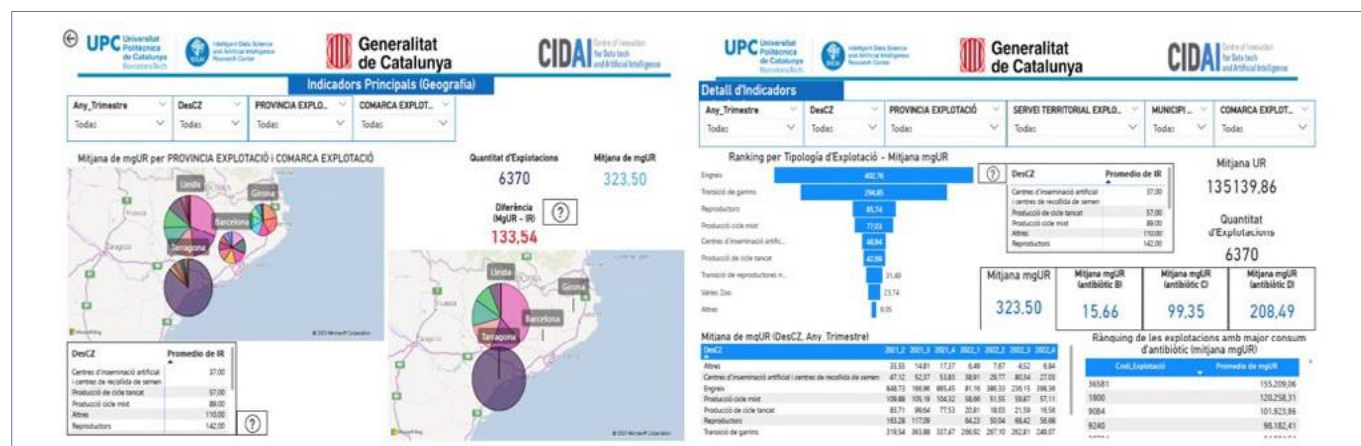
Figure 27. Main indicators view



Details of Indicators

Here the analysis is enhanced through geographic visualizations and detailed Key Performance Indicators (KPIs) of antibiotic consumption by typology (B, C, and D).

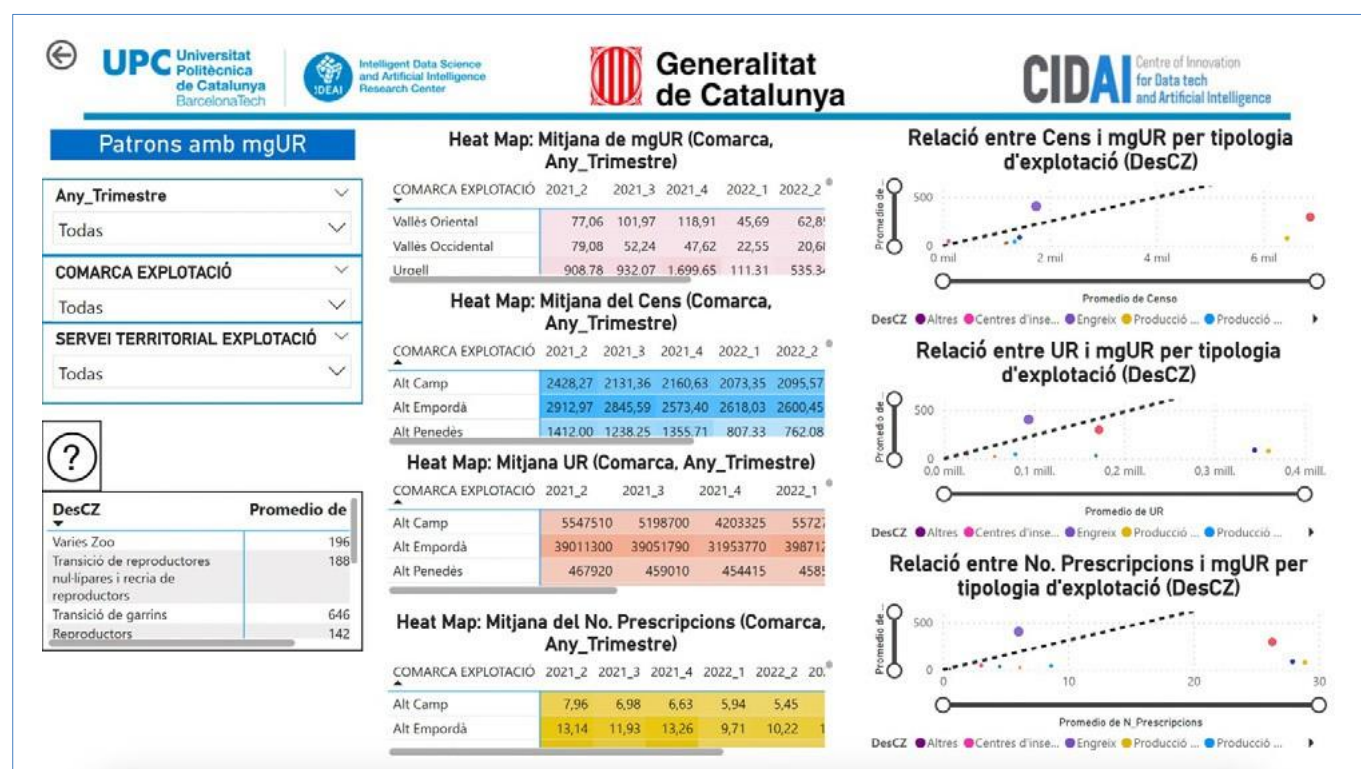
Figure 28. Details of indicators



The patterns and relationships view

This view displays the explanatory connections between farm census data and antibiotic consumption, between urinary retention incidents and antibiotic usage, and lastly, between the number of prescriptions and the rate of antibiotic consumption. These tables are categorized according to the zootechnical classification of swine farms.

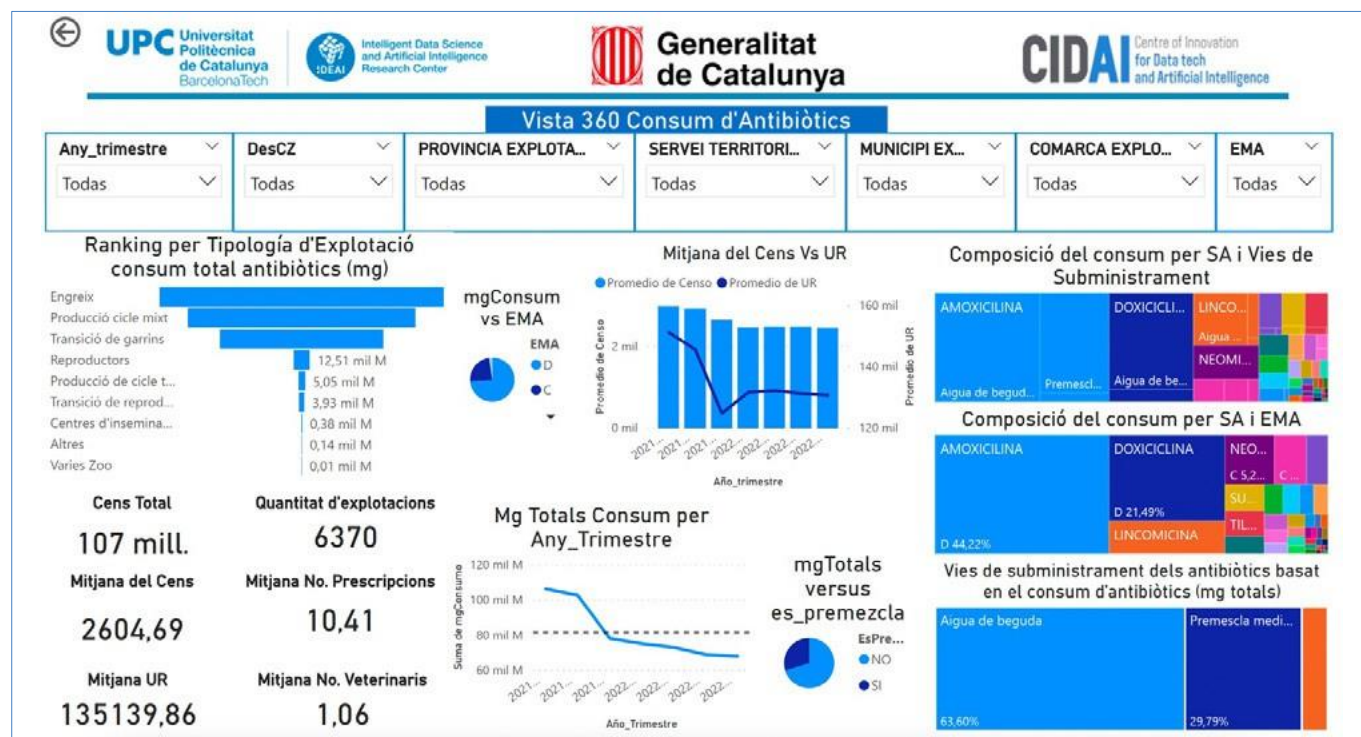
Figure 29. The patterns and relationships view



Detailed view (ZOOM) of antibiotic consumption in Catalonia

This view provides an in-depth look at antibiotic use showcasing the routes of administration, the most utilized antibiotic families, trends, and composition graphs. Filters facilitate the exploration of results by time periods, farm locations, and the EMA classification of antibiotics.

Figure 30. Detailed consumption view



Section 2 - Use of antibiotics in pig farms in Catalonia

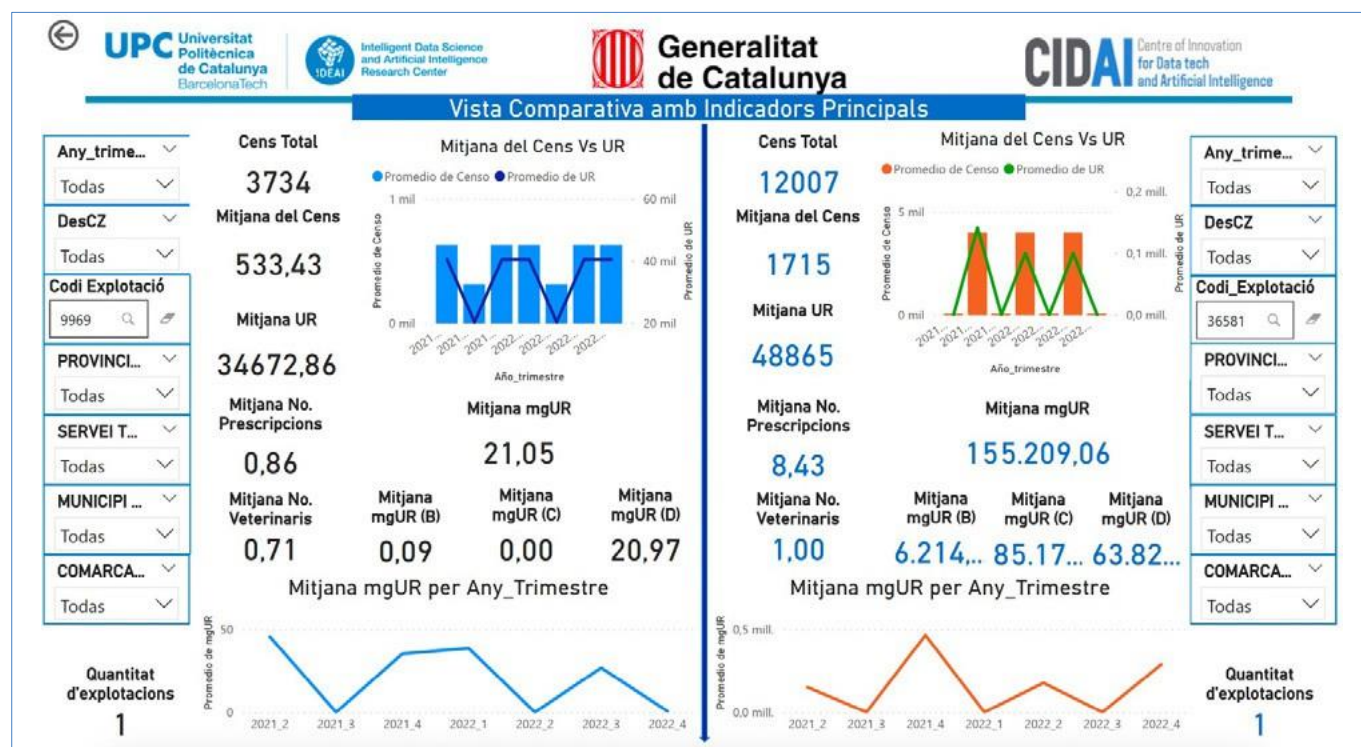
Detail at the farm code level: the views are akin to those depicted above, yet they contain individualized information for each farm code in the database. Due to data privacy concerns, these views are omitted from this report.



Final view

This view facilitates the simultaneous comparison of the majority of the previously described information. It serves as an influential tool in the decision-making process, as it enables the comparison of various farms and time periods, among other factors. (Note: The two farms compared in the figures have been anonymized to ensure data privacy).

Figure 31. Comparative view with main indicators



2.8. Conclusions

In the development of this initial use case, the study's primary goals were met. An analysis involving various scenarios and experiments was conducted to delineate the correlation between antibiotic consumption variables and farm characteristics, identifying the zootechnical category as a significant discriminant in antibiotic usage. Additionally, the study uncovered deviations in antibiotic consumption patterns in pigs, revealing four distinct groups of farms within Catalonia's pig farming sector. Moreover, Principal Component Analysis experiments have shown that the base indicator, termed mgUR by experts, does not correspond with the national reference indicator index (RI) for fattening farms.

Computer vision for individualized monitoring of pigs on farms

Cont



4.1. Introduction

This initiative stems from the overarching requirement for all High Impact Projects to extend beyond their pilot phase, by offering tools that enable scalability and seamless market integration. It embodies the AI LAB concept outlined in the Catalan Artificial Intelligence Strategy¹, tailoring it to a particular application setting. Thus, the High Impact Project serves as an experimental environment for the tool itself. The imperative to progress past the pilot phase and furnish business models based on innovation validation responds directly to the demands of Catalonia's productive stakeholders, who seek a transparent understanding of the investment returns associated with validating technologies in tangible experimental settings.

