Evaluating Farm Management Practices: A semi-supervised clustering approach to assess the impact of technologies

Caspar Matzhold[1], Katharina Schodl[1], Franz Steininger[1], Peter Klimek[2], Christa Egger-Danner[1]

[1]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria

[2]Complexity Science Hub Vienna, 1080 Vienna, Austria

Introduction

Background:

- Diversity of data and information available on farms is continuously increasing.
- Animal health and welfare is influenced by many factors.

Objective of Research:

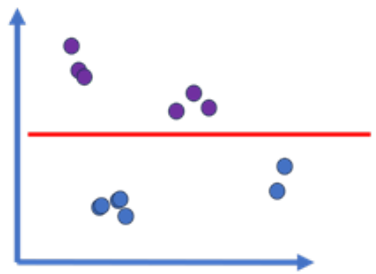
- Integration of diverse data sources.
- **Farm Management ⇄ Health**
- To gain insights that will assist in:
  - Planning of diverse farm management strategies that will improve bovine health.
  - Identifying the most effective management style for a specific strategy through benchmarking.

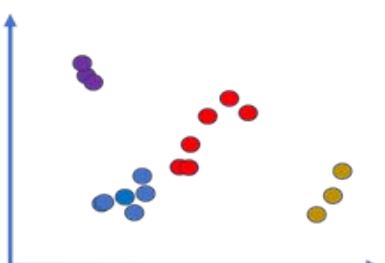
Challenges:

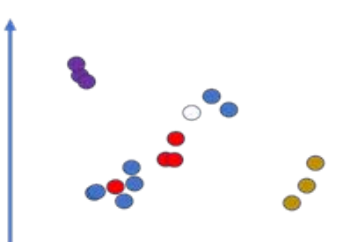
- Access & Cleaning & Validation of diverse data sources.
- Development of a methodological approach to transform high-dimensional data into knowledge on clusters, representing a certain management style.

Methodological Approach to Extract Clusters

Approach	Description	Challenges and Opportunities
Supervised	Knowledge-Driven: existing knowledge to identify clusters	•Distinctive clusters e.g., high discriminatory power •Potential oversight of novel feature interactions
Semi-Supervised	Combination of knowledge-based feature selection with unsupervised clustering methods	•Feature-selected distinctive clusters that exhibit high discriminatory power •Potential novel cluster formation is anticipated
Un-supervised	Data-driven: Machine learning methods to identify patterns / clusters within the data	•Uncover new patterns of similarity (Clusters) •Less accurate result due to low discriminatory power







Semi-Supervised: Key-Feature-based Clustering Approach

- **What is a key-feature?**
  - A key-feature is known to have a significant impact on farm management.
    - Milking System AMS: Yes/No
    - Altitude: High/Low

- **Why should we use it?**
  - **Reduce dimensionality:** Define similarity spaces for effective cluster extraction.
  - **Increase discrimination power:** Improve capability to distinguish between clusters.
  - **Targeted analysis:** To define the pattern for the similarity measurements allows for more detailed analysis of the effects of features of interest, such as sensor systems.

