Identification of the most impacting environmental variables on dairy cow's milk yield using Machine Learning methods

Daniele Pietrucci (1-2), Marco Milanesi (1), Balasubramanian Chandramouli (3), Giovanni Vignali (1), Riccardo Negrini (4), Marco Macini (5), Monia Santini (6), Giovanni Chillemi (1)

(1) Dipartimento per la Innovazione nei sistemi Biologici, Agroalimentari e 8 Forestali (DIBAF), Tuscia University, Viterbo, Italy, Viterbo, Italy

(2) Istituto di Biomembrane, Bioenergetica e Biotecnologie molecolari, IBIOM, CNR, Bari, 7 Italy, Bari, Italy

(3) Super Computing 9 Applications and Innovation, CINECA, Bologna, Italy, Bologna, Italy

(4) Associazione Italiana 10 Allevatori, AIA, Rome, Italy, Rome, Italy (4)

(5) Fondazione Centro Euro-Mediterraneo sui 11 Cambiamenti Climatici, Advanced Scientific Computing Division, Lecce, Italy, Lecce, Italy

(6) Fondazione Centro 12 Euro-Mediterraneo sui Cambiamenti Climatici, Impacts on Agriculture, Forests and 13 Ecosystem Services (IAFES) Division, Viterbo, Italy, Vierbo, Italy

daniele.pietrucci@unitus.it

Introduction

- Several reports show that **the effects of climate change are tangible for different aspects of life on earth** (e.g., World Health Organization, National Oceanic and Atmospheric Administration, Intergovernmental Panel on Climate Change)
- Such effects are particularly true for **the food production sector**, **in terms of economic losses and decreases in food quality** (Unanaonwi, 2014; Maulu et al., 2021; Singh et al., 2021; Chandio et al., 2022)
- Plus, **the effects of climate change are well-documented for milk production** (Sheik et al., 2017; Cheng et al., 2022)

The effects of climate change on milk production

Literature shows that environmental changes can lead to:

- **Reductions in milk yield**
- **Reduction in milk quality:** fat and protein content
- **Poorer cows' health:** higher somatic cell counts during heat-stress conditions: somatic cell counts are used as a proxy of animal health

(e.g., Nasr & El-Tarabany, 2017; Sheik et al., 2017; Cheng et al., 2022; Toghdory et al., 2022)

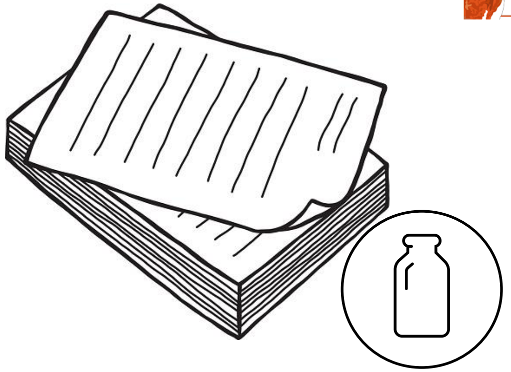
Our research question

- The short term effects of the climate on milk yield and production has been reported in literature (Biswal et al., 2019)
- It is not well-known **the individual impact of the climatic variables and how quickly these climatic conditions can affect milk production and quality.**
- We developed a Machine Learning pipeline to identify the most important climatic variables that may influence milk yield and quality and their long-term effect

Methods
Data collection

- The “**production**” dataset: dairy cows’ records from 1990 to 2020 (AIA, ANAPRI, LEO Project)

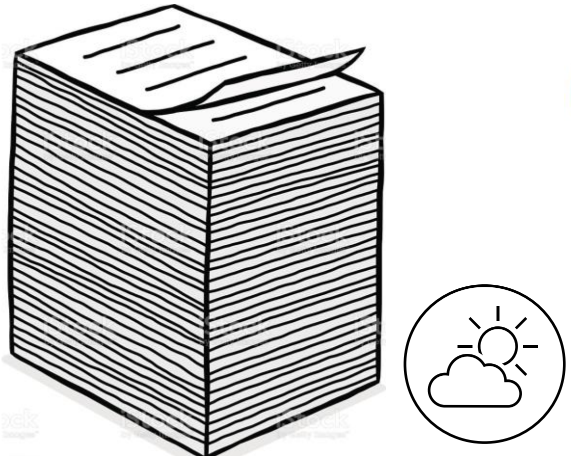
- The “**climatic**” dataset: daily meteorological information from 1990 to 2020 (downloaded from HL and CMCC's DDS)



Libro Genealogico Pezzata Rossa Italiana A.N.A.P.R.I. - Udine
Ente selezionatore A.N.A.P.R.I. Associazione Nazionale Allevatori Pezzata Rossa Italiana

Livestock Environment Opendata

A.I.A.