At the methodological level, the AI LAB follows five distinct stages:

1. mapping and identifying challenge owners in the territory;
2. co-creating the solution proposal;
3. implementing the technical prototype;
4. designing and implementing the data space prototype;
5. creating the business model and sustainability plan.

At the project level, following the conducted mapping, a proposal has been made to take action in Lleida to bolster the farming community. Various meetings with local stakeholders, including the Federation of Agricultural Cooperatives of Catalonia, representatives from the Diputació de Lleida, and regional companies, have highlighted the Centre for Porcine Studies as a pivotal collaborator for knowledge exchange and the joint development of a high-value technological solution for both the pig industry and the region. The project development team has been established as follows:

- Eurecat
- Computer Vision Centre (CVC)
- Centre for Porcine Studies (CEP)
- University of Lleida
- Barcelona Supercomputing Center (BSC)

Challenge

The primary challenge of this case study was dual:

1. Firstly, to pinpoint the key stakeholders in the agricultural and livestock sectors of Lleida who are keen on developing and adopting technological solutions for integration into a technology innovation process.
2. Subsequently, after identifying these actors and challenges, a technological solution was devised and implemented at the Porcine Studies Centre of Lleida, orchestrated by the Eurecat project. Following the AI LAB approach, this solution consists of 3 different deliverables: Technology prototype, data space, and governance, sustainability and business model.
 - **Technology prototype:** A Computer Vision-based system has been developed to individually monitor the life cycle of each pig within a stall, aiming to detect behavioural patterns. The initiative is orchestrated by the Computer Vision Centre.
 - **Data space:** Simultaneously, the development of a data space for the agricultural sector has been recognized as a challenging task. This initiative has been spearheaded by the University of Lleida.
 - **Governance, sustainability and business model:** The project's development phase included the challenge of creating a governance and sustainability model. This model aimed to distinguish the roles of the involved agents and outline a roadmap to guarantee the project's continuity and operation. It also sought to define steps that would encourage the creation of other similar initiatives. The Barcelona Supercomputing Center (BSC) has spearheaded this effort.

¹ "The CATALONIA. AI LAB serves as a collaborative and experimental space, acting as an open portal to global ideas and initiatives that are gathered and indexed: social innovation by and for the people. It aims to enhance the technology sector by providing access to infrastructure and tools, and concurrently, to foster open innovation with businesses and public entities through involvement in pilot projects." CATALONIA.AI. The Artificial Intelligence Strategy of Catalonia. Government of Catalonia Ministry of Digital Policies and Public Administration Secretariat of Digital Policies. 2020

Goals

Within the specified contextual framework, the project objectives are as follows:

- Transfer the knowledge and technology from cow monitoring to a pig farm setting (CVC, 2022).
- Gather valuable information that enables farm managers to assess animal behaviour, thereby facilitating actions that ensure and enhance animal welfare.
- Create a proof of concept for a data space that adds value to both data and algorithms.
- Formulate a governance and sustainability model for technological advancements.



Socioeconomic impact

The challenges tackled in this initiative are closely aligned with the key stakeholders of the region and the tangible assets of CIDAI partners. Moreover, the project's impact is in harmony with the Shared Agendas strategy endorsed by the Government, thereby ensuring a contribution to the long-term goals set out in the Smart Specialization Strategy of Catalonia. Additionally, the execution of this project has strengthened the digital transformation of the agri-food sector, in line with the planned activities of the Digital Innovation Hub for Agri-Food and Forestry of Catalonia. Regarding the specific innovations developed, the following impacts have emerged:

- The initiative, as envisioned by AI LAB, confirms a scalable tool that can be replicated across various sectors, not just agriculture, within the Catalan industrial landscape. This enhances the portfolio of strategic methodological tools for integrating AI into the industrial ecosystem. The computer vision-based solution has enabled the identification of a new strategic workstream that yields objectively verifiable metrics, which are in correlation with specific animal welfare indicators.
- The prototype data space developed is a trailblazer in Catalonia's "agro" sector, marking the beginning of a crucial tool for data value enhancement. It introduces a novel strategy for pig farms, encapsulating the concept of "from bits to kilos."
- Governance, sustainability, and business models, even at an initial stage, serve as essential reference tools for the industry. They guide how to quickly capitalize on investments made during the prototyping process and offer a versatile roadmap applicable across various industry verticals.,

The impacts mentioned previously serve as a foundation for potential future developments, leading to solid partnerships under the innovation transfer framework².

² The CIDAI team involved in this initiative has developed various spin-off proposals, which have been presented to the grant call for knowledge transfer demonstration activities ([PEPAC 2023-2027](#)).

4.2 Monitoring the life cycle of pigs to improve animal welfare

Context of the need for a technological solution

Livestock farming involves the breeding, raising, and trading of animals. While this sector remains largely traditional, it has begun to embrace new technologies like Big Data, the Internet of Things, and Industry 4.0. These advancements have given rise to the term 'Smart Farming,' which denotes the use of Information and Communication Technologies in agriculture and livestock management.

From the perspective of farmers and ranchers, smart agriculture ought to offer high added value by providing decision support tools for more efficient farm management. This concept is closely linked to three interconnected technological fields:

1. **Information management systems:** This refers to structured systems designed for the collection, processing, and dissemination of various of data essential for managing farm operations and functions.
2. **Precision agriculture:** Precision Agriculture involves the management of spatial and temporal variables to enhance economic performance and environmental impact. It encompasses support tools for integrated management aimed at optimizing inputs and conserving resources. This is facilitated through the use of GPS systems, ground observation systems, and aerial imagery obtained from drones and satellites, which enable the creation of maps with various spatial variables related to production, soil type, organic matter content, soil moisture levels, and more.
3. **Agricultural automation and robotics:** This refers to the application of robotic technologies, automated control systems, and artificial intelligence

across all levels of agricultural and livestock production, including the utilization of agricultural robots and drones.

Ultimately, smart farming applications are designed not only for large conventional farms, but can also serve as significant drivers for farms, cooperatives, and federated cooperatives, addressing the full spectrum of granularity into which the ecosystem is organized.

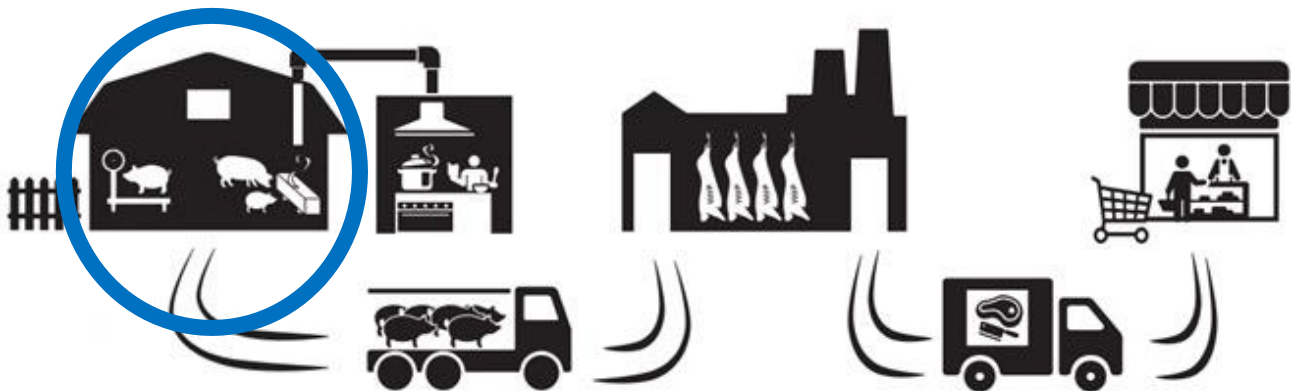
This use case established assesses how this paradigm assists in identifying and automating elements within the processes of the pork industry. From the perspective of production infrastructure, the pork sector encompasses: the farms where piglets are born, the farms where pigs for fattening are raised, the slaughterhouses where pigs are processed, and the logistics companies responsible for transporting the meat to supermarkets. In the absence of consumer associations, this cycle represents the complete value from farm to fork³.

The pork sector holds significant importance in Catalonia. According to the Institut d'Estadística de Catalunya, the pig sector, with approximately 6.5 million pigs, ranks second after poultry, which has over 43 million birds. Pig rearing takes place on fattening farms, where most of the pig population consists of fattening pigs. This fattening process lasts about four months, and when the pigs reach a final weight of around 110 to 120 kg, they are sent to the slaughterhouse.

In this project, various aspects of pig fattening were studied, and a set of sensors (cameras) were installed both on the farm and in the slaughterhouse. These sensors collect data to help evaluate, monitor, and improve the process.

³ From farm to Fork Strategy. For a fair, healthy and environmentally-friendly food system. European Union. 2020.

Figure 1. Diagram of the pig's life cycle. Enclosed in blue, the prototyping space of this project.



An essential factor in the pig fattening process is their diet. Traditionally, pigs are fed with dry feed, but there are trends experimenting with alternative types of feed, such as liquid feed, which is easier for the animals to digest. Some initiatives even consider introducing this type of feed on farms. Figure 1 presents a diagram of the pig's life cycle.

Pigs are fattened on farms. They enter the fattening farm when they are 10-11 weeks old and weigh about 20-25 kg. When they reach a weight of 110-120 kg, after approximately four months, they are transferred to the slaughterhouse. It is crucial to monitor the weight of each pig on the farm because this determines when they can be sent to the slaughterhouse. The slaughterhouse prefers pigs with a homogeneous weight to avoid adjusting their machines. Given this, it would be ideal to have a system that monitors the pigs' weight during the fattening process. The simplest solution is to use a scale that measures their weight in real time as they pass over it. However, for accurate weighing, the pig must remain still for a short period to obtain a stable weight reading.

Ultimately, apart from tracking the weight of pigs in the stall, there is a growing interest in monitoring their movements and behaviours. This concept is tied to the idea of animal welfare on farms: pigs should be kept in the best possible conditions during their fattening process, with objectively verifiable indicators.

Prototyping proposal

The proposed solution includes the following technological modules:

1. Image capture system: cameras installed in an experimental stall.
2. Pig segmentation system.
3. Individualized identification system for each pig.
4. Pig movement tracking system.

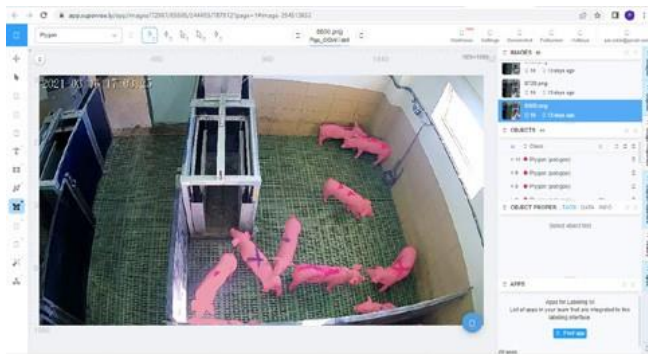
One method to monitor the animals is by using a camera installed in the stall, as illustrated in Figure 2, which shows a camera located at the CEP facilities. As part of the project, this camera was chosen to monitor the pigs in real time and study their movements.

Figure 2. CEP pig fattening farm. Experimental pen with pigs marked by identification. Above left is the scale. On the right, the water container.



For this reason, it was decided to implement an algorithm to detect pigs in the images. It was concluded that pixel-level detection was necessary to track the animals even when they are close to or touching each other. An instance segmentation neural network, specifically Mask R-CNN, was selected for this task. To train this network, a reference set of about 100 images was manually annotated (see Figure 3).

Figure 3. Pig identification. Manual annotation for the creation of the reference set for neural network training.



Once the network was trained, the detection of the animals proved to be very efficient, as illustrated in Figure 4. In this figure, the detections of the pigs are indicated with different colours.

Figure 4. Pig stall with R-CNN animal identification



However, accurately detecting pigs in the image is not enough for tracking; it is also necessary to relate the pigs in one video frame to the pigs in the subsequent frame. This process is known as re-identification and can be performed in several ways:

1. Using a model to predict the movements of the animals (e.g., with a particle or Kalman filter).
2. Using an algorithm to generate a unique identifier (a kind of fingerprint) for each pig. With these identifiers, the detections from two consecutive frames can be matched.

In the final implementation, a combination of different techniques is used to account for various scenarios. For example, due to the placement of the cameras, pigs may disappear from view temporarily because of occlusions by other pigs, and tracking must be re-established.

Once the tracking algorithm was implemented, an analogy was made that studying the life of a pig in a stall could be likened to analysing a football match. By tracking the animals, statistics can be generated similar to those produced for a game of football (see Figure 5).

Figure 5. Representation of the movement of individuals on a football pitch



There are several parameters worth estimating to gain an initial understanding of the animals' behaviour in the stall, including:

1. Paths of the pigs (distance, maximum speed).
2. Time spent in the scale area.
3. Time spent standing (potentially sleeping).
4. Time spent in the water area.
5. Heat maps by areas of preference.

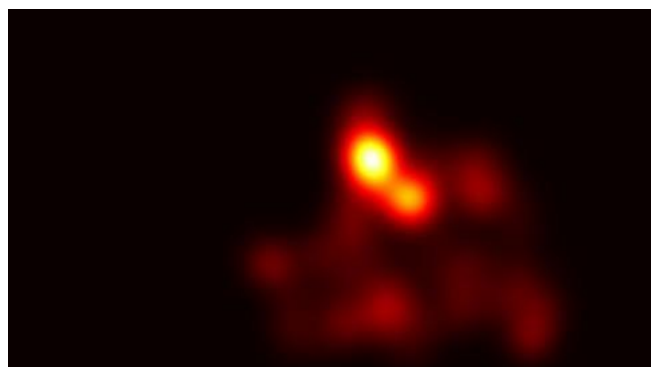
To illustrate these calculations, reference images were generated to summarize the paths taken by the pigs throughout the day. Figure 6 shows the paths of two different pigs.