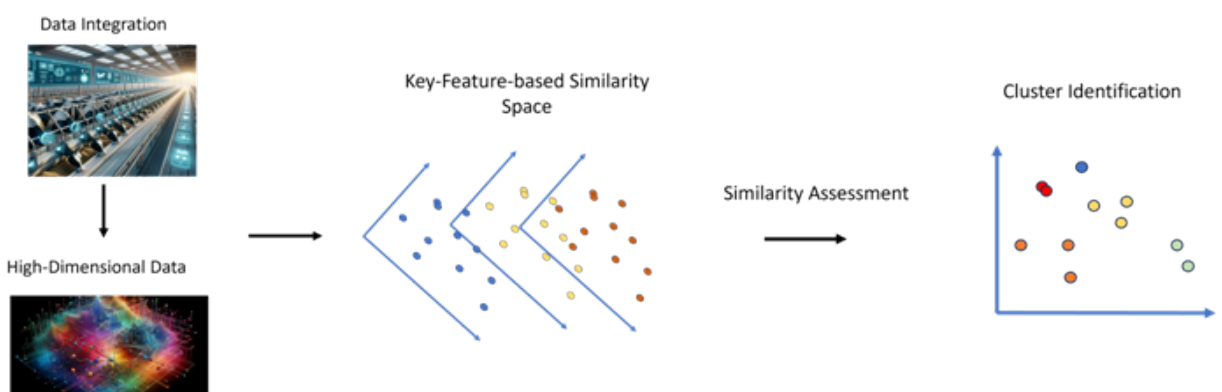
➢ **How to use it?**

Methodology: 1) Cluster Extraction

1. **Define a Set of Key-Features**
   - Spanning the similarity space.

2. **Assess Similarity**
   - Similarity is assessed among the key-features using similarity measurements such as Cosine distance for numeric features and Jaccard distance for binary features.

3. **Cluster Identification:**
   - The obtained vectors of similarity are then used as inputs for the unsupervised clustering algorithm HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) to group the clusters into sets of similarity.



Methodology: 2) Cluster Evaluation

1. **Disease Risk Assessment**:
   - Disease risks are analyzed by calculating **incidence rates (R)** and **odds ratios (OR)** for diseases across clusters.
   - ➢ *A disease risk is defined as the probability of a diagnosis occurring.*

2. **Pertinent Feature Evaluation**:
   - Clusters are characterized using **z-scores** for each feature, assessing statistical significance with t-tests or Fisher's exact test.

➢ **Cluster Evaluation Applied Within and Between Cluster Groups:**

- **Inter-Cluster Analysis**: Analyze trends across different cluster groups to discern overarching patterns.
- **Intra-Cluster Analysis**: Investigate interactions of features within similar clusters to delve into impacts beyond key-features.

Data

1) Farm dataset:

- 3.284 (8.000) farms described by **80 Hot-Encoded** Features describing:
  - Feeding-, Milking-, Housing-Systems,..
  - Husbandry Management, Pasture Strategies,..
  - Technology Applications like AMS and Sensor System,..

Farm Dataset	Counts
Farm	3.284
Cow	98.037
Feature	80

2) Health dataset:

- Extracted from the national cattle registry, comprises veterinary diagnoses and documented events observed during calving.
- Incidence Rates are calculated by summing the **unique counts** of diagnoses observed **on a farm over a 2-year** period.
- Diagnoses were **categorized into 3 groups**:
  1. Udder diseases (*Chronic and Acute Mastitis*)
  2. Fertility disorders (*Anestrous, Uterus Inflammation, Ovarian cysts*)
  3. Milk fever (*Hypocalcemia*)

Diagnose Group	Unique Counts over 2 year
Udder diseases	23.923
Milk fever	11.294
Fertility disorders	39.087
Sum:	**74.304**

Study - Objective

Demonstrate Clustering Approach:

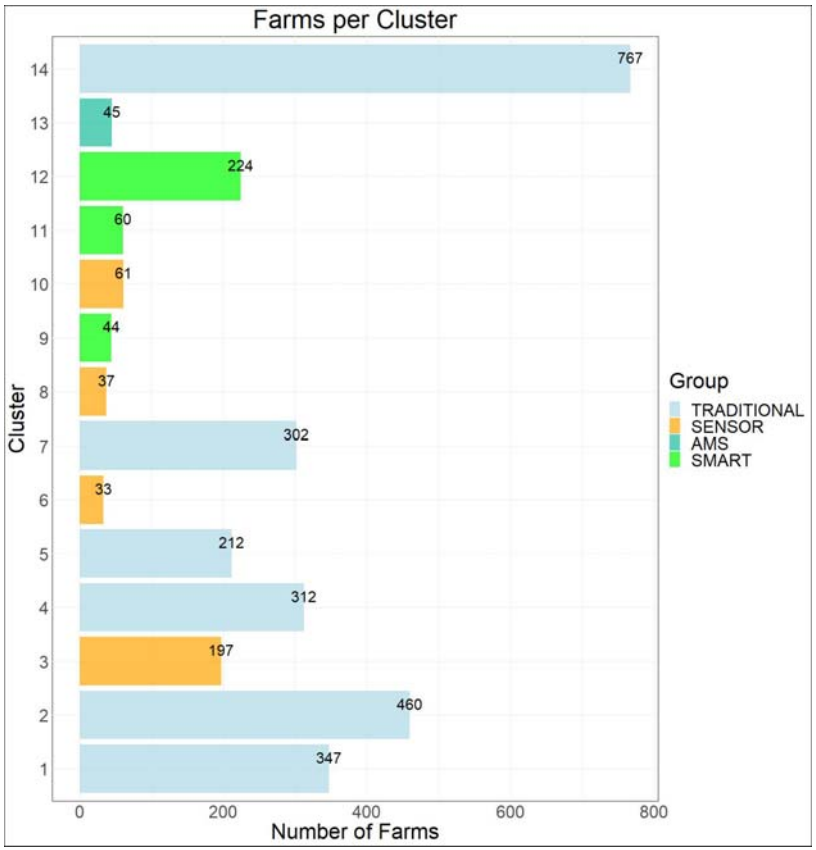
Show the effectiveness of a key-feature-based clustering method to extract and evaluate the impact of varied management practices on bovine health.