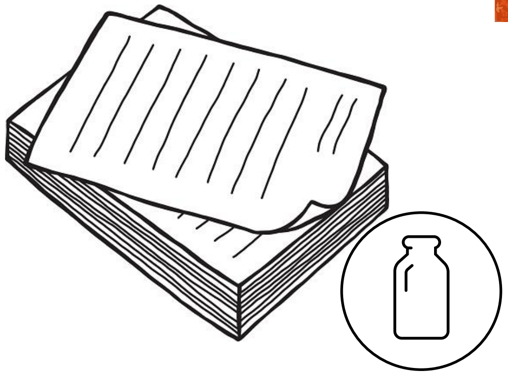


cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici

Highlander High performance computing to support smart land services

Methods
Data collection

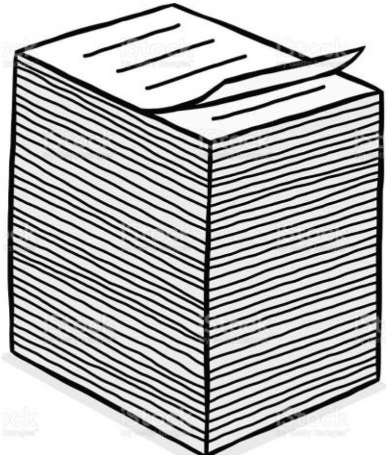
- “**Production**” dataset variables:
  - milk yield
  - fat
  - protein
  - SCC (Somatic Cells Count)
  - **date of functional control (FC)**
  - **latitude and longitude of the farm**
  - **EBV**
  - **Parity numbers**
  - **Farm ID / Animal ID**
  - **Days in Milk**

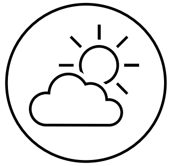


- Data were quality checked and filtered
- For each variable (milk yield, fat, protein, SCC) a Linear Model was used to mitigate the effects of genetic factors or farm management

Methods
Data collection

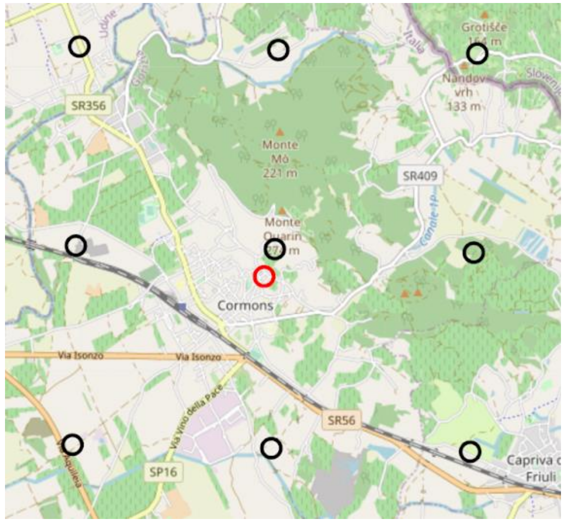
- “**<u>Climatic</u>**” dataset variables:
  - **date of the measurement**
  - **latitude and longitude**
  - Temperature (min, max, mean)
  - Relative humidity (min, max, mean)
  - Wind
  - Precipitation (total)
  - Cloud cover (total)
  - Discomfort Index (min, max, cumulated)

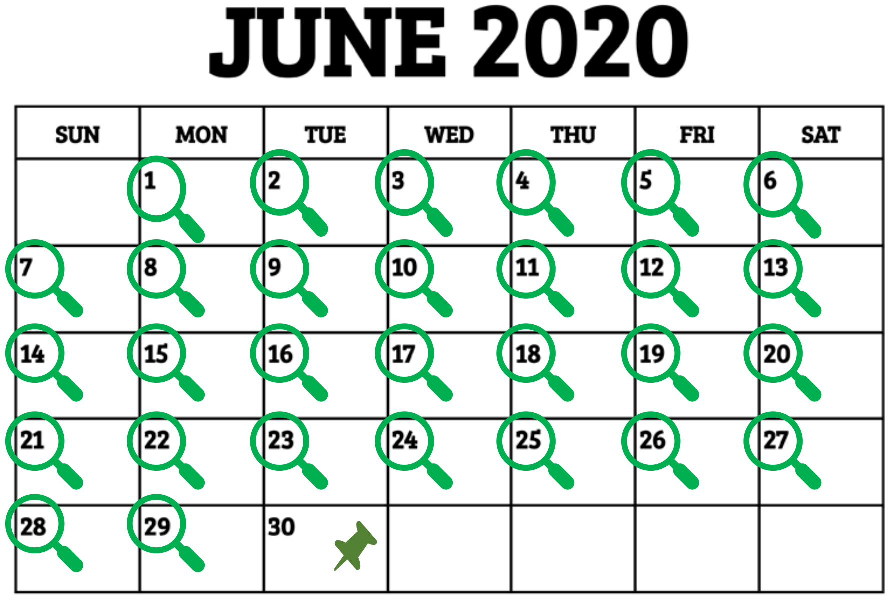




Methods
Data and plan of analysis

- The production and climatic data were paired
- The production of each animal **was paired with the climatic variables assessed for up to 30 days before the functional control date**