Figure 6. Path of the pigs inside the stall. The trajectories followed by the centre of mass of two different pigs throughout the day are depicted in blue and orange. The image of the empty stall serves as a reference for the pigs' movements.



Heat maps have also been generated to visualize where the pigs spend their time. Figure 7 shows the results for a specific pig. This approach requires contextualization with respect to the rest of the pigs, which has not been studied in this project.

Figure 7. Heatmap of a pig inside the stall



4.3. PoC of data space in the agri-food sector

We are at the dawn of a new digital revolution, often called the fourth industrial revolution, largely driven by recent advancements in Artificial Intelligence (AI). AI models have the potential to transform industries and societies by automating complex tasks and analysing vast amounts of data.

However, for these models to function effectively, they require extensive data repositories. Typically, the more data (and higher quality data) we gather, the more precise and capable the models become. Acquiring and processing this data is a complex and costly endeavour, yet it serves as the backbone of this new digital era. Consequently, it is common to hear the phrase, "data is the new oil," reflecting its critical value.

Yet today, we frequently encounter "data silos"—data sets stored in completely isolated systems and infrastructures. This isolation makes it impossible for other entities to access and utilize the data, leading to a significant waste of its potential value.

Data spaces are ecosystems designed to address the issue of data confinement. They enable the secure sharing and utilization of resources between different entities. These spaces go beyond mere data storage; they incorporate functionalities that allow data of various types and origins to be shared and used securely. They establish a well-organized environment with clear rules, guidelines, and robust security measures.

On one hand, dataset owners can generate revenue by sharing their data with third parties while retaining control over its use and privacy. On the other hand, data spaces act as catalogues where entities and individuals can search for and utilize a diverse array of data and algorithms that operate within the ecosystem itself.

For instance, consider a pig farm that uses a camera to monitor the activity of the animals. While the images captured over time may have limited value to the farm owners themselves,

they could be incredibly useful to other individuals or entities. Therefore, these images can be offered in a data space for a certain fee, allowing the farm to profit from sharing the data while providing valuable resources to others.

The images are now accessible to any organization or individual interested in utilizing them. A second entity, specializing in analysing livestock behaviour, can use these images and manually annotate them, indicating the exact position of each animal within the images. These annotations constitute new data with significant value, enhancing the original images' worth. Together with the annotations, the images can now be applied to a broader range of applications. These annotations are also made available in the data space.

Finally, a third-party entity with an artificial intelligence model designed for pig detection and tracking can obtain this data and use it to develop or enhance its model. Optionally, this entity can also offer the improved model in the data space, enabling other interested parties, including the original farm owners, to benefit from its capabilities.

In this way, we can easily monetize and enhance the value of our data while fostering the creation of more advanced, mutually beneficial solutions.

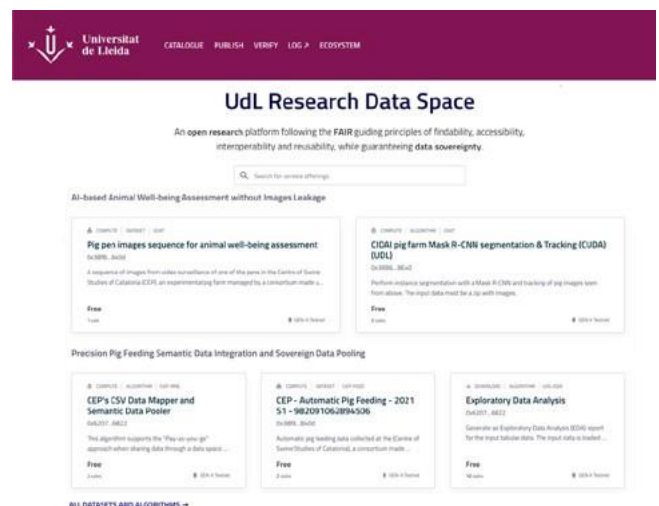
Pontus-x (Gaia-X, n.d.) serves as an excellent example of a data space that facilitates the offering and usage of resources and services. It is a decentralized ecosystem built on Gaia-X, a recent European initiative aimed at creating an open, transparent, and secure data ecosystem in Europe. This space employs the Ocean Protocol to manage interactions with data and services. Due to its transparent and open-source nature, this decentralized technology returns control of data and services to their rightful owners.

This environment leverages Blockchain technology to represent resources and manage user interactions. When we publish a dataset or service within this ecosystem, we automatically generate an NFT (non-fungible token) that uniquely represents our resource and includes all the information needed for secure use. This NFT "lives" on the Blockchain, ensuring decentralized access and giving the owner full control. Additionally, transactions for acquiring and utilizing data and services in this ecosystem are conducted through virtual currencies (tokens), enhancing transparency, security, and integration.

Within the scope of this project, the data space offers several functionalities.

1. Upon accessing the portal of a Pontus-X based data space, such as the one at the University of Lleida shown in Figure 8, we encounter a catalogue of data and algorithms available either for purchase or for free. For datasets, there are two access modes that the owner must select. The first mode, "download," allows the data to be downloaded after purchase. The second mode, "computation," does not permit direct access to the data but allows its usage through an algorithm within the same ecosystem. This ensures that the data never leave the environment, maintaining privacy and control. The owner also has the authority to decide which algorithms can be executed on their data.

Figure 8. UdL data space portal



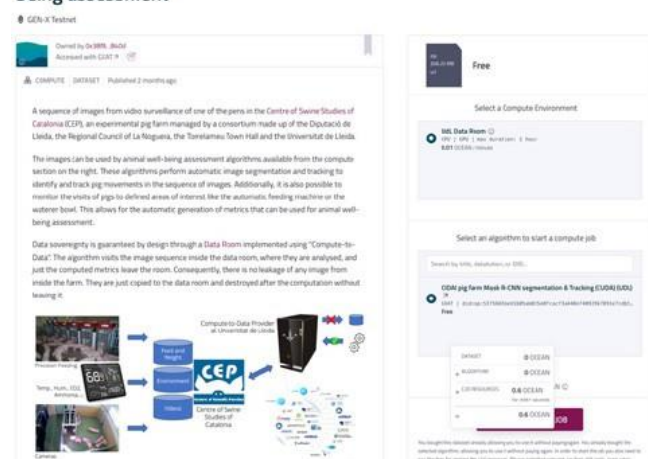
2. When we want to use an algorithm with specific data, we must pay the price associated with the data, the algorithm, and the cost of computation, all of which go to the respective resource owners: The data owner, the algorithm creator, or the computing infrastructure provider. This process is executed within the platform itself, and we can track the status and results in the "computation tasks" section of our profile (as shown in Figure 9).
3. On the other hand, to publish data or algorithms, we only need to fill in the necessary information. This includes selecting the type of resource (data or algorithm), providing basic information like

the title and description, setting a price, and specifying the location of the data or algorithm, among other details. It's important to note that Pontus-X is not a data store; it only manages access to the data. Therefore, we are responsible for uploading the data to a web service, and when we publish them, we must indicate how to access them. We can use access methods such as a single URL to download the data or protocols like IPFS, which allow data storage in an open and decentralized network.

4. When offering data for download, the platform generates a secure link upon purchase that never exposes the original data address, ensuring control remains intact. Additionally, the data has a set validity period from the time of purchase. After this period expires, the buyer will need to repurchase the data to regain access.

Figure 9. Details of the data space

Pig pen images sequence for animal well-being assessment



5. Ultimately, the algorithms are executed using the Docker virtualization environment. This allows us to utilize any Docker image available in private or public image repositories, such as Docker Hub. When publishing the algorithm, we need to specify the order of execution and the code file to be executed. The execution will occur on the machines provided by the environment, although there is also the possibility of using another provider with different hardware specifications.

4.4. Governance, sustainability, and operation analysis

Introduction

One of the major problems with the pilots is that they do not manage to overcome the pilot phase, since they are not consolidated within a sustainable business model. This inability to consolidate within a sustainable business model adversely affects both citizens and producers. They often fail to see the tangible impact of their participation in innovation processes, leading to a loss of trust in the institutions promoting these initiatives. Consequently, this results in diminished legitimacy for these institutions and a decline in citizen interest in open innovation processes.

Throughout the project execution, the BSC (Barcelona Supercomputing Center) has undertaken specific initiatives to promote the technological transfer of various tools and solutions to the real market, evaluating their exploitability and economic impact. The following specific objectives are proposed:

- Examining the intellectual property of the projects and ensuring their protection.
- Crafting comprehensive business plans.
- Technology Readiness Level (TRL) evaluation
- Approaching the proposals to the market and to the institutions that are part of the Catalan Deep Tech ecosystem.

To achieve these objectives, the data obtained during the technical development of the work conducted in the various High Impact Projects of CIDAI will be utilized. Additionally, the expertise in innovation management and intellectual property rights of the Technology Transfer team at the Barcelona Supercomputing Center will play a crucial role.

The methodology used to prepare this report is detailed in ANNEX (A3).

Traceability algorithm and welfare indicators

Intellectual property protection analysis

Pre-existing Intellectual Property (IP) Rights

University of Lleida			
COMPONENT (name of pre-existing technology/know-how/software)	TYPE (is it a methodology, software, hardware?)	OWNERSHIP (who has Intellectual Property rights over this technology?)	CONDITIONS OF USE AND OPERATION (what conditions apply to this technology? Has it been assigned a licence?)
UdL Research Data Space	Software	University of Lleida	Apache License 2.0
Exploratory Data Analysis Data Space Algorithm	Software	University of Lleida	Apache License 2.0
CEP's Mapper for Daily Pig Weight CSV Data Space Algorithm	Software	University of Lleida	Apache License 2.0
CEP's Mapper for Automatic Pig Feeding CSV Data Space Algorithm	Software	University of Lleida	Apache License 2.0

Computer Vision Centre			
COMPONENT (name of pre-existing technology/know-how/software)	TYPE (is it a methodology, software, hardware?)	OWNERSHIP (who has Intellectual Property rights over this technology?)	CONDITIONS OF USE AND OPERATION (what conditions apply to this technology? Has it been assigned a licence?)
AI-based algorithm for tracking cows within an enclosure.	Software	Shared between the CVC and Grupo Alcañiz	Agreement between the CVC and the company for commercial operation by the company

The following table shows the pre-existing IP used by members to implement tasks with third-party ownership (third-party knowledge/software/etc., including open source).

COMPONENT (name of pre-existing technology/know-how/software)	TYPE (is it a methodology, software, hardware?)	OWNERSHIP (who has Intellectual Property rights over this technology?)	CONDITIONS OF USE AND OPERATION (what conditions apply to this technology? Has it been assigned a licence?)
Minimal Viable Gaia-X Portal	Software	deltaDAO	Apache License 2.0
Ocean Marketplace	Software	Ocean Protocol Foundation Ltd.	Apache License 2.0
ydata-profiling	Software	YData Labs Inc.	MIT License
rmlmapper-java	Software	RDF Mapping Language (RML)	MIT License
Mask R-CNN (Detectron2)	Software	Yuxin Wu and Alexander Kirillov and Francisco Massa and Wan-Yen Lo and Ross Girshick	Apache License 2.0
Deep-Person-REID	Software	Zhou, Kaiyang and Xiang, Tao	MIT License

Resulting IP (foreground IP)

COMPONENT (name of pre-existing technology/know-how/software)	TYPE (is it a methodology, software, hardware?)	OWNERSHIP (who has Intellectual Property rights over this technology?)	CONDITIONS OF USE AND OPERATION (what conditions apply to this technology? Has it been assigned a licence?)
Methodology for the deployment of sustainable data spaces and data sovereignty in the agri-food sector.	Methodology	Shared with the rest of the High Impact Project participants	Apache License 2.0
AI-based algorithm for tracking pigs within a pen.	Software	CVC, CEP/UdL	Apache License 2.0

Context: the market

Characteristics of the livestock market

The livestock market is evolving. Today, livestock farms vary greatly in terms of size and level of technology. Some highly advanced farms use computerized control tools, while older farms may lack modern facilities or even mobile coverage. Currently, many farms use management programs like Porc-tec (from Agrotec Soft) for productivity parameters and big data for feed management. Additionally, they occasionally participate in innovation projects.

Large integrators are expanding their influence by acquiring farms, thereby increasing their market power and instigating price wars. They are implementing tools to enhance productivity and manage costs effectively. For example, they use precision feeding to save on feed and drugs, and manure control to reduce tax liabilities.

At the individual farm level, it is often observed that farmers tend to be less inclined toward innovation compared to other agents in the agri-food sector, such as industrial companies that are more accustomed to managing costs and productivity.

Therefore, we can expect a gradual increase in the technification of farms. This progress, however, will be slow due to the complexity and numerous uncontrollable parameters inherent in the livestock sector.

Importance and role of innovation

The importance of monitoring animals and collecting parameters to assess their behaviour and welfare, as well as to measure the production costs of farms / livestock holdings, is growing and will continue to do so. This is crucial, particularly in terms of performance / productivity and optimization, such as monitoring water consumption, feed usage, and more.

This increased level of control is mandated by administrations, which are requiring more rigorous monitoring of gas and waste emissions (such as CO₂, ammonia, etc.), as well as meticulous record-keeping. Consequently, it will become essential to acquire and implement tools to aid in these tasks.

Meanwhile, consumers are placing greater importance on the quality and origin of their meat, often willing to pay a premium for it. As a result, knowing the traceability of the animals has become increasingly crucial. Modern computer tools can significantly aid in providing this information.

Innovation, the use of software, and monitoring are essential for the future. For farmers, it is crucial that innovations provide clear cost-benefit advantages. Adding value to data is key. However, the reality is that farmers often have limited time / resources to analyse the data they already generate. A practical mindset prevails. They require straightforward, useful data that is easy to interpret and can help them make decisions aimed at boosting productivity.

Software

Attitude and general assessment

When it comes to new technological solutions, farmers / cooperatives are showing receptive attitudes. The concept of individualized animal monitoring is recognized as intriguing, but they don't yet see the practical value of the metrics currently available. However, they believe that other important data, potentially provided by this tool, could significantly improve productivity. While everyone acknowledges the importance of enhancing animal welfare, the primary goal remains increasing productivity, not necessarily obtaining certifications. There is some scepticism regarding the tool's approach. Farmers find it challenging to monitor a large number of animals with fixed cameras, considering it a high investment for data they deem uninteresting. They also have other investment priorities, such as cooling systems for sheds and laser technology for silos.