Focus on Technology Impact:

Assess the influence of Automatic Milking Systems (AMS) and sensor technologies on bovine health.

Key-Features
Herd size
Milk yield average per year
Altitude
Pasture
Access to open areas
Free-stall housing system
Automatic Milking System
Sensor

Results

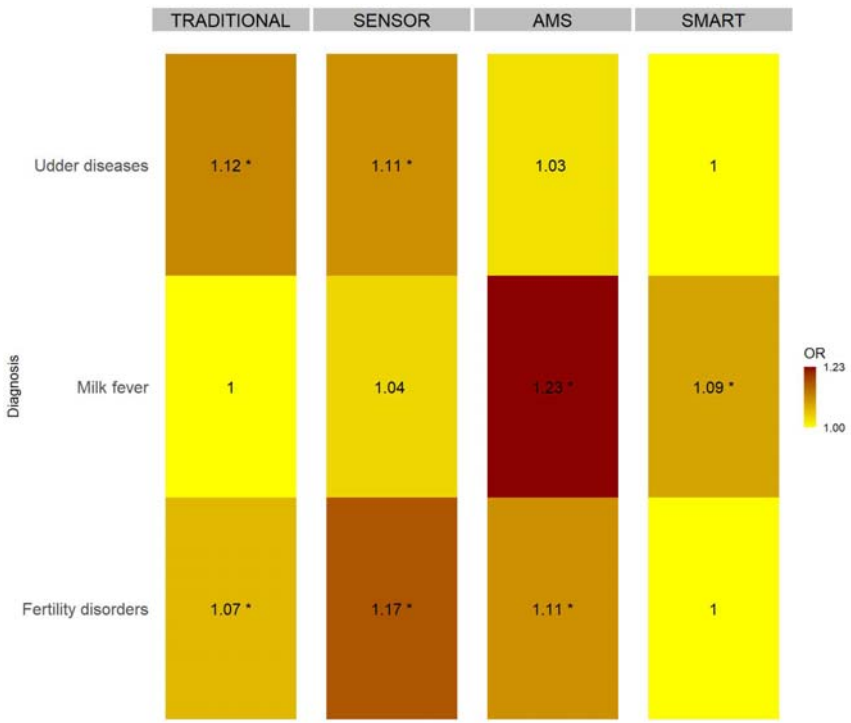


- Extracted 14 distinct farm management clusters.

- **Assignments:** 3.101 of 3.284 farms were successfully assigned to a cluster.