Methods
Data and plan of analysis

Pilot dataset: farms located in **Friuli-Venezia-Giulia**

- The number of FC analyzed in the Machine Learning approach to detect the best prediction model were:
  - **2,332,083** FC (1375 farms * 105,285 animals)
- The total number of climatic variables analyzed is:
  - **198**
- Given the high amount of data and dimensions, we applied a **Machine Learning (ML) approach to identify the best climatic variables affecting milk production**.

Methods
Our pipeline

1. **Remove collinear climatic variables** to maintain the most informative ones before data analysis
2. **Identify the best algorithm family** to use climatic variables as milk production predictors
3. **Optimize the parameters of the best algorithm** to increase its accuracy
4. **Evaluate the importance of all variables to subset the most important**
5. **Explain how each variable can affect milk production and quality**

Climate variables → Milk yield / quality / SCC

Features — Targets

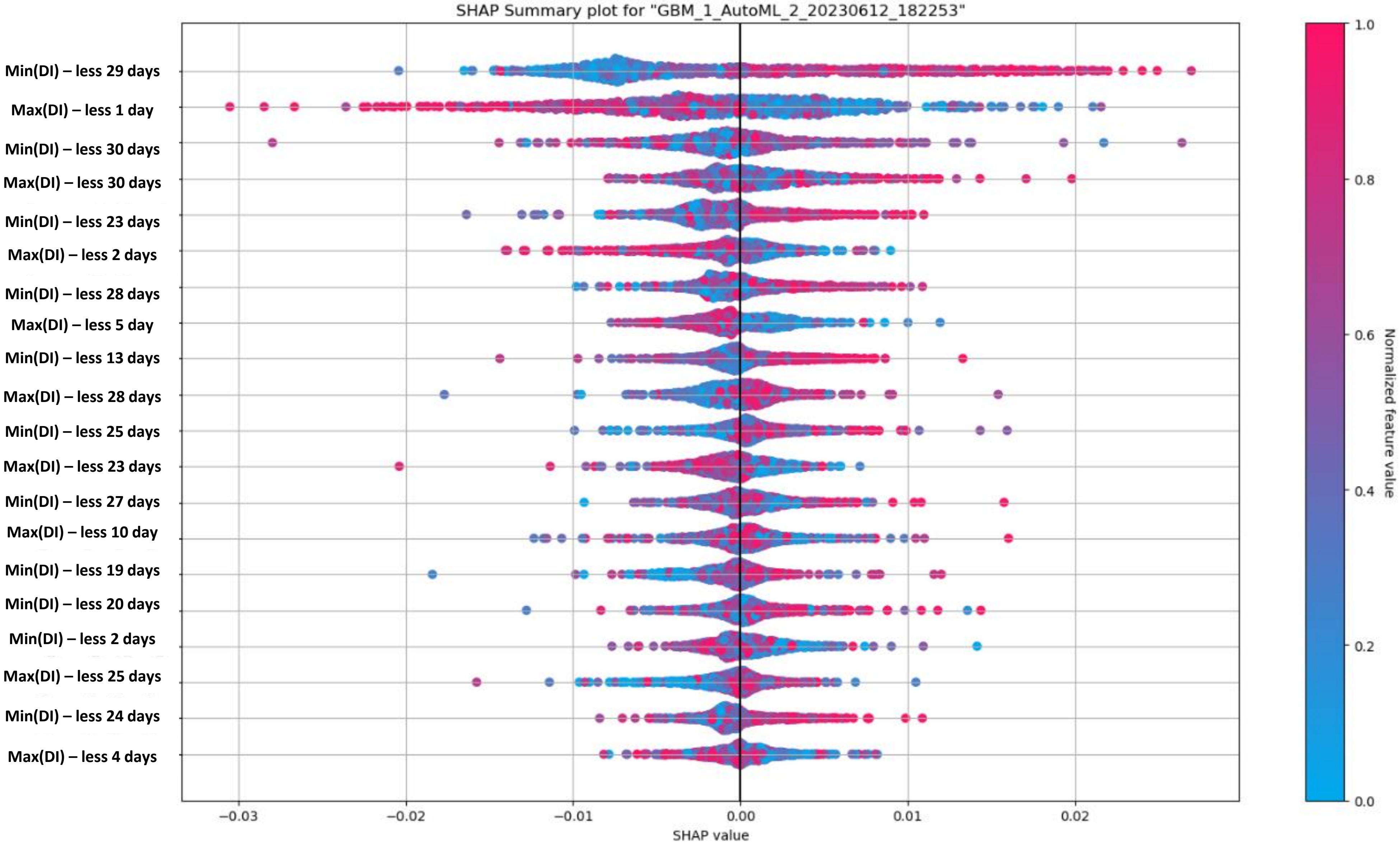
Results

Feature	Agorithm	Proxy	Root Mean Square Error	Mean Absolute Error	R-squared	Range
Milk yield	Random Forest	Production	3.90	2.91	0.06	[-32.93, +32.55]
SCC	XGBoost	Health	0.45	0.34	0.20	[-2.33, +2.70]
Protein	Gradient Boosting Machine	Milk quality	0.20	0.15	0.22	[-1.89, +2.77]
Fat	Gradient Boosting Machine	Milk quality	0.20	0.14	0.20	[-3.79, +3.32]

- Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used to predict the model accuracy
- They provides an estimate of the typical magnitude of prediction errors. Lower values indicate better model performance

Results

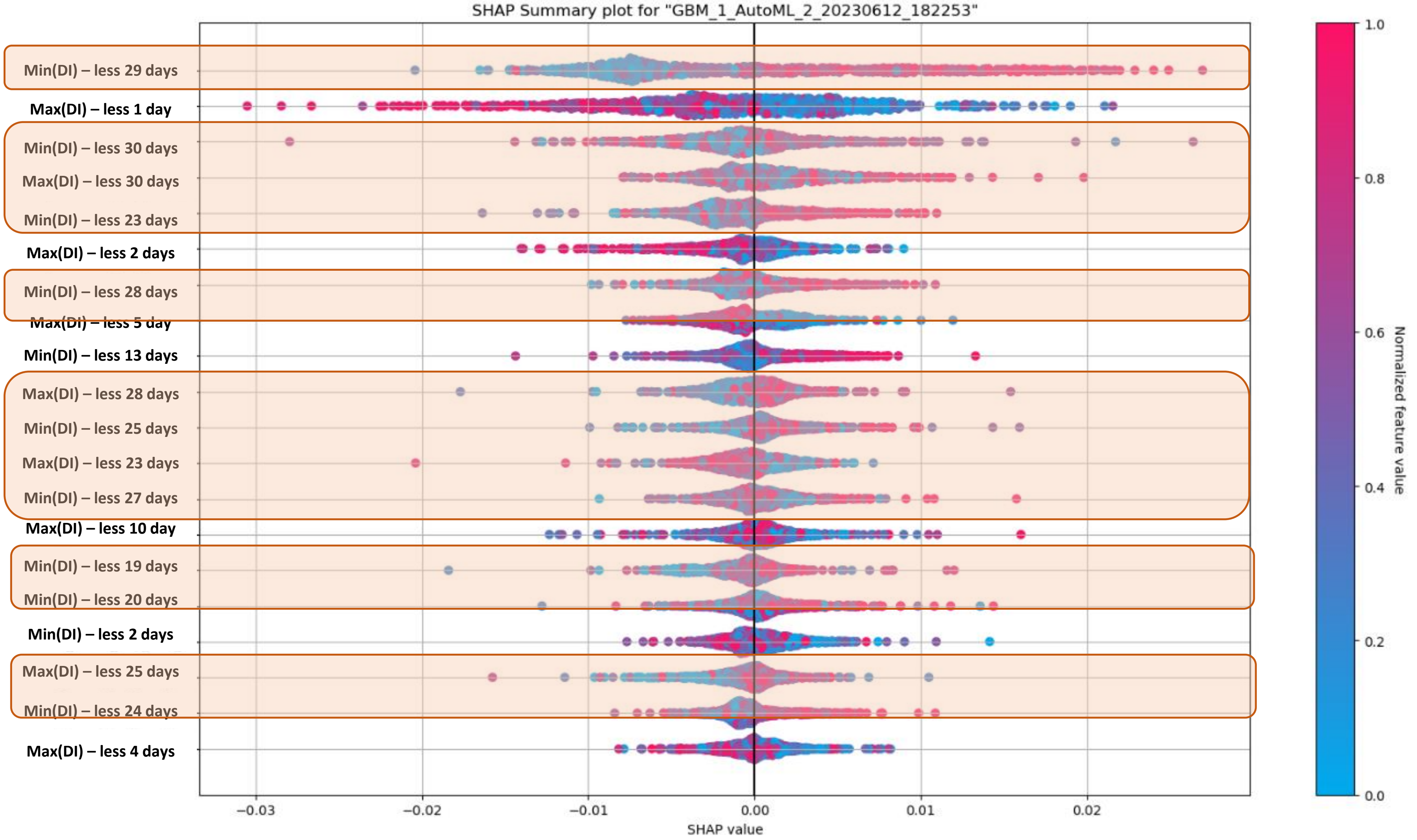
Protein %



- For each of the four phenotypic traits, the most important variables in determining the predictive model were identified using the SHAP (SHapley Additive exPlanations) algorithm