Consultants have a much more positive assessment of the tool, recognizing its significant potential. They are more cognizant of the technological challenges involved in tracking and identifying animals in the pen, and they believe these challenges have been effectively addressed. Moreover, they have a clearer understanding of the various possible parameters or key performance indicators (KPIs) related to animal behaviour and welfare that can be derived from the data collected thus far. These insights are seen as highly valuable and useful for farmers.

All the sectoral agents recognize that the current developments represent the initial and essential stage of a broader process. They understand that the subsequent steps are crucial to demonstrating the tool's overall usefulness and potential.

Assessment of current information

The valuation for each of the currently calculated metrics is shown below:

- **Pig travel** (distance). No initial interest.
- **Time spent in the scale area** (where pigs eat and can be weighed). It is not useful to them. They are interested in data such as: whether they eat or stop eating, the amount of feed they consume, and, most importantly, their weight.

- **Time spent in the water zone** (to estimate water consumption). This data is not of interest to them, as animals might linger there to cool off. Instead, they are interested in the actual water consumption data.
- **Time spent "sleeping"** (or idle time) and heat maps (tracking where the pig spends its time). They are not inherently of interest. However, these data points become valuable if they indicate abnormal behaviours such as passivity, high activity, aggressive behaviour, tail biting, etc., which could be linked to diseases or welfare issues.
Such insights can inform decisions on actions like separating animals, administering medication, adjusting shed temperature or air conditions to reduce stress, and more.

End users struggle to perceive the value or utility of the current measures provided by the tool. They see potential, particularly if the tool can offer more relevant data and/or correlate the existing variables (e.g., inactivity) with more specific and actionable parameters that can guide decisions towards optimization / better organization / increased productivity.

The consultants have a much more favourable view, believing that the current information can be used to derive valuable KPIs. These KPIs could help identify issues such as climate problems, conflicts, and diseases, facilitate decision-making, and provide traceability throughout the animal's life cycle.

Information expectations. More interesting / desirable measurements and applications

The relevant indicators identified include:

- **Weight.** This is the most interesting and motivating metric for using the tool, particularly for fattening and transition/weaning farms. Knowing the weight of the animals at any given moment, and having historical data (growth curves), is highly valuable for numerous applications.
- **Detecting Deviations from the Standard**, e.g. not reaching weight. Determining the underlying causes (such as issues with water, feed, diseases, etc.) and implementing corrective measures.
- **Deciding when to cover animals** (e.g. from 150 kg).
- **Making decisions to change the type of feed**, looking for the most appropriate according to their weight

(feed: starter, growth, finishing feed), they could also be placed in different pens according to their weight and adapt the feed to each pen.

- Forecasting and selection/marketing of animals to go to the slaughterhouse (currently done visually).
- Amount of water they drink and amount of feed they eat. Consumption management and disease detection.
- If they drink water or not / If they eat or not. Detecting febrile states and diseases.
- Passive behaviours: observing changes in hierarchy, reactions to being beaten, and movement inactivity. Detecting febrile states and diseases.
- Hyperactive behaviours: identifying signs of nervousness and stress. Controlling house conditions such as temperature and air currents to manage stress and disease.
- Aggressive behaviours (tail biting). Managing medication to reduce the probability of infections.
- Individualized temperature monitoring: Detecting fever and managing disease.

Along these lines, they are interested in other indicators related to abnormal behaviours and diseases, although they are unsure whether these can be obtained through the tool or how they might be treated. Potential indicators include: visual indicators: images could be used to identify issues like diarrhoea, changes in animal colour (e.g., a more yellow hue might indicate liver problems), neck fat (correlating with weight), body condition before and after calving (for mother farms); behavioural indicators: monitoring moods such as aggressiveness, fighting, changes in pen hierarchy, ear or tail biting, and passivity, and detecting dead animals. Health Indicators: If the tool can incorporate sound, it might detect coughing or breathing issues (indicating whether to vaccinate or medicate), screaming or fighting (indicating stress or disease), and mother-piglet communication.

Additionally, measures and records requested / to be requested by the Administration are of interest. These are typically general data of the pens / sheds rather than individualized data. Starting in 2025, the following will be required:

- CO₂ Levels.
- Ammonia levels.
- Ambient temperature records.
- Light levels per pen (minimum 40 lux at animal head level).

They anticipate that in the future, the Administration may request additional records regarding animal welfare parameters. To date, data medicines and vaccines are reported to the Administration; however, their traceability remains unknown. They also submit information on the Administration's portals regarding droppings, CO₂, and NH₃ levels, but they lack a means of verifying their accuracy. This would enable effective control.

For companies and cooperatives, the ultimate goal of data should always be to enhance the conversion rate of farms and, consequently, their productivity. The enhancement of animal welfare is relevant insofar as it aids in achieving this objective (anticipating diseases / optimizing resources). Ultimately, it is crucial to relate any parameter to the feed conversion rate, as the aim is to expedite the animal's journey to the slaughterhouse while minimizing water and feed consumption. If improvements in animal welfare or reductions in discomfort lead to enhanced feed conversion rates, then such information becomes highly significant and valuable. It is essential to explore how to establish these correlations.

Currently, they believe that tool will not be useful for certifications, welfare, or other purposes, as the data are obtained by inspectors and the farms do not require any additional information. They are also uncertain whether individualized follow-up can be beneficial given that the Administration requests data and records per shed or pen. However, if the tool is able to measure parameters directly related to animal welfare, it may prove useful for future requirements from the Administration.

The consultants consider a broader range of applications, including animal welfare, hygiene, problem detection on farms, certifications, and more. They emphasize the importance of collecting extensive data to facilitate multivariate statistical analysis, aiming to identify significant correlations, even with variables that initially may not seem important.

Behavioural KPIs should be connected to secondary technical variables, with the more variables included, the better. Veterinarians should identify all potential direct and indirect variables, as this adds value to the data. This comprehensive approach enables a thorough analysis, allowing for the identification of all possible correlations.

Expectations about data / interface / operation

In relation to data, they need reliable, simple, and easy-to-interpret data. They prefer fewer, but highly relevant and practical data points that facilitate decision-making. They are interested in the ability to select KPIs and measures from a catalogue and find it very appealing to customize the KPIs / information to be contracted, with pricing based on the data selected.

Regarding the interface / outputs / and data processing, it's recommended to make them highly visual, offering evolutionary information that allows for historical tracking and comparisons between different parameters / farms. Once the raw data is acquired, it can be integrated with business variables (e.g., through the power value program), and KPIs / data can be downloaded to create control tables, graphs, and production follow-up reports. Additionally, the same application could be used to monitor certain variables (e.g., pen losses) and include a space for data entry by the farms.

Pipeline / process assessment

Installation of sensors / cameras

The more open and positive sector agents do not foresee significant issues with installing these devices. However, they mention certain aspects that need to be considered (which can be seen as barriers):

- If the buildings or pens are made of uralite, drilling holes in them is prohibited by law.
- The cameras must be protected or designed to withstand pressure cleaning, disinfection, and exposure to gases in the buildings. This implies that, in addition to the initial investment, there may also be ongoing equipment maintenance costs.
- There are farms located in remote locations with no internet coverage.
- The most sceptical individuals view the investment as very expensive, particularly for large facilities. They find it impractical to have fixed cameras and suggest that an application providing information from photographs or sounds on a cell phone would be more effective.

In both scenarios, the increase in productivity and optimization resulting from data-driven decisions must offset the costs and investments required. The benefits need to be clear and substantial to motivate farmers, farms, or cooperatives to invest.

For end users, the optimal investment-return ratio is crucial in deciding whether to participate in this initiative. The consultants believe this potential barrier (investment cost) can be addressed through creative solutions, such as installing cameras in select pens rather than the entire facility and offering servitization models to assist farmers in decision-making. For example, renting cameras for the first fattening period with the option to purchase once the solution's effectiveness has been proven.

Obtaining data / timing

Generally, the fact that it takes a month to adjust the model and start obtaining data is not a concern for farmers, as long as the data they receive is reliable and useful.

However, a key question arises: will the model need a month to gather data for each fattening cycle, or only for the initial setup? If it's required for each cycle, the tool loses its appeal, since monitoring parameters at the beginning of the fattening period is crucial.

Those with more technical knowledge understand that the larger the volume of data integrated into the software, the faster the model will adjust. This potential barrier can be overcome if more users join the project.



Data space

End-user assessment and expectations

Initial barriers

Initially, farmers and cooperatives are hesitant to share their farm data. They would be willing to do so if the data is anonymized and presented in an aggregate form on the server, as they do not want competitors to access their productivity data. The data protection policy is very transparent about how the data is used.

Level of interest / usefulness

Those who find the tool's information useful also see the value in accessing a server or cloud with aggregated data from farms using the software. This would allow for comparative analysis of their results against farms in the area or province, different types of farms, farms with similar feed types, and more, with data such as costs, number of piglets per sow, conversion rates, and defined KPIs.

Additionally, they consider the data space valuable for disease prevention. However, they note that farms are often reluctant to share data on issues like viruses or health problems, despite recognizing that sharing this information can enhance the algorithm's accuracy and stability.

Conversely, those who are sceptical about the tool do not see the value in accessing images or data from the data space.

Operation

Functioning with a reasonably priced subscription is positively valued.

Consultant assessment and expectations

General attitude

Stakeholders appreciate the approach and believe it has significant potential. The methodology incorporates democratic criteria, based on a token economy.

Sovereignty and data sharing

This is a crucial issue to define, as data sharing is perceived as the fundamental barrier. While it's not easy to resolve, creating a reliable environment is essential to maintain control over the data. It's imperative that data is always provided in an aggregated form.

Federated Learning, which allows the use of third-party data in algorithms without needing to physically store them, is seen as a promising solution. This method ensures that data are not exported, downloaded, or copied. Despite its complexity, stakeholders suggest offering a guarantee or certification seal by a trusted company to alleviate farmers' concerns. Consequently, a data governance policy must be established, covering consents, voting, data management delegation, and negotiation, in line with European regulations. However, they believe that end users may be reluctant to share internal data (e.g., productivity) but will likely share other data that can benefit the sector (e.g., water consumption reduction, anomalous behaviour detection).

Ecosystem

They recognize the importance of an ecosystem that adds value to the solution and prevents it from being easily copied.

This ecosystem could function as a marketplace where data producers, analysts, interpreters, AI or software developers, investors, and others can collaborate. Such a marketplace would benefit all parties involved, including both service providers and end users.

Incentive system

It's important to create incentive models that motivate and do not generate distrust. Direct payment for data may increase distrust.

However, with the support of public administration, cooperation could be more likely as it would be perceived as benefiting the entire sector, even if the service is marketed privately. The administration could also assist in developing initial models.

Another option is a mixed model: farmers provide data to help advance the sector and, in return, receive a discount on the service, encouraging their participation.

Potential users

Farmers

Farmers may be interested in the solution to reduce day-to-day management time, make informed decisions, gain peace of mind, improve productivity, and cut costs (although many expenses are covered by the integrator).

They seek a tool that is simple, practical, and cost-effective. They prefer minimal, reliable, and useful data that can drive decisions. A few alarms might suffice, such as alerts three days before a potential animal illness; or notifications when a pig shows promising growth and quality. They lack the time and resources to analyse extensive data and often have limited staff or training. Therefore, it's crucial to introduce the solution when the technology is mature / robust / and provides tangible value, even if initially small. Farmers are wary of adopting new technologies that have previously failed due to immaturity (e.g. drones).

Farms with more potential

The more technologically advanced farms will be able to better utilize these solutions. The trend towards increased technification is expected to continue, as consumers value and demand traceability of animals. These solutions will also be beneficial for larger farms, where tracking many animals and setting alarms or making small improvements in KPIs can significantly impact final productivity.

Currently, the most technologically advanced farms include: pig fattening farms, lactation farms, mother and great-grandmother farms, as well as poultry and beef farms.

Integrators

Integrators theoretically have significant potential as key agents. They are accustomed to managing and analysing data, comparing their farms to identify which ones perform well and which do not. However, they typically lack insight into the inner workings of each farm. This solution can be very beneficial, helping them understand the internal operations, identify problems or reasons for varying results, and implement optimization and improvement measures. Additionally, integrators often have available resources. Despite this, as profit-driven entities, they may be reluctant to invest additional resources in productivity improvements or may have other priorities. Similar to farmers, they tend to prefer waiting until the tool is more mature.

Cooperatives

Cooperatives are also interested in helping their farmers by offering this service to members, provided they see a direct return.



Final assessment

End users are open to learning more about the tool. They find it interesting and appreciate the ability to continuously measure animal behaviour and welfare. However, these measures must be carefully selected to ensure the tool is economically profitable, leading to improved KPIs, cost reduction, space optimization, and better conversion rates, etc. The improvements need to be clearly measurable.

It's important to note that animal husbandry involves many interrelated variables, including the human factor, making it challenging to isolate parameters that facilitate decision-making.

The consultants are confident in the medium-term potential of the software and data space, identifying numerous benefits. They believe that in a second phase, interrelationships between variables will be discovered, leading to savings and increased productivity for farms. However, they acknowledge that the journey will be long and that various barriers will need to be overcome. Moreover, they emphasize the importance of understanding farmers and presenting the tool when it is more mature, with compelling arguments that motivate them—namely, minimal yet useful data for management. They propose developing creative marketing models throughout the process to break down barriers for farmers. These could include servitization models for camera installation, incentive systems to encourage data sharing, and leveraging support and sector-specific aids from the Administration to develop and validate the models, as outlined in the next steps shown in Figure 10 and described below.