- Categorized based on **technological usage** into 4 groups:
  - TRADITIONAL
  - SENSOR
  - AMS
  - SMART (Sensor & AMS)

Inter-Cluster Results:

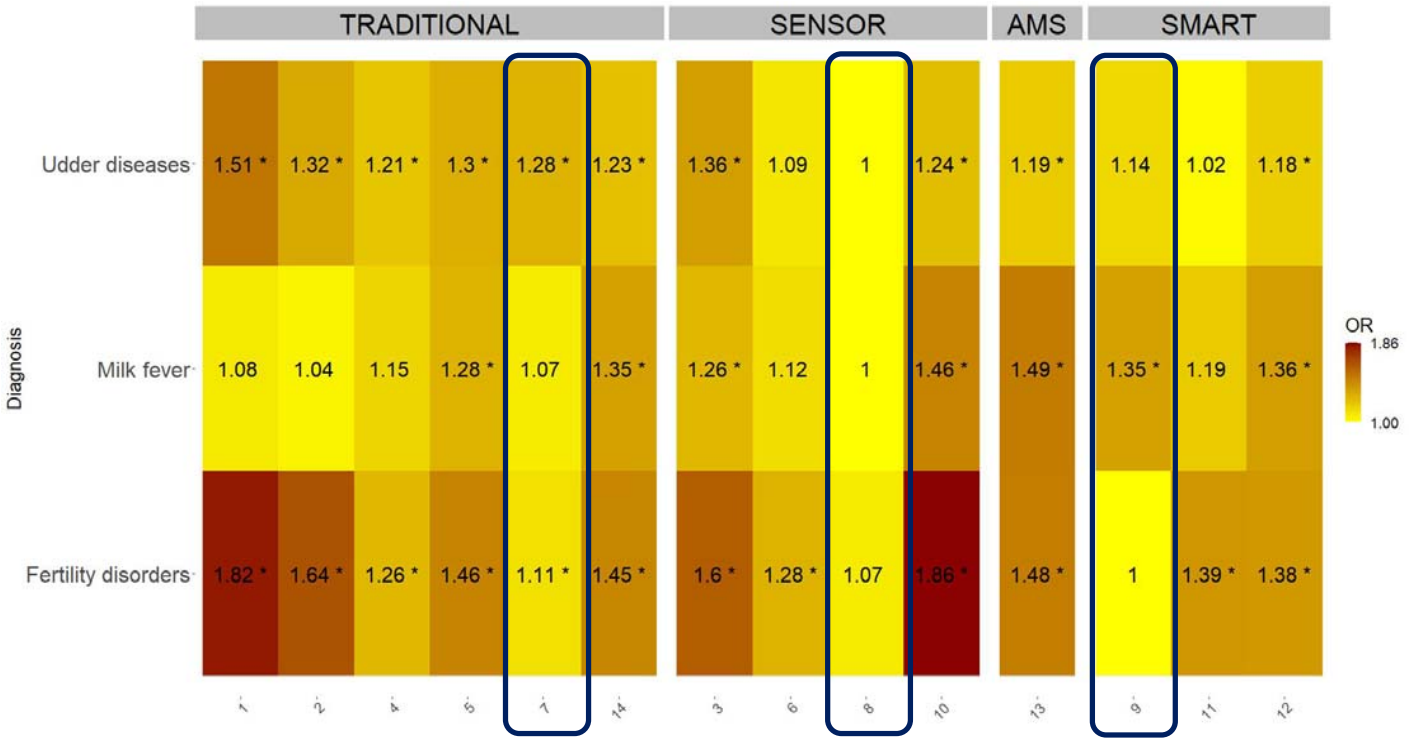


Grouped Disease Risks: Displayed as heat maps with yellow indicating the lowest disease incidence rate and red indicating the highest risk (Odds Ratio)

➢ On a group level:

- We find a **trend towards lower disease risks in technologically advanced** farms, especially in SMART farms combining AMS with sensor systems.
- SMART lowers incidence rates for udder diseases and fertility disorders.
- Traditional farming lowers incidence rate for milk fever.

Inter-Cluster Results:



Disease Risk: Displayed as heat maps with yellow indicating the lowest disease incidence rate and red indicating the highest risk (Odds Ratio)

➢On a cluster level:

•High variation in disease risks across and within clusters.