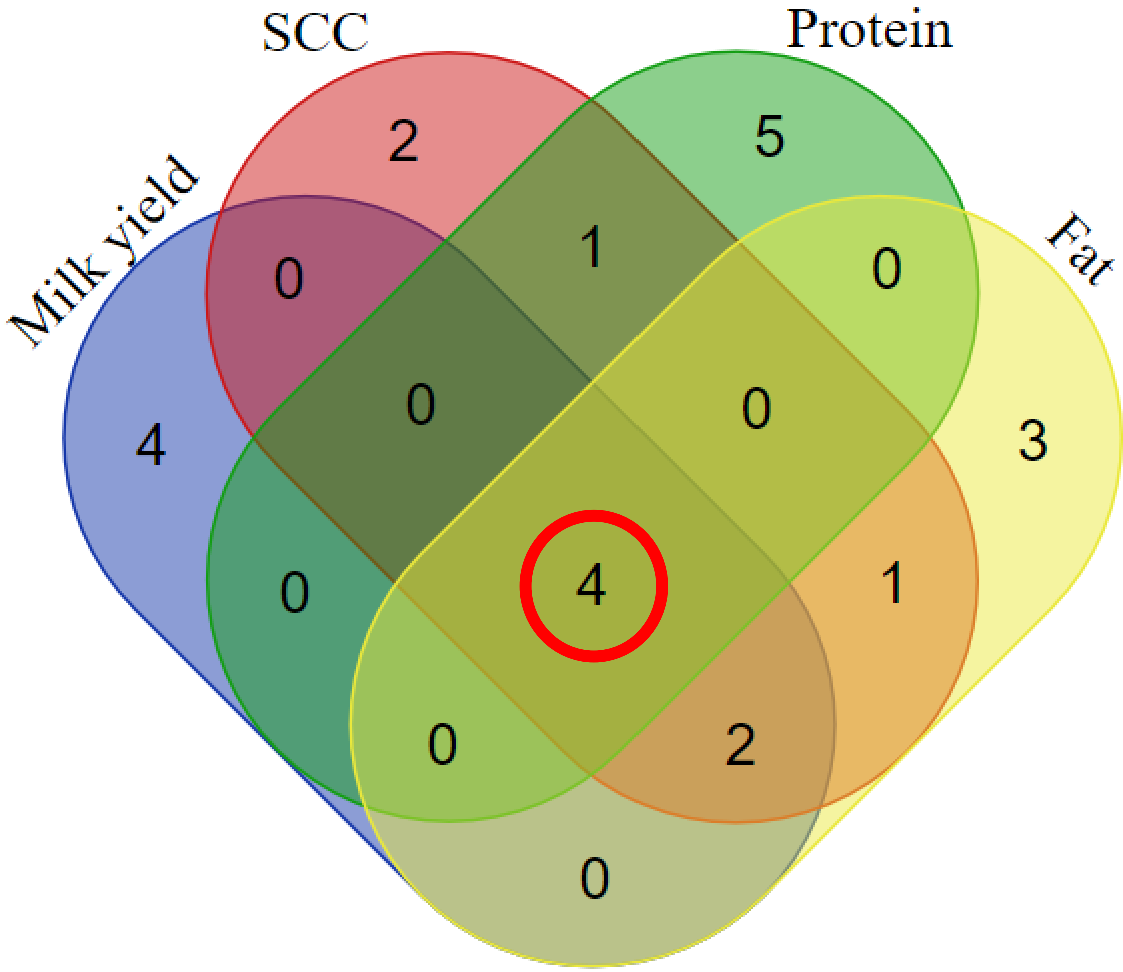
Results

Protein %



- Among the most important variables in explaining the prediction, **data from days well before the functional check are found**. **This suggests a potential long-term effect on milk production and quality**