Figure 10. Diagram of next-steps



- **Current Stage / Technological Stage.** Achieve the technology to obtain the first level animal behaviour KPIs, from the individualized follow-up of each identified pig.
- **Stage 2 / Technical Stage.** Exhaustively list all potential variables related to animal welfare, abnormal behaviours, sanitary conditions, conversion rates, and productivity (direct or indirect) that can help identify correlations with first-level KPIs and segment the models as needed (e.g., by growth stages). The more variables, the better (e.g., antibiotic consumption, feed, movements, inactivity, etc.). The less obvious, the better, to prevent easy replication of the solution. Veterinarians can provide more technical data than farmers, adding significant value to the models. Farmers can also contribute.
- **Stage 3 / Data Collection Stage.** Obtain data from as many farms as possible (to strengthen the model) / validate the data collected.
- **Stage 4 / Creation and Validation Stage.** Conduct a multivariate statistical analysis to identify all significant correlations between variables. Build segmented mathematical models (e.g., for mothers, reproduction, fattening, and fattening duration) that enable the extraction of summarized, useful, and directly relevant KPIs for farmers.
- **Stage 5 / Marketing Design Stage.** Design effective sales models for farmers / integrators / and cooperatives, tailoring our approach to their needs and thinking of formulas that motivate the contracting of both the software and the DataSpace (incentive systems).
- **Stage 6 / Technological + Legal Stage.** Parallel stage to the previous one (not correlative). Formulate the technological approach of the DataSpace. Establish data protection policies, consents, etc.
- **Stage 7 / Marketing Stage.** Obtain clients for the tool / contracting (farmers, integrators and cooperatives).
- **Stage 8 / Operation Stage.** Obtaining the KPIs of each client farm (software). With the data of those who agree to share them, feed the databases to strengthen / validate / sophisticate the models.

Data governance model

Establishing a governance model is crucial for this project in the swine industry, particularly where data spaces are utilized, for several key reasons. First, it ensures that data is collected, stored, and shared ethically and responsibly, protecting the privacy of pig farms and other stakeholders. Additionally, it helps ensure compliance with privacy and data protection regulations, thereby avoiding potential legal sanctions. Clear guidelines for data management are also established, promoting transparency, building stakeholder trust, and facilitating ethical and responsible research. In short, the governance model is vital for maintaining integrity, compliance, and the continued success of the project in this sector.

Organizational structure

Establishing a clear and stable organizational structure in this swine project, utilizing data spaces, is critical for several reasons. First, it helps define roles, responsibilities, and hierarchies, ensuring that each team member understands their functions and how they integrate with others. This promotes efficient work distribution and effective communication, preventing confusion and duplication of effort. It also facilitates decision-making and problem-solving, as responsibilities are clearly assigned. Additionally, it provides long-term stability and cohesion, ensuring continuity of the project.

To this end, it will be essential to establish:

- **Data Manager.** Appoint a data manager, preferably a Chief Data Officer (CDO), who will oversee the data governance of the project.
- **Data Governance Committee.** Set up a committee that includes representatives from the university, scientists, ethics and privacy experts, and other relevant stakeholders.
- **Roles and Responsibilities.** Define clear roles and responsibilities for data managers, including scientists, technicians and field staff.

Policies and procedures

Data Collection Policy

Establishing a data collection policy for this swine sector project is essential to ensure consistency, privacy protection, and regulatory compliance in the use of data spaces. This policy ensures that data are collected and shared ethically and effectively, safeguarding stakeholder privacy and maintaining the integrity of the research.

Installation of cameras:

- Cameras will be installed on pig farms to monitor animal activity.
- The placement of the cameras will be coordinated with the farm owners to ensure they respect the privacy and safety of the animals.

Informed consent:

- Pig farms must provide informed consent for the collection of images and data related to the animals.
- This consent should be obtained in writing and maintained as a record.

Access and Use Policy

Establishing an access and use policy is essential for this swine project to regulate data use, ensure security and privacy, and comply with regulations. This policy helps prevent abuse and ensures the ethical use of data, setting clear guidelines for users and protecting the interests of all parties involved, including the pig farms.

Access Authorization:

- The collected images will be the property of both the pig farms and the University of Lleida. They can only be used by the interested parties in accordance with the data space agreements.

Purpose of Use:

- The data may only be used for the specific purposes of monitoring and recording animals, as well as improving artificial intelligence models for tracking pigs.

Data Sharing Restriction:

- The data cannot be shared outside the data space without the explicit authorization of the pig farms and the University of Lleida.

Data retention

The data retention guidelines in this project ensure efficient and secure data management, minimize the risk of leaks, and comply with regulations. Additionally, these guidelines help avoid unnecessary data retention, thereby protecting the privacy of data subjects.

Data Retention:

- The original images and related annotations will be retained for a specific period, as agreed upon with the farms and interested parties.
- After this period, the data will be securely deleted, ensuring they are not recoverable.

Security Policy

The security policy agreement is crucial for ensuring the continuity and protection of the swine project. It safeguards data, prevents losses, and builds stakeholder confidence. This is essential for maintaining the integrity and long-term success of the project.

Data Security:

- Data will be protected against unauthorized access, misuse, and loss.
- Security measures will include user authentication, encryption, and access control.

Camera Security:

- Cameras will be protected to prevent tampering or unauthorized access.
- Procedures will be established for the regular maintenance of the cameras to ensure they are operational and collecting data correctly.

Protection of Anonymized Data:

- Anonymized data will be protected to avoid identification of individual animals or pig farms.

Data quality

The data quality guidelines ensure that the swine project maintains data integrity and reliability, which are critical to its continuity and success.

Data Standardization:

- Standards will need to be defined for data collection, including video resolution, file format and camera calibration.

Data Validation:

- Implement validation processes to ensure that the data collected is accurate and reliable.

Data Annotation:

- Images should be annotated using the Supervisely environment or similar tools, following established annotation practices.

Visual Validation:

- The annotations should be visually validated by a team of experts to ensure they accurately reflect the reality observed in the images.
- The accuracy of the annotations will be checked for the exact position of the animals and the segmentation masks.

Consistency Check:

- The data should undergo a consistency check to ensure that annotations are uniform throughout the dataset and that labels are applied consistently.

Model Reliability Tests:

- The annotated data will be used to test and validate the reliability of the Machine Learning model (Mask R-CNN).
- The model's performance will be evaluated based on its ability to accurately detect and track pigs using the annotations.

Feedback and Correction:

- If errors or inaccuracies are detected in the annotations during validation, the relevant annotations will be corrected and updated.

Privacy and Ethics - Privacy Impact Assessment (PIA)

A Privacy Impact Assessment (PIA) is essential for this project to protect data privacy, ensure legal compliance, gain stakeholder trust, minimize risks, and ensure ethical research. By identifying and mitigating risks, the PIA helps safeguard the privacy of swine farms and all individuals involved, ensuring compliance with applicable privacy laws.

Risk identification

Pig Farm Identification Risk:

- There is a risk that the collected and shared data could be used to identify pig farms and their locations, potentially posing a security risk to the owners.

Risk of Exposure of Sensitive Data:

- Tracking data may contain sensitive information, such as husbandry practices or animal health, which could expose confidential details about pig farms.

Risk of Reidentification of Individuals:

- If images and data are not properly anonymized, there is a risk that individuals could be identified in the context of the images, which would affect their privacy.

Impact assessment

Impact on the Privacy of Pig Farms:

- The risk of identification could compromise the security and privacy of pig farms, potentially leading to resistance to data sharing.

Impact on People's Privacy:

- Inadequate anonymization could expose individuals to re-identification, compromising their privacy.

Risk mitigation

Data Anonymization:

- Ensure that images and data are effectively anonymized before being shared in the data space, so that pig farms and individuals cannot be identified.

Limitation of Sensitive Data:

- Avoid collecting or sharing sensitive data that is not necessary for the project's objectives.

Informed Consent:

- Continue to obtain informed consent from swine farms for data collection and use, ensuring they understand the risks and benefits.

Education and Awareness:

- Provide information and training to all stakeholders on the importance of data privacy and secure practices.

Ongoing Supervision:

- Implement a process for continuous data privacy monitoring, regularly reviewing and updating privacy measures as needed.



Measurement and monitoring

Establishing performance metrics and conducting regular audits are similarly essential for assessing effectiveness, ensuring regulatory compliance, improving practices, and protecting data privacy in such a project. These measures provide quality control, demonstrate regulatory compliance, and ensure transparency, which is critical to ethical research and stakeholder confidence.

Performance metrics

Tracking Model Accuracy:

The accuracy of the AI model (Mask R-CNN) in detecting and tracking animals in the images.

Model Performance:

- Processing speed and response time of the model during practical use.

Legal Compliance:

- Degree of compliance with data privacy and animal protection regulations.

Stakeholder Satisfaction Index:

- Assessment of stakeholders (farms, annotation entities, etc.) on the management of their data and their degree of satisfaction with the project.

Policy Compliance:

- Measure of adherence to established data governance policies.

Improvements Implemented:

- Number and impact of improvements made in response to audits and privacy assessments.

Number of Active Users:

- The number of users regularly accessing the data space to use data and resources.

User Growth Rate:

- The increase in the number of users over time, indicating the attractiveness of the data space.

Frequency of Use:

- How often users access and use the data available in the space.

Retention Rate:

- The number of users who continue to use the data space after their first access.

Data Downloads:

- The number of times data has been downloaded from space.

User Feedback:

- User feedback and rating of the ease of use and usefulness of the data.

Developed Projects or Applications:

- Number of projects or applications that have been created using the space data.

Impact on Scientific Research or Practical Applications:

- Indicators of the results and contributions of the use of data in scientific research or practical applications.

Regular Audits

Regular audits are crucial for ensuring compliance, identifying problems, and improving security and privacy. It is recommended to perform them at least once a year to maintain the system's integrity.



Communication

Establishing staff awareness and training, along with external communication protocols, is critical in a project like this for several reasons. Firstly, it ensures that everyone involved understands and complies with privacy policies and data protection regulations. Additionally, it fosters awareness of the ethics and importance of privacy protection. Lastly, effective external communication helps build trust with stakeholders and the public, demonstrating a commitment to privacy and ethical responsibility in scientific research. This approach helps avoid potential legal issues, protects the project's image, and maintains integrity and transparency in data management.

Awareness and training

To ensure continuity over time, the project should provide comprehensive training to staff and stakeholders to ensure understanding and adherence to data governance policies and procedures. This includes regular training sessions, training materials, and online resources for continuous learning and proper implementation of these policies and procedures.

External communication

For continuity beyond the High Impact Project period, it's crucial to inform external stakeholders, such as farms, through clear and transparent communication. Distributing explanatory documentation detailing how data will be collected, used, and protected to ensure privacy is recommended. Additionally, holding educational sessions and briefings to address specific questions and concerns can be useful. Establishing open communication channels for stakeholders to raise concerns or seek clarifications is also important. This approach ensures that farms are well-informed and confident in the management of their data within the project.

Feedback and Improvement:

Continuous improvement protocols ensure that the swine project adapts to changes, improves, and maintains its effectiveness over time, ensuring continuity and success. To this end, the following steps are recommended:

1. **Gathering Feedback.** Collect feedback from stakeholders through surveys, meetings and comments.
2. **Assessing Feedback.** Analyse feedback to identify trends and areas for improvement.
3. **Updating Policies and Procedures.** Adjust policies and procedures according to feedback and changing needs.
4. **Communication and Training.** Communicate changes and provide training on updates to team members and others involved.
5. **Implementation and Monitoring.** Ensure new policies are implemented and track their impact.
6. **Ongoing Assessment.** Maintain feedback as an ongoing practice, with periodic collection and adjustment.



4.5. Conclusions

The developed project leads to the following conclusions:

1. The AI LAB model emerges as an effective methodological tool for addressing territorial challenges, efficiently integrating the ecosystem into the AI injection process within the specific sectoral activity.
2. The technological solution for monitoring pig activity in the barn introduces a new scenario for the automatic and individualized traceability of fattening factors for each animal. This enhances the quality of the processes and provides a potential benefit for opening market lines where the traceability of animal welfare during the fattening process can be objectively validated.
3. The pilot development model not only requires the specific implementation of a technological solution but also a mechanism to create value from the generated data. The data space within the agricultural environment provides the necessary infrastructure to make this possible.
4. The governance, sustainability, and business model offer a preliminary assessment of the potential value of the technology used.

The operation of the digital values generated in this project, including data and algorithms, opens opportunities for the sector to benefit from the integration of AI within farms, enhancing the quality of the fattening processes. This model exemplifies the "from bits to kilos" approach that Catalonia's ecosystem can leverage through CIDAI and various innovation hubs of the High Impact Project.



Annex

Content



A1. Details of the experiment – Theoretical basis

Data collection

First, it is necessary to develop an understanding of the application domain and the most relevant prior knowledge by identifying the objective of the KDD process from the customer's point of view. Key points of this phase include: Identifying the relevant data sources for analysis; Designing the data collection process; Establishing appropriate methods and techniques for efficient data acquisition; Selecting representative samples of the population to ensure generalizable results; Collecting data and documenting the procedures.

Finally, ethical and legal aspects must be considered, ensuring privacy, obtaining informed consent, and complying with data protection regulations and policies.

Preprocessing: Selection, cleaning, and transformation

This stage of the KDD process, known as data preprocessing, data cleaning, or data preparation, is the most relevant and time-consuming part. It focuses on preparing the data for further analysis. During this process, basic descriptive statistical techniques and graphical representations are used to understand and visualize the data. Sometimes, multivariate descriptive techniques, such as Principal Component Analysis (PCA), are also applied to detect outliers or nonlinear patterns.

Visualization techniques play a crucial role in identifying important issues during data preprocessing. These techniques may vary in sophistication, but all are useful in ensuring the quality and adequacy of the data prior to analysis.

Although well-established preprocessing methodologies have not been formalized, some guidelines can be provided. For this section of the methodology, the preprocessing process will be simplified into three points: data selection, cleaning, and transformation, guided by the methodology proposed in Gibert K (2016) and illustrated in Figure 1.