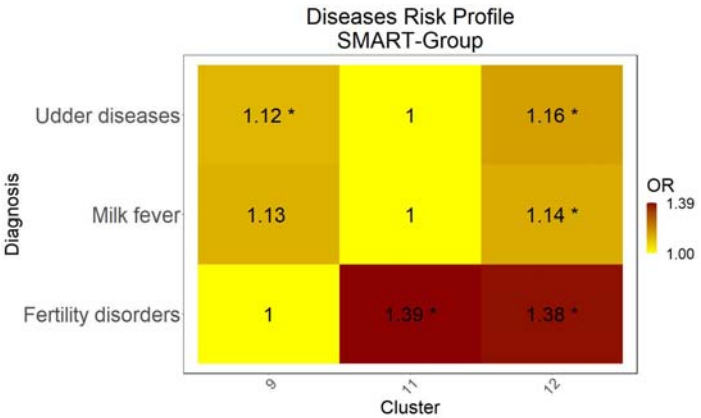
➢Pertinent Features Analysis of healthy clusters reveals that the feature combination

•**organic farming** and

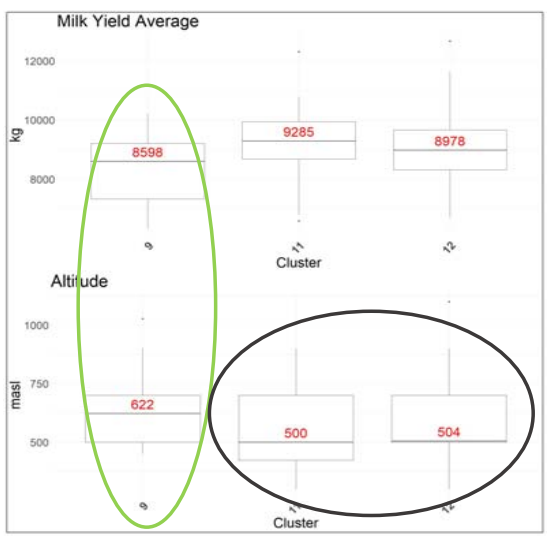
•**more freedom for movement**

has a positive impact on bovine health.

Intra-Cluster Results: SMART Group Analysis



Pertinent Numeric Features



Pertinent Binary Features

Cluster 11	Cluster 12	Cluster 9
Access To Open-Air Areas: Always	Access To Open-Air Areas: No	Pasture
Open-Air Floor: Paved	Management: Conventional	Pasture: Seasonal
Manure Removal: Manual	Summer: Corn Silage	Management: Organic
Calves: Outdoor	Manure Removal: Automatic	Access To Open-Air Areas: Always
DHI:Milk Yield Average	Access To Open-Air Areas: Partial	Open-Air Floor: Paved
Alley Flooring: Slats	Barn Design: Outdoor Climate House	Summer: No Silage
Calving Box: With Sick Animals	Concentrates: PMR	Altitude
Ventilation System: Lying Area	Forage Type: Mixed Ratio	Multiple Feeding Places
Sick Bay: Together With Calves	Sick Bay: No	Concentrates: TMR
Ventilation System: Feeding & Lying Area	Calves: Outdoor Climate House	Claw Trim: Tipping Crush
Concentrates: Individual	Ventilation System: No	Forage Type: Template
Claw Trim: Professional	Feed Pusher: Yes	Calves: Warm Stable
Lure Feed Concentrates	Calving Box: Separate	Ventilation System: No
Claw Trim: Standing	Alley Flooring: Solid Concrete	Milking Take-Off: Yes
Deep Bed Cubicles	Management: GMO-Free Feeding	High Bed Cubicles
Management: Conventional	Claw Trim: 3 Times A Year	Claw Trim: Low To No Training
Concentrates: TMR	Cooling System: No	
Ventilation System: Feeding Area		
Concentrates: Transponder Only		

Table 1. Intra-Group Analysis: Pertinent features comparison. Significant positive features of each cluster are arranged in descending order of importance, as determined by their z-values. Features with high z-scores are considered as pertinent features that characterize a particular cluster.