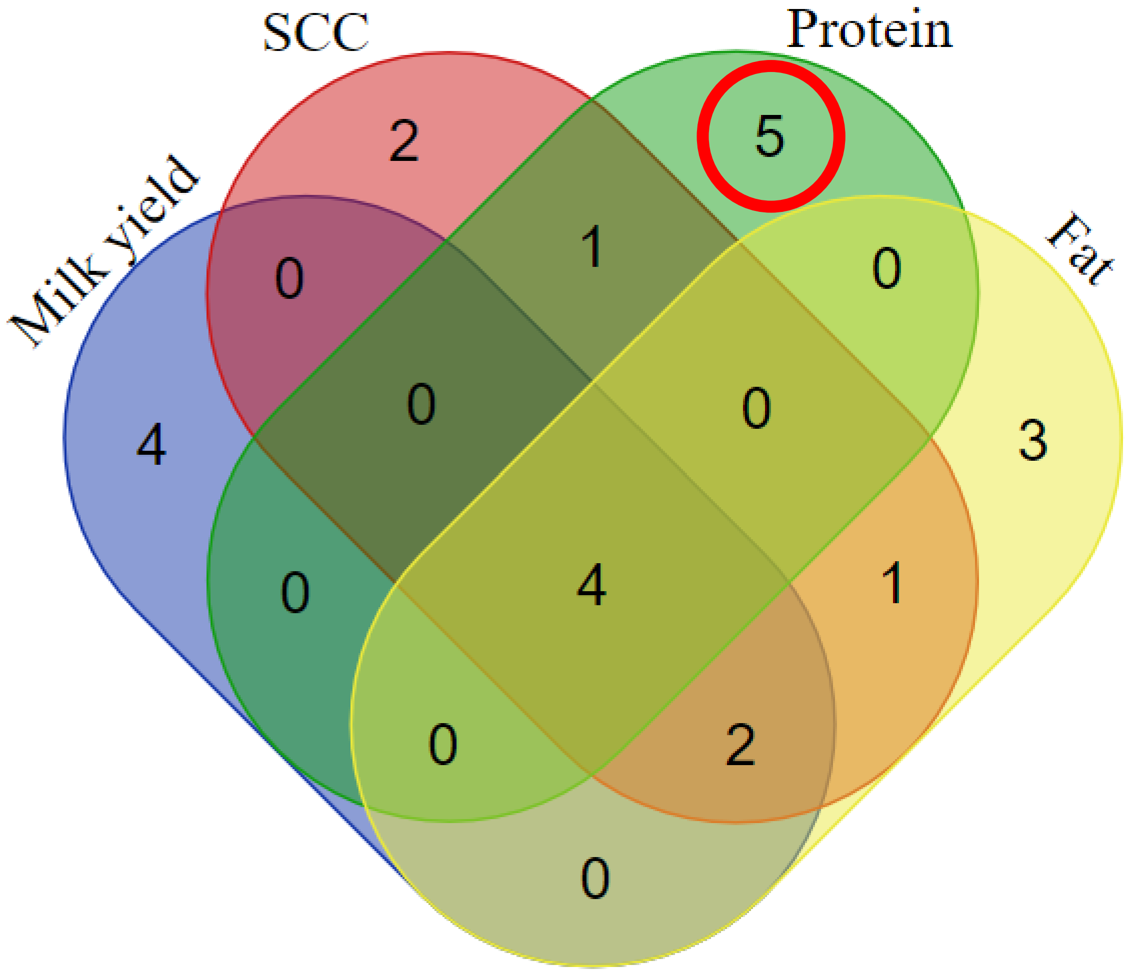
Results



- Several days emerged as important for each feature: **1**, **2**, **25** and **29** days before the Functional Control

- Some dates were specific for each variable, such as, for example, day 10, 16, 18, 25 and 26 for Proteins

Results



- Several days emerged as important for each feature: 1, 2, 25 and 29 days before the Functional Control
- Some dates were specific for each variable, such as, for example, day **10**, **16**, **18**, **25** and **26** for Proteins

Conclusion

- We have developed a **machine learning model** that allows us to predict milk production data, quality indicators (protein and fat), and somatic cell count based on environmental data
- The model highlights the **long-term** effects that climatic variables can have on these parameters, while also emphasizing the **short-term** effects as previously reported (Biswal et al., 2019)
- Furthermore, the model enables the identification not only of the important variables but also of the specific days of interest
- These data can be used to construct more complex models and build **risk maps for specific areas or seasons**

Take-home message

Extending the model to different dairy cattle breeds to test species-level resilience with varying capacities to adapt to climate change

Optimizing the predictive model on specific areas (e.g. Italian regions) and developing risk maps for the upcoming years

Generating risk maps that can be useful for prediction and provide farmers with the ability to make proactive decisions

Thank you!

Smarter livEstock Breeding through Advanced Services Tailoring Innovative and multi-sourcE data to users' Needs

https://www.sebastien-project.eu

Centro Nazionale Agritech, finanziato dall'Unione Europea – NextGenerationEU

Spoke 5: Sustainable productivity and mitigation of environmental impact in livestock systems

Thank you!