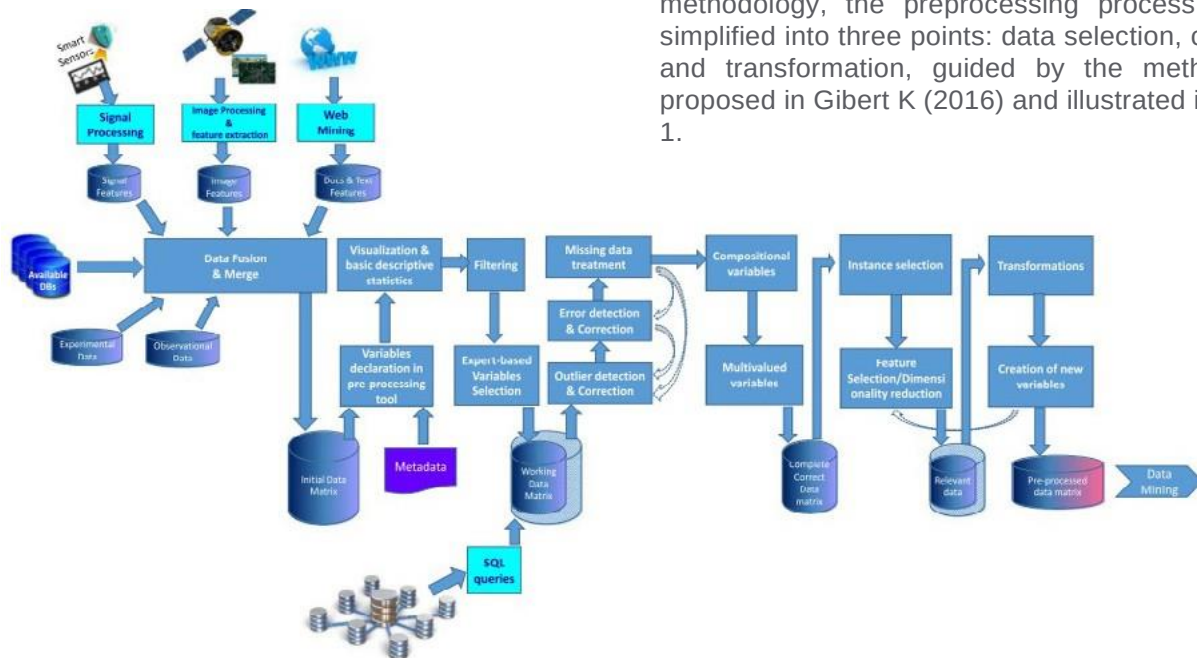


Figure 1. Preprocessing process (Gibert K, 2016)

Data selection

Data selection involves identifying and collecting the relevant data sets to be used in the knowledge discovery process. This stage requires a clear understanding of the project objectives and requirements, as well as the available data sources. Some key aspects of this stage include:

1. **Defining objectives:** Before starting data selection, it is essential to clearly understand the project objectives and requirements. This involves defining the type of knowledge you want to extract from the data and the questions you want to answer. Establishing clear objectives helps focus data selection on information relevant to achieving these goals.
2. **Identifying data sources:** At this stage, it is necessary to identify the available data sources that could contain the information needed for the analysis. These data sources may include databases, CSV files, transaction logs, data collected from sensors, online data, among others. It is important to assess the quality and integrity of these data sources to ensure they are reliable and suitable for analysis.
3. **Assessing relevance:** Once the data sources have been identified, it's essential to evaluate their relevance in relation to the project objectives. This involves examining the variables and attributes in each data set to determine if they provide the necessary information to answer the posed questions. Some data sources may need to be discarded if they are not relevant, while others might require transformations or combinations with other sources to enhance their usefulness.
4. **Size and representativeness of the data:** At this stage, it's essential to consider the size and representativeness of the data sets. The appropriate data size will depend on the complexity of the analysis and the resources available. Ensuring that the selected data sets are representative of the population or phenomenon being studied is crucial to avoid bias or incorrect conclusions.
5. **Ethical and legal considerations:** When selecting data, it is crucial to consider ethical and legal aspects. This includes ensuring data privacy and confidentiality, as well as complying with applicable data protection regulations and policies. In cases involving sensitive data or personal information, obtaining informed consent from the individuals whose data will be used may be necessary.



Data cleaning

Once the data have been selected, data cleaning is performed. This critical step in the data preparation process aims to identify and correct problems or inconsistencies in the data sets. Key aspects of data cleaning include:

1. **Missing data.** During data collection, it is common to encounter missing values. These can arise from various causes, such as data entry errors, transmission problems, or simply missing information. At this stage, it is crucial to identify the missing data and decide how to handle it. This may involve removing rows or columns with missing data, imputing values using techniques such as averaging or interpolation, or employing more advanced methods like machine learning to predict missing values. In cases of structural missing data, experts can provide rules to manually impute them (knowledge-based logical rules) (Gibert K, 2016).
2. **Outliers** Outliers are observations that differ significantly from the rest of the data. These values may result from measurement errors, incorrect data entry, or even indicate exceptional or anomalous situations. It is important to identify outliers and determine whether they should be corrected or eliminated. This can be done using statistical methods, such as range limits or criteria based on standard deviation, or by using more sophisticated algorithms, such as outlier analysis.
3. **Inconsistencies and errors.** Data sets often contain errors or inconsistencies, such as incorrect or contradictory values. For example, a field recording a person's age may contain negative values or values greater than 100 years, which is unlikely. At this stage, it is necessary to identify and correct such errors. This can be done either by eliminating observations with serious problems or by using specific techniques to correct known errors.
4. **Duplicate data.** Data sets may contain duplicates, meaning there are observations that are identical or very similar to other entries in the set. The presence of duplicate data can bias the results of the analysis and affect the accuracy of the models. Therefore, it is important to identify and eliminate duplicate data to avoid unnecessary biases in the analysis.

Data transformation

Data transformation involves applying operations or techniques to convert data into a form suitable for subsequent analysis. This may include normalization of values, discretization of continuous variables, encoding of categorical variables, dimensionality reduction, or any other transformation necessary to adapt the data to the specific algorithm or method to be used in the next stage of the knowledge discovery process.



Data mining

Data mining is the stage where algorithms are used to analyse and discover patterns in data. These algorithms efficiently search for specific patterns or models within a vast pattern space, which is often infinite. However, due to computational limitations, constraints are imposed to limit the search space and the subset of patterns that can be explored. These constraints are necessary to ensure a computationally efficient and feasible process. In this research, a selection of the data mining model to be applied to the case study was made. The literature mentions a varied typology of models, and their selection depends on the nature of the data, the objectives of the study, and the intended use of the results according to the selected model. The authors Gibert et al. (2018) provide a conceptual map that supports the proper selection of methods in data mining. Here, techniques where a variable response is not involved, and the relationships between variables (PCA) and individuals (clustering) are studied.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of data while retaining relevant information. It is part of a set of techniques known as unsupervised learning, which are not based on a predictive model. Instead, PCA seeks to extract information to identify patterns, such as subgroups within the data.

The PCA method operates algebraically with the *eigenvectors* and *eigenvalues* of a numerical data matrix, using variance maximization criteria to express its principal components. This allows the observation of the inertia or explained variance of the dimensions, indicating whether strong relationships exist between the variables in these planes and suggesting the number of axes to be studied. Typically, components that explain at least 80% of the total variance—thereby capturing the real information—are selected.

The PCA process identifies the directions in which the variance is greatest. Since the variance of a variable is measured on the same squared scale, variables with a larger scale will dominate the analysis if all variables are not pre-standardized to have a mean of 0 and a standard deviation of 1. Therefore, it is essential to standardize the data before calculating the principal components, which also helps correct for possible outliers, as they can directly influence the variance of the variables.

Clustering

Cluster analysis, or clustering, is a statistical method used in data processing to organize all elements of a database into groups, or clusters. These clusters are formed based on how closely the elements are associated and the characteristics they share. The goal of cluster analysis is to identify similar groups of subjects, where the "similarity" between each pair of subjects is defined by sharing certain traits across the entire set of characteristics.

The method used for clustering is Ward's hierarchical bottom-up method. This approach aims to find compact spherical groups that minimize the total within-group variance. At each stage, pairs of groups with the smallest distance between them are merged, and all possible pairs of groups are evaluated. The groups whose merger results in the smallest increase in information loss are then combined.

Following this methodology, dendrograms—tree diagrams showing the clusters of observations and levels of similarity—are created. The optimal number of classes is determined by observing the resulting dendrogram and considering the characteristics of the data and the objectives of the analysis. An expert can determine the optimal number of classes by examining the tree diagram or using other algorithms, such as the elbow method. It should be noted that the greater the distance between two links in the dendrogram, the greater the difference in terms of characteristics.

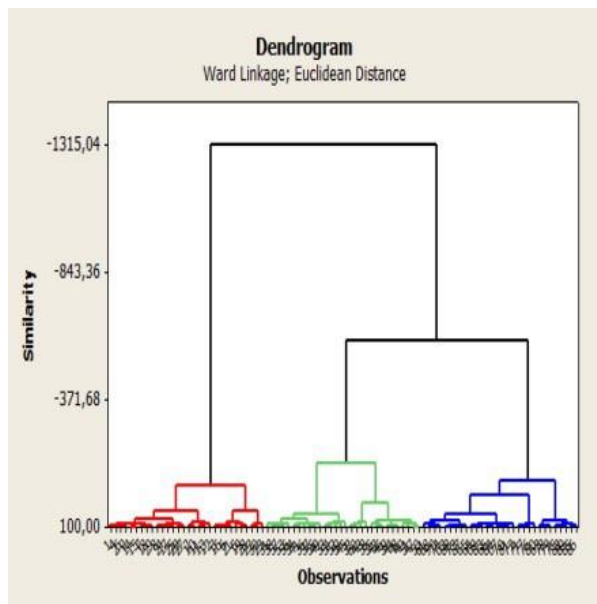


Figure 2. Dendrogram of Iris Flower data from R (<https://www-eio.upc.edu/teaching/DocenciaMultivariant/>)

To perform cluster analysis on mixed data (numerical and categorical), it is necessary to establish a suitable metric, for which there are several algorithms. In this study, Gower's metric and Euclidean distance, which measure the distance between observations, were used.

Interpretation and assessment

The profiling of the classes (clusters) is used to describe the conditional distributions of these classes, identify their specificities, and interpret them based on the variables included in the clustering process. Some of the most commonly used tools for this purpose include:

- **Calculation of centroids.** The calculation of centroids generally involves the mean and the mean of aggregate data, conditioned by each class, for numerical variables. In the case of categorical variables, centroids are determined based on the mode or through tables that describe the composition of the categories conditional on each class.
- **Descriptive statistics tools.** Multiple boxplots, bicharts, pie charts and multiple barcharts to illustrate the composition of each class-conditional variable.

Validation

Validation refers to the evaluation and verification of the results obtained during the discovery and modelling stages. It is an essential step to ensure the discovered patterns, models, or rules are reliable, accurate, and useful in real-world contexts.

Validation entails the application of suitable metrics and techniques to address the quality and performance of KDD results. This process may involve partitioning the dataset into training and testing subsets, comparing model predictions with actual data, calculating measures of precision and accuracy, and conducting tests to ascertain the significance of the results.

Statistical tests of the relationship between variables:

- **Kruskal Wallis test.** The Kruskal-Wallis test is a non-parametric statistical methodology utilized to assess whether significant differences exist between groups concerning a continuous variable. The data are organized and ranked, with the sum of the ranks computed for each group, resulting in a test statistic. If this test statistic is found to be statistically significant, it can be concluded that there are notable differences between the groups regarding the measured variable.
- **Chi-square test.** The Chi-square test is a statistical method employed to determine whether a significant association exists between two categorical variables. This test is grounded in comparing the observed frequency of each category with the expected frequency, under the assumption of independence between the variables. First, a chi-squared test statistic is calculated and compared to the theoretical chi-squared distribution. Secondly, if the p-value associated with the test statistic is less than the predefined significance level, the null hypothesis is rejected, leading to the conclusion that there is a significant association between the variables.

A2. Experimentation Details – Application to Use Case 1

Data collection

To evaluate the potential influence of various factors on antibiotic consumption, while considering production orientation, the following information was necessary:

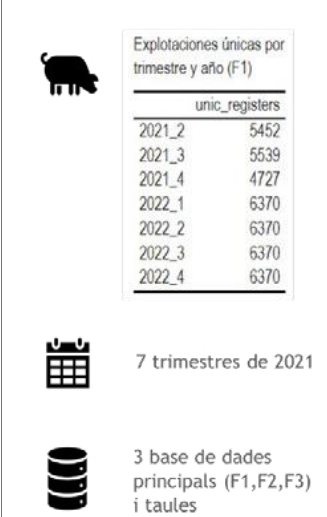
1. Farm size (average farm occupancy)
2. Date of farm (date of origin or date of last productive orientation)
3. If the farm is part of an integrating company
4. Number of animal origins
5. Distance from the origin of the animals to the destination farm
6. Age of the animals at entry
7. Staggered inputs (all-in/all-out system)
8. Biosafety survey result

Based on consumption data categorized by farm and species, the influence of various factors associated with the prescribing veterinarian on antibiotic consumption was evaluated using data from:

- Total antibiotic consumption and critical antibiotics / farm/ veterinary /year
- The number of prescriptions, the number of farms, and the distribution of farms by veterinarian
- Number of prescribing veterinarians per farm
- Identification of repetitive treatments by type of farm and veterinarian
- Types of antibiotic used

Data on antibiotic consumption is available categorized by farm, species, and veterinarian prescriptions for the year 2021. These data are provided by the Service of Animal Feeding and Livestock Production Security of the Government of Catalonia through the MAPA (Ministry of Agriculture, Fisheries and Food). Furthermore, the prescribing veterinarian and the classification of the antibiotic (B, C, and D) are accessible.

Figure 3. Database summary. Data source: CIDAI (6 July 2023)



	unic_registers
2021_2	5452
2021_3	5539
2021_4	4727
2022_1	6370
2022_2	6370
2022_3	6370
2022_4	6370

The original database has a total of 25 tables. These are divided into main tables and auxiliary tables — Excel type (xlsx):

- F1, F2, and F3 are the primary files that are each subdivided into seven additional tables, organized year and quarter.
- There are four auxiliary tables containing information, including location, results of the biosecurity survey, animal mortality and other pertinent details.