Conclusion

- We presented a **semi-supervised** clustering approach based on **key-features** that enables differential risk analysis between and within groups of clusters.
- We demonstrated how the approach can be used to **study the influence of specific features,** such as **technological systems**, on bovine health.
- Inter-Cluster Analysis **reveals trend** towards:
  - Lower disease risk for farm-integrated technological systems.
  - Lower disease risk, especially for fertility disorders, for an organic farm management that provides more freedom for movement.
- Intra-Cluster Analysis of the SMART-Group **reveals effects beyond key-features:**
  - Differences in heat, feeding, and pasture management are associated with variations in disease risks.

Practical Relevance

- **Improving Animal Health and Welfare**:
  - **Trends Analysis:** Understanding trends **supports better planning** and **decision-making** in livestock management.
  - **Benchmarking:** Identifying differences between similar management styles can aid in identifying the **best management style.**

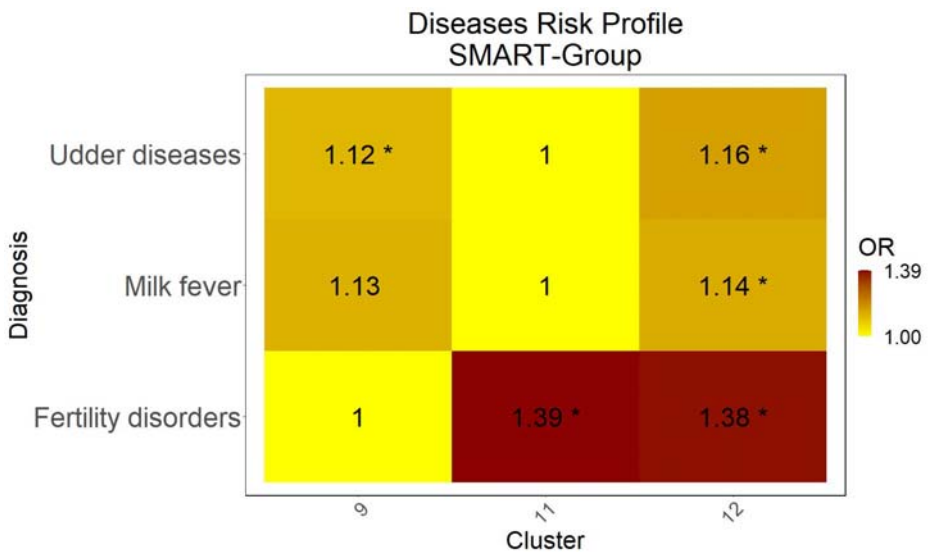
- **Complexity of Disease Management**:
  - There is no universal best farm management strategy for all diseases, as disease risks result from several interacting factors.
  - Farm Management $\rightleftharpoons$ Disease Risk

- **Our Key-Feature-Based Clustering Approach**:
  - Offers a **flexible framework** to gain insights into trends and differences between similar management styles, enhancing overall farm efficiency and animal health.

Thank You!

BitKuh

Intra-Cluster Results: SMART Group Analysis



Disease Risk: Displayed as heat maps with yellow indicating the lowest disease incidence rate and red indicating the highest risk (Odds Ratio)

➢ Pertinent Features

- **Organic vs. Conventional Farms:**

Organic farms (Cluster 9) show decreased occurrence of fertility disorders due to:

- Less intensive milk production.
- No corn-silage feeding.
- More freedom for movement.

- **Conventional Farm Comparison at similar altitude:**

Cluster 11 shows an overall decreased disease risk compared to Cluster 12, that can be attributed to:

- Cooling Systems: Reduces heat stress, lowering disease risks.
- Outdoor Access: Improves welfare and health.
- Concentrates Individual: Feeding: Meets specific nutritional needs, enhancing health.

Background:

- Diversity of data and information available on farms is continuously increasing.
- Animal health and welfare is influenced by many factors.

Objective of Research:

- Integration of diverse data sources.
- **Farm Management ⇌ Health**
- Provide benchmarking and targeted decision support aiming to improve farm efficiency and bovine health.

Challenges:

- Access & Cleaning & Validation of diverse data sources.
- Development of a methodological approach to transform high-dimensional data into knowledge on groups of similarities, representing a cluster.