Livestock Enviroment OpenData Project
https://opendata.leo-italy.eu/portale/home

https://highlanderproject.eu/

References

- Biswal, J., Kennady V., and Habibar R,. "Seasonal Variations and Its Impacts on Livestock Production Systems with a Special Reference to Dairy Animals: An Appraisal." Vet. Sci. Res 2 (2020): 24-30.
- Chandio, A. A., Nasereldin, Y. A., Trang Anh, D. L., Tang, Y., Sargani, G. R., & Zhang, H. (2022). The Impact of Technological Progress and Climate Change on Food Crop Production: Evidence from Sichuan—China. *International Journal of Environmental Research and Public Health*, *19*(16).
- Cheng, M., McCarl, B., & Fei, C. (2022). Climate change and livestock production: a literature review. *Atmosphere, 13(1*), 140.
- Maulu, S., Hasimuna, O. J., Haambiya, L. H., Monde, C., Musuka, C. G., Makorwa, T. H., ... & Nsekanabo, J. D. (2021). Climate change effects on aquaculture production: sustainability implications, mitigation, and adaptations. *Frontiers in Sustainable Food Systems*, *5*, 609097.
- Nasr, M. A., & El-Tarabany, M. S. (2017). Impact of three THI levels on somatic cell count, milk yield and composition of multiparous Holstein cows in a subtropical region. Journal of Thermal Biology, 64, 73-77.
- Sheikh, A.A., Bhagat, R., Islam, S.T., Dar, R.R., Sheikh, S.A., Wani, J.M., & Dogra, P. (2017). Effect of climate change on reproduction and milk production performance of livestock: A review. *Journal of Pharmacognosy and Phytochemistry, 6*, 2062-2064.
- Singh, P., Vaidya, M. K., & Pathania, K. (2021). Economic impact of climate change on agriculture: Present, past and future. *Journal of Pharmacognosy and Phytochemistry*, *10*(1S), 19-27.
- Toghdory, A., Ghoorchi, T., Asadi, M., Bokharaeian, M., Najafi, M., & Nejad, J. G. (2022). Effects of Environmental Temperature and Humidity on Milk Composition, Microbial Load, and Somatic Cells in Milk of Holstein Dairy Cows in the Northeast Regions of Iran. *Animals. 12*(18): 2484.
- Unanaonwi, O. E. (2010). The effects of climate change on food production. *Research Journal of Agriculture and Environmental Management*, *3*(11), 538-541.