Table 1. Description of the original files

Types of tables	Filename	Description
Main	F1	Summary of F2 and F3 information. The first columns are common to all 3 files and serve to list the data.
	F2	It contains information on the weight and UR of the farms.
	F3	Contains dosage information for each prescription and veterinarian information.
Auxiliary	Staggered	Additional information on farms
	Collection of carcasses	Livestock carcass data by farm for the first quarter of the year 2022
	Biosecurity surveys	Results of the biosecurity survey in some quarters in F1
	Farm records	Geographical information and legal data of the farms

For the data analysis the seven tables F1, F2, and F3 were initially merged a single table, resulting in one table containing quarterly data for 2021 and 2022, respectively.

Preprocessing and transformation

Selection, cleaning, and transformation

Before commencing the analysis of the missing values and description of the data, the consolidation all data from F1, F2, and F3 was completed. For this, Structured Query Language (SQL) was used in the Rstudio software with the sqldf() function of the sqldf package (RStudio_Team, 2020). Consequently the total of 21 subtables, categorized by year quarter, will be condensed into three.

Data quality

The study of missing values in database is continued. The following figures illustrate the percentage of missing data for each variable across the various database files.

Figure 4. Graph of the missing values of file F1



Figure 5. Graph of the missing values of file F2

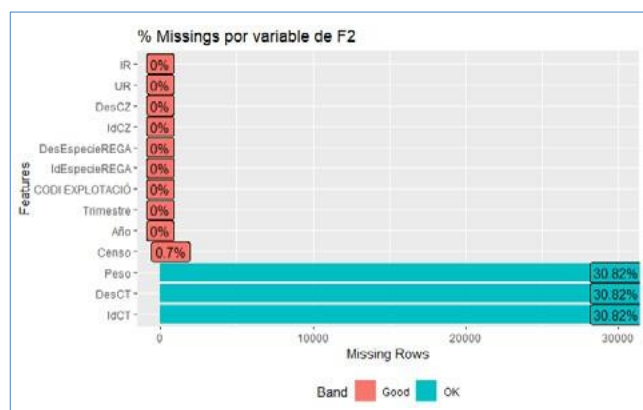


Figure 6. Graph of the missing values of file F3

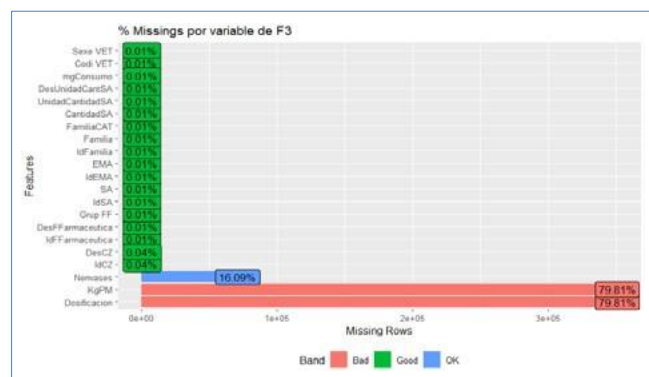


Figure 7. Graph of the missing values of file Staggered

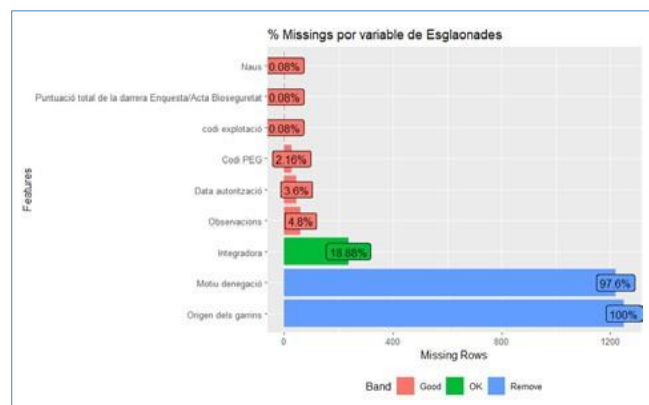


Figure 8. Graph of the missing values of file Record of farms

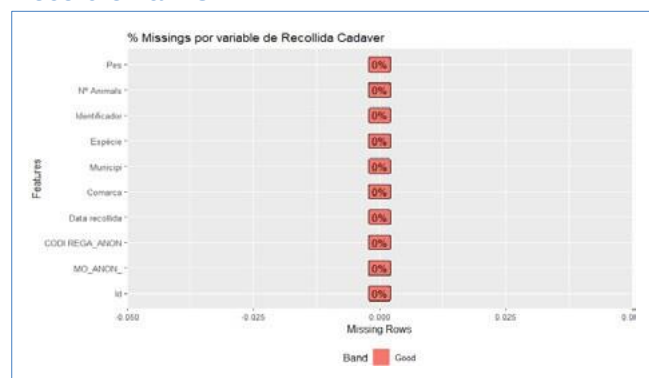


Figure 9. Graph of the missing values of file Carcass collection

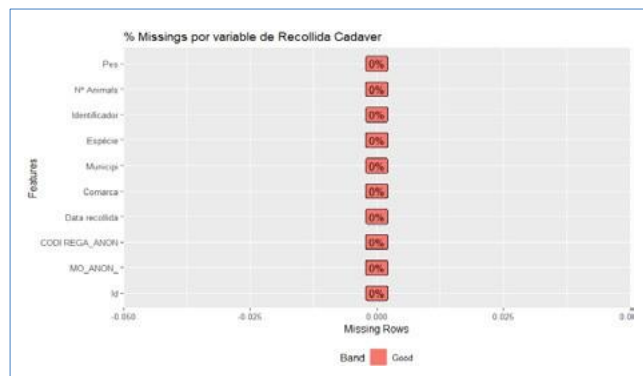
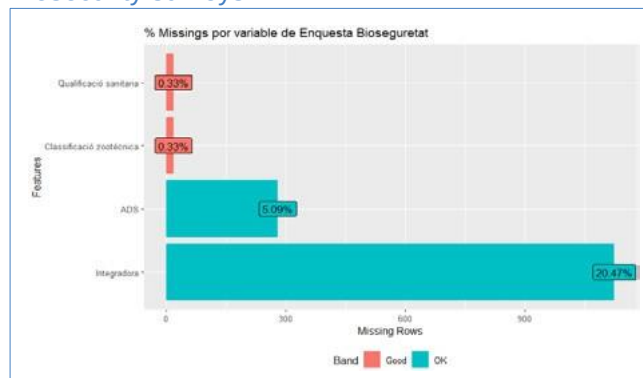


Figure 10. Graph of the missing values of file Biosecurity surveys



The graphs indicate certain variables exhibit minimal or no missing data, whereas others display a substantial percentage of missing data. The latter variables were excluded from analysis and not utilized during the study; consequently, the missing data they contained was not addressed.

Handling missing values

Based on discussions with the expert and the end-user, logical case-based rules have been established to impute the missing values in most of the reported cases. The variables mgUR12m (F1), Ncontainers, Dosage, KgPM (F3), and Origin of piglets (Staggered) were excluded from the analysis due to their high percentage of *missing* data.

It is important to note that there were no missing data in the auxiliary table for carcass collection, hence no table with logical rules exists for this.

Table 2. Logical rules for processing the F1 file

Variable	New value	Logical rule
N_Prescriptions	0	If (N_Prescriptions =missing) and (Census=0)
mgUR	0	If (mgUR=missing) and (Census=0)
Census	0	If (Census=missing) and (N_Prescriptions=0)

Table 3. Logical rules for processing the F2 file

Variable	New value	Logical rule
Weight	0	If (Weight=missing) and (Census=0)
Census	0	If (Census=missing) and (UR=0)
IdCT	"99"	If IdCT=missing
DesCT	"Uncategorized"	If DesCT=missing

Table 4. Logical rules for processing the file F3

Variable	New value	Logical rule
IdCZ	"100"	If IdCZ=missing
VET code	"9999"	If VET Code=missing
Sex VET	"NA"	If Sex VET=missing

Lastly, 288 null records were removed, meaning that all columns were empty (*missing*). [Table 5. Logical](#)

[rules for processing the file Staggered](#)

Variable	New value	Logical rule
Reason for rejection	"no description"	If Reason for rejection =missing
Integrator	"no description"	If Integrator =missing
Comments	"no description"	If Comments =missing
Authorization date	"01/01/1900"	If Authorization date =missing
PEG code	"PEG999"	If PEG Code =missing
Buildings	"no description"	If Buildings=missing
farm code	"9999"	If farm code=missing
Total score of the last Biosecurity	Average of all values of the variable	If Total score of the last Biosecurity Survey/Act =missing

Table 6. Logical rules for processing the file Record of farms

Variable	New value	Logical rule
ZOOTECHNICAL CODE	"9999"	If ZOOTECHNICAL CODE=missing
INTEGRATOR	"no description"	If INTEGRATOR=missing
ADS NAME	"no description"	If ADS NAME=missing
PRODUCTIVE CAPACITY	"no description"	If PRODUCTIVE CAPACITY=missing
BREEDING METHOD	"no description"	If BREEDING METHOD=missing
SUBSPECIES	"no description"	If SUBSPECIES=missing
SUBFARM END DATE	"no description"	If SUBFARM END DATE=missing
ZOOTECHNICAL CLASSIFICATION	"9999"	If ZOOTECHNICAL CLASSIFICATION=missing
SUBFARM STATUS	"no description"	If SUBFARM STATUS=missing
FARM POSTCODE	"no description"	If FARM POSTCODE=missing
SUSTAINABILITY CRITERION	"no description"	If SUSTAINABILITY CRITERION=missing
SMALL CAPACITY	"no description"	If SMALL CAPACITY=missing
CENSUS UPDATE DATE	"01/01/1900"	If CENSUS UPDATE DATE =missing
CAPACITY UPDATE DATE	"01/01/1900"	If CAPACITY UPDATE DATE=missing
ZOOTECHNICAL CLASSIFICATION DATE	"01/01/1900"	If ZOOTECHNICAL CLASSIFICATION DATE=missing
SUBFARM STATUS CHANGE DATE	"01/01/1900"	If SUBFARM STATUS CHANGE DATE=missing

Table 7. Logical rules for processing the file Biosecurity surveys

Variable	New value	Logical rule
Zootechnical classification	"no description"	If Zootechnical classification=missing
Health rating	"no description"	If Health rating=missing
ADS	"no description"	If ADS=missing
Integrator	"no description"	If Integrator =missing

Handling outliers

Outliers were identified in the dosage amounts from several farms; therefore, an expert in the field was consulted and indicated that these represented outliers in the population. Consequently, the analyses were conducted without excluding outliers. It was essential to monitor extreme values of antibiotic consumption cases of malpractice in antibiotic use as these cannot be included in the study, given that our aim is to identify the characteristics of these farms, among other factors.

Variable engineering

Automated coding has been conducted not based on the end user, but by selecting the most significant variables from the entire database, including both main and auxiliary files. Consequently, three new files have been created with data aggregated by medication type, location, mortality records, and biosecurity surveys, to establish a more comprehensive database that will serve as input for the visualization panel in Power BI.

Table 8. Summary of files resulting from variable engineering

Name	Description
MasterTableF1	Table with the variables of F1 with 11 more variables added from F3
MasterTableF1-Extended	Table with MasterTableF1 variables and 10 more variables from auxiliary tables
F1Master-clust-Extended	MasterTableF1-Extended with the variable c4 of the groups added

As each table is nested within the next, the first five observations of the added variables for each file are presented below.

Figure 11. Image of the first 5 records of MasterTableF1

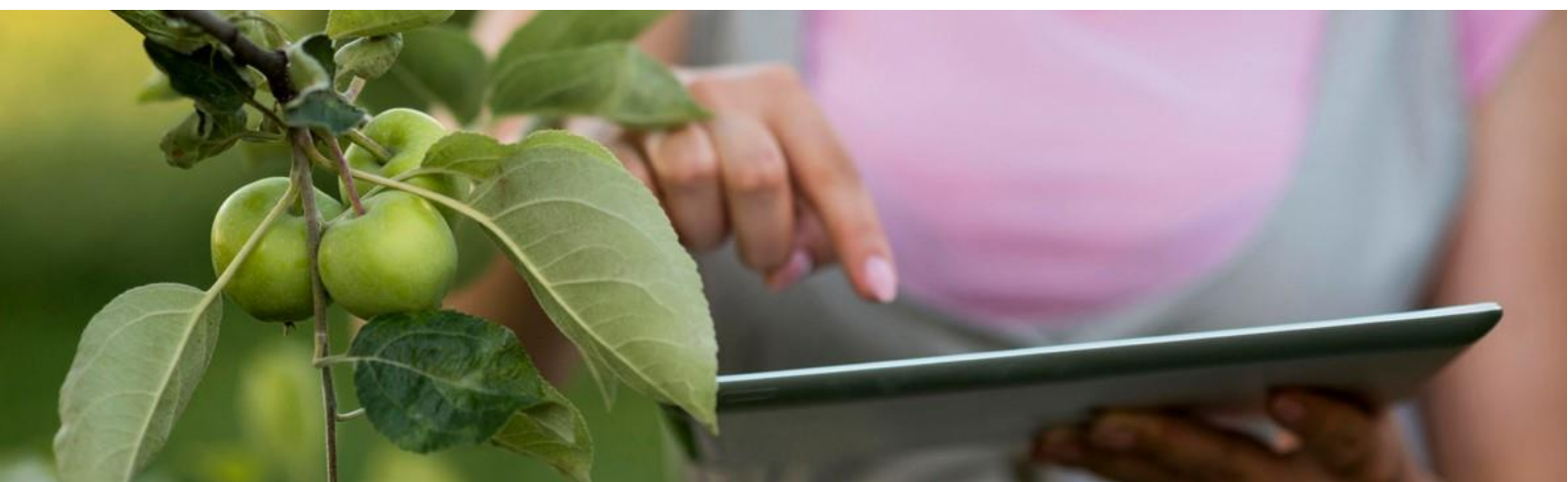
Anio_Trimes	Peso	peso2	mgConsumo_B	mgConsumo_C	mgConsumo_D	pct_mgConsumo_B	pct_mgConsumo_C	pct_mgConsumo_D	CantidadSA_B	CantidadSA_C	CantidadSA_D
2021_2	0	0	0	0	0	0	0	0	0	0	0
2021_2	65	65	250000	6300100	22311000	0,87	21,83	77,3	200	1366,7	2194
2021_2	73,8235294	142,5	44250	7020000	10112564	0,26	40,87	58,87	385	2000500	3289,6
2021_2	43,8048799	38,3333333	1505000	19895150	38330000	2,52	33,31	64,17	830	504963,5	6270
2021_2	65	65	0	0	5566000	0	0	100	0	0	400
2021_2	65	65	50000	150000	750000	5,26	15,79	78,95	100	300	300

Figure 12. Image of the first 5 records MasterTableF1-Extended

Codi_postal	Servei_territorial_explotació	Provincia_explotació	Comarca_explotació	Municipi_explotació	Cadavers_Total_animals	Promedio_bloc1	Promedio_bloc2	Promedio_bloc3	Prom_punt_total
8630	Barcelona	Barcelona	Baix Llobregat	Abrera	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA

Figure 13. Image of the first 5 records of MasterTableF1-Extended

Promedio_bloc1	Promedio_bloc2	Promedio_bloc3	Prom_punt_total	c4
NA	NA	NA	NA	4
NA	NA	NA	NA	3
NA	NA	NA	NA	3
NA	NA	NA	NA	3
NA	NA	NA	NA	4
NA	NA	NA	NA	4



Summary of new calculated variables

The transformed variables corresponding to the three antibiotic type aggregate files created during variable process are described here.

Table 9. Summary of the new variables transformed in the variable engineering phase

Variables	Description	File from which they derive
Year_Quarter	Concatenation of the Year and Quarter variables	F1
Weight2	Weight calculation with UR/Census	F1/F2
mgConsumption_B	Consumption of type B antibiotic in mg	F3
mgConsumption_C	consumption of type B antibiotic in mg	F3
mgConsumption_D	consumption of type B antibiotic in mg	F3
pct_mgConsumption_B	consumption percentage of antibiotic type B	F3
pct_mgConsumption_C	consumption percentage of antibiotic type C	F3
pct_mgConsumption_D	consumption percentage of antibiotic type D	F3
QuantitySA_B	Amount of type B antibiotic according to doses	F3
QuantitySA_C	Amount of type C antibiotic according to dose	F3
QuantitySA_D	Amount of type D antibiotic according to dose	F3
Total_animal_carcasses	Total number of pig carcasses per year and quarter	Collection of carcasses
Average_Block1	Block I average score per year and quarter	Biosecurity surveys
Average_Block2	Block II average score per year and quarter	Biosecurity surveys
Average_Block3	Block III average score per year and quarter	Biosecurity surveys
Total_point_avg	Total average score per year and quarter	Biosecurity surveys
C4	Classes of data identified and assigned to clustering	



Data mining

We have worked with factorial and biplots techniques to extract the present relationships of the variables through a multivariate analysis of the data and to understand in a dimensionality reduction map the relationships of the incident variables in the case study.

PCA Analysis of the overall relationships between the variables

The first PCA experiment (Lebart, 1995) with the numerical variables of the F1 summary table showed a total independence of the UR and mgUR variables with respect to the rest. To see if the results were affected by the size of the farm (size effect), it was decided to experiment by scaling the variables of some of the experiments by the variable Census to eliminate this effect.

$$UR = Censo \times Peso$$

$$mgUR = \frac{mgTotal}{UR}$$

Thus, principal component analysis was performed with different numerical variables from the F1, MasterTableF1 and MasterTableF1-Extended files depending on whether the data were scaled by the Census.

Table 10. Summary of the 4 PCA experiments, data and numerical variables

Experiment	Data	Numerical variables	Justification
1	F1 initial	IdCz, Census, UR, mgTotal, N_Prescriptions, mgUR, RI	Initial experiment with F1 supplied by the end user
2	F1 scaling	IdCz, UR, mgTotal, N_Prescriptions, mgUR, RI	The same experiment as before, but this time using a per capita study by the census to eliminate the potential size effect of the farms.
3	MasterTableF1 with unscaled absolute consumptions	IdCz, Census, UR, mgTotal, N_Prescriptions, mgUR, RI, Year_Quarter, Weight, mgConsumption_B, mgConsumption_B,	Enrichment of experiments 1 and 2 by engineering variables
4	MasterTableF1 with the proportions of consumption	IdCz, UR, mgTotal, N_Prescriptions, mgUR, RI, Year_Quarter, Weight, pct_mgConsumption_B, pct_mgConsumption_B, pct_mgConsumption_B.	At the expert's suggestion, experiment number 4 was repeated, altering the absolute consumption based on the proportions



A3. Methodology of the governance and sustainability model

The methodology used to draw up the Governance and Sustainability Plan was based on four main cornerstones. The first cornerstone involved analysing the protection of intellectual property (both existing and anticipated). In this phase, we collaborated with all institutions involved in the project to register all intellectual property contributed to and arising from the project. The second, third, and fourth cornerstones were developed through interviews with key actors in the sector, including livestock farms, cooperatives, and sector associations. These cornerstones focus on the general assessment of the algorithm, the assessment of the Dataspace, and the suggestions made by the interviewees.

Intellectual property protection analysis

-

The main objective of the analysis is to identify, document

and protect the intellectual property associated with the High Impact Project.

This process will ensure that the value of the project is maintained and that intangible assets are properly managed:

1. **Preserving project value.** Identifying, documenting and protecting intellectual property preserves the investment and resources that went into developing the project.
2. **Risk minimization.** The IPR assessment helps to minimize the risks associated with unauthorized disclosure or misuse of the technology developed. This analysis allows preventive measures to be taken to protect the project's intangible assets.
3. **Negotiating and working with external bodies.** Intellectual property analysis is essential for negotiating and establishing clear agreements on property rights. This transparency facilitates collaboration and avoids future legal disputes.
4. **Legal compliance.** Attention to intellectual property protection ensures that the project complies with relevant laws and regulations.
5. **Facilitate future development.** IP analysis provides a basis for deciding on licensing strategies, commercialization or other forms of technology value enhancement.

6. **Alignment with strategic goals.** The analysis is aligned with the strategic objectives of the project, ensuring that IP-related decisions are consistent with long-term goals.

The methodological steps are as follows:

- **Identification of technologies and IP elements:**
 - Make an exhaustive list of all technologies and IP elements related to the project.
 - Specify the name of each technology or element.
- **Authorship:**
 - Authorship: Identification of the authors of each technology or element.
 - Details of each author's specific contribution to the development of the intellectual property.
- **Ownership:**
 - Identify the ownership of each technology or element.
 - Identify the entities or individuals that own the intellectual property.
- **Conditions of use/Operation:**
 - Establish the conditions for the use and operation of each technology or element.
 - Include any intellectual property restrictions or licences.
- **Foreground forecast:**
 - Create a similar table for technologies or IP elements that are expected to be generated in the future (foreground).
 - Include the same categories: Name of technology, authors, ownership, conditions of use/operation.

Overall assessment of the solution (algorithm)

Presentation of the solution

By installing sensors and cameras in a pig pen, the software provides the position of each pig at every moment in time. This allows for individual

monitoring of the animals' state (welfare, diseases, etc.) and their behaviour (time spent drinking, eating, walking, etc.). As a result, a series of statistics or KPIs can be obtained, which are useful for owners/managers of livestock farms at various levels. These include space optimization, productivity, disease control and management, compliance with regulations, and obtaining certifications or quality seals.



Assessment

- **Assessment and level of general interest in livestock farms:** Advantages and disadvantages, potential and utilities.
- **Assessment of information:** Assessment of the statistics and KPIs provided. Also other information that may be of interest (short, medium and long term).
- **Interface assessment:** Importance of customization vs. standardization.
- **Pipeline assessment (process from the time a farm wants to participate until results are obtained):** Camera installation (assessment and implications), data processing and obtaining KPIs, timings.
- **Overall assessment of the solution:** Most relevant elements, profile of farms that may be interested, permeable to the introduction of the technology. Requirements and conditions. Price that they would be willing to pay.
- **Assessment of the shared data space (dataspace):**
 - **Data governance.** When a user decides to 'contract' a service, they must sign an agreement that grants them access to the server or cloud where data is managed. To utilize the algorithm, the user needs to either supply the required images or, if they lack the resources, install sensors or cameras on their property. These images are then automatically uploaded to the server. Once this is done, the company can generate the specified KPIs or statistics. By subscribing to the dataspace, farms—whether they subscribe to the algorithm service or not—can access a shared image/video and data repository from other farms to gather valuable insights. Additionally, all data contributed helps to train the algorithm, enhancing the model's capabilities and stability.
 - **Assessment of the data space.** Advantages and disadvantages, suggestions for improvement.
 - **Data space shared data access utilities.** For livestock farms, for cooperatives and associations
 - **Paid access vs. public and free access.** The possibility of the service being a one-time payment or recurring payments, versus the possibility of it being a free service with public access.

Suggestions and conclusions

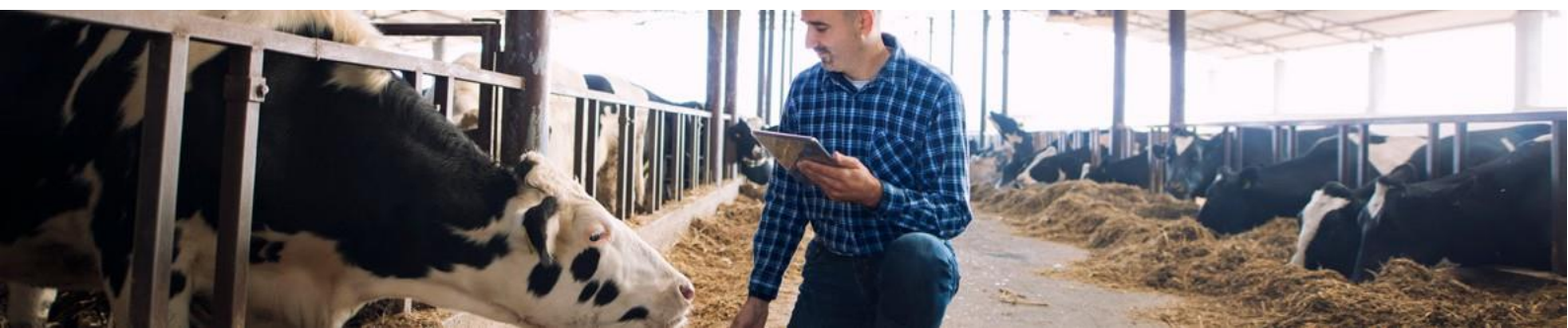
Using a methodology built on four key cornerstones, we have crafted the Governance and Sustainability Plan for the High Impact Project in the agricultural sector. This plan spans from intellectual property assessment to engagement with critical stakeholders, with each step meticulously planned to lay the groundwork for future growth. Our goal with this strategy is to safeguard our intellectual assets while developing an algorithmic solution that has a beneficial and enduring environmental impact.

Suggestions

1. **Optimizing the Intellectual Property Analysis Procedure.** Exploring tools or technologies that can streamline the collection of intellectual property data, such as intellectual property management platforms or automated registration systems.
2. **Expanding Participation:** By involving more institutions and organizations in intellectual property analysis, we can achieve a richer and more varied perspective.
3. **Ongoing Development of Foreground IP:** Creating a system for the ongoing assessment of new intellectual property, regularly updating records as new technologies or IP components emerge.
4. **Promoting Collaboration within the Shared Dataspace:** Launching programs to foster cooperation and active data sharing among businesses and farms in the Dataspace, thereby enhancing the data pool's volume and quality.
5. **Enhancing Transparency in Data Management:** Fortifying data governance practices to ensure that all users comprehend and consent to the data usage policies within the Dataspace.

Conclusions

1. **Added Value of Intellectual Property Analysis:** The methodology for analysing intellectual property is crucial for maintaining the project's worth and mitigating risks, offering a robust framework for the stewardship of intangible assets.
2. **Significance of Sectoral Interviews:** Engaging with pivotal industry players through interviews enhances our comprehensive grasp of the project, bringing fresh insights that inform our evaluation of the algorithm, the Dataspace, and other vital elements.
3. **Algorithmic Solution With Beneficial Outcomes:** The algorithmic solution we propose brings considerable advantages to livestock farms, enhancing animal care and welfare, while also producing useful statistics that can elevate productivity and product quality.
4. **Dataspace as a Key Asset:** The Dataspace serves as a vital asset, providing access to collective data, aiding in the evolution of the algorithmic model, and offering practical benefits to both livestock farms and other sector participants.
5. **Future Outlook:** Looking ahead, we aim to delve into new technologies, strive for ongoing process enhancements, and cultivate a collaborative atmosphere that drives innovation and sustainable growth within the livestock industry.



References

[Contents](#)



- Catalonia, Gov. (no date). *Les agendes compartides de la RIS3CAT 2030 (The shared agendas of RIS3CAT 2030)*. Retrieved from <https://fonseuropeus.gencat.cat/ca/ris3cat/2030/agendes-compartides/index.html>
- Computer, C. d. (2022). Retrieved from <https://cdn.eurecat.org/PDF/CIDAI/PoC/POC-CVC-110523.pdf>
- Gaia-X. (no date). *Pontus-X*. Retrieved from Streamlined interoperability across data ecosystems: <https://www.pontus-x.eu>
- Fayyad, U. P.-S. (1996). *From Data Mining to Knowledge Discovery in Databases*. 17.
- Gibert K, S.-M. M. (2016). A survey on pre-processing techniques: relevant issues in the context of environmental data mining. 29(6).
- Gibert, K. I.-M.-R. (2018). Which method to use? An assessment of data mining methods in Environmental Data Science. *Environmental modelling & software*, 110, 3-27.
- Lebart, L. M. (1995). *Statistique exploratoire multidimensionnelle (Vol. 3)*. Paris: Dunod.
- RStudio_Team. (2020). *RStudio: Integrated Development for R*. Recollit de <http://www.rstudio.com/>
- Gibert, K. G.-R.-S. (2008). The role of KDD Support-Interpretation tools in the conceptualization of medical profiles: An application to neurorehabilitation. *Acta Informatica Medica*, 16(4), 178.

Promoters



Members