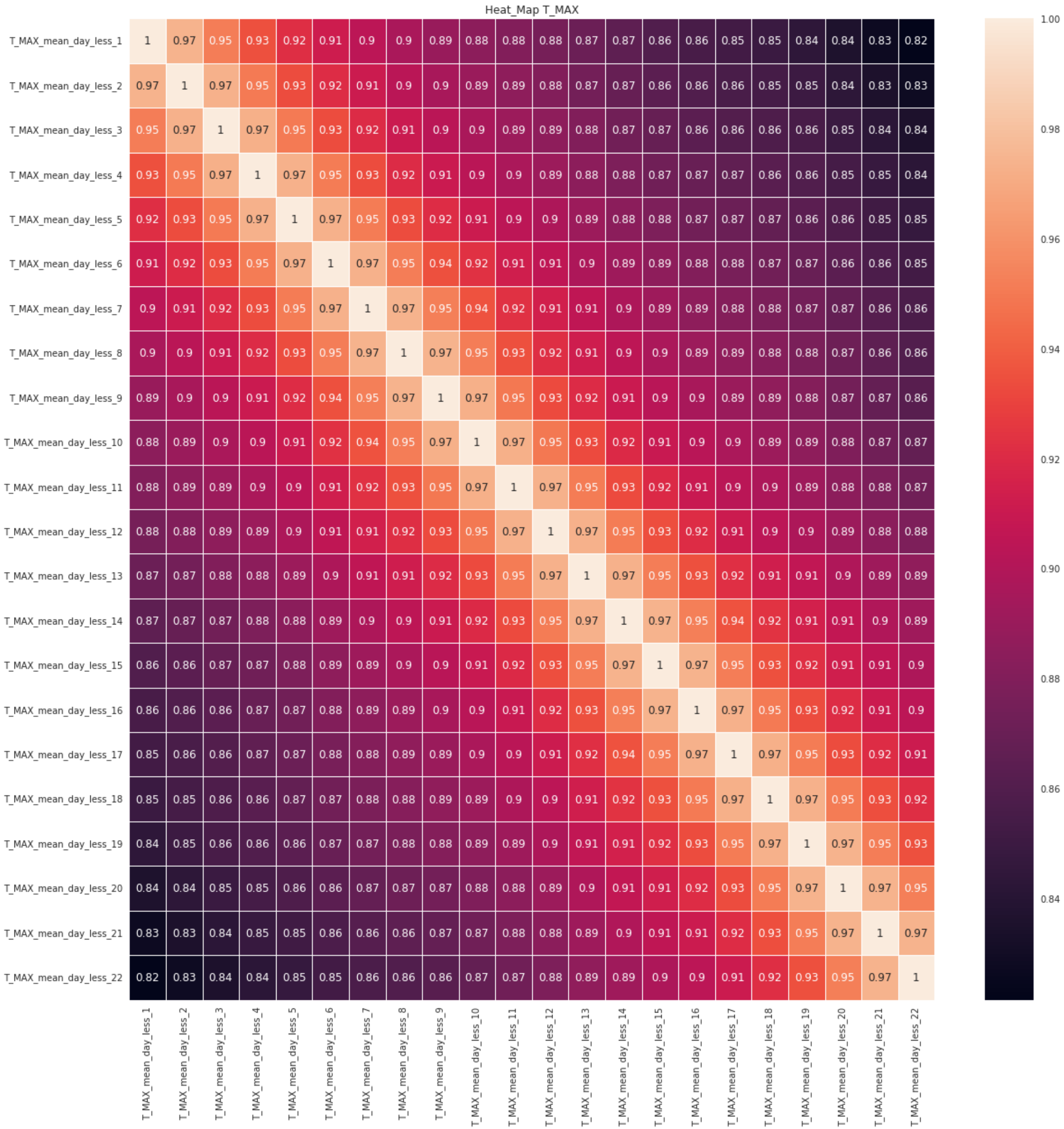
Methods
Collinear variables

- **Prior Feature Selection**
- Collinear variables were removed

- Only the most important and less informative features were maintained

- 50 variables were selected for temperature, humidity index and wind speed

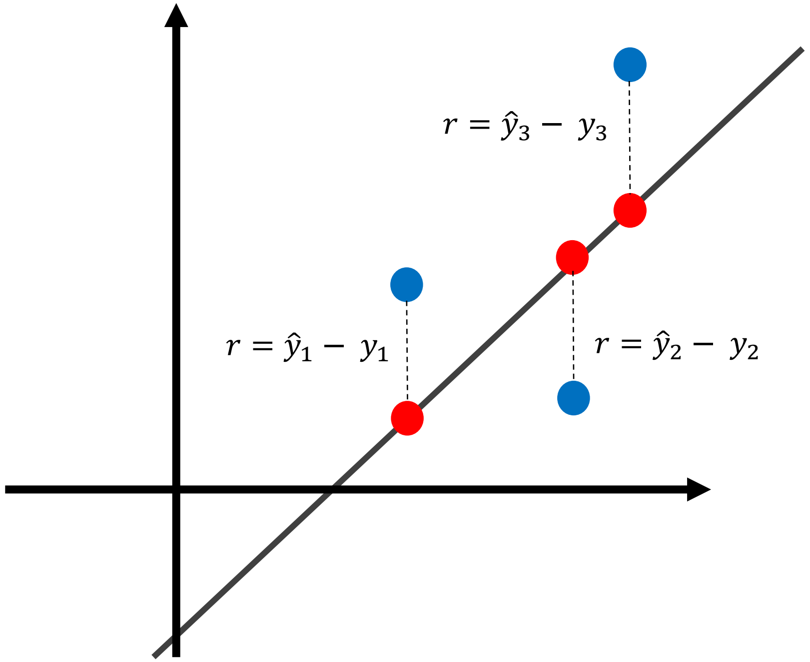
- 50 variables were selected for Discomfort Index



Methods
Linear model

Four phenotypic data were analyzed:

- Milk yield (kg/hg)
- Somatic Cell Counts
- Protein content (%)
- Fat content (%)



$$\underbrace{Phenotypic\ data = Parity + Days\ in\ Milk + Cow\ Age + EBV}_{} + Farm + Animal\ ID$$

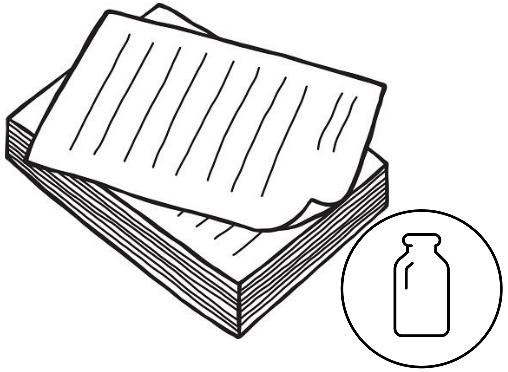
Fixed effects (Parity + Days in Milk + Cow Age + EBV) — *Random effects* (Farm + Animal ID)

- The residuals of the linear model were used to mitigate the effects of genetic factors or farm management
- They represent the data to be predicted for the artificial intelligence model

Methods
Data collection

- The “**<u>production</u>**” dataset: dairy cows’ records from 1990 to 2020 (AIA, ANAPRI)

- Pilot dataset: milk production from “Pezzata Rossa” in Friuli-Venezia-Giulia



Libro Genealogico Pezzata Rossa Italiana A.N.A.P.R.I. - Udine

Ente selezionatore
A.N.A.P.R.I.
Associazione Nazionale Allevatori Pezzata Rossa Italiana

Livestock Environment Opendata

A.I.A.

Methods
Our pipeline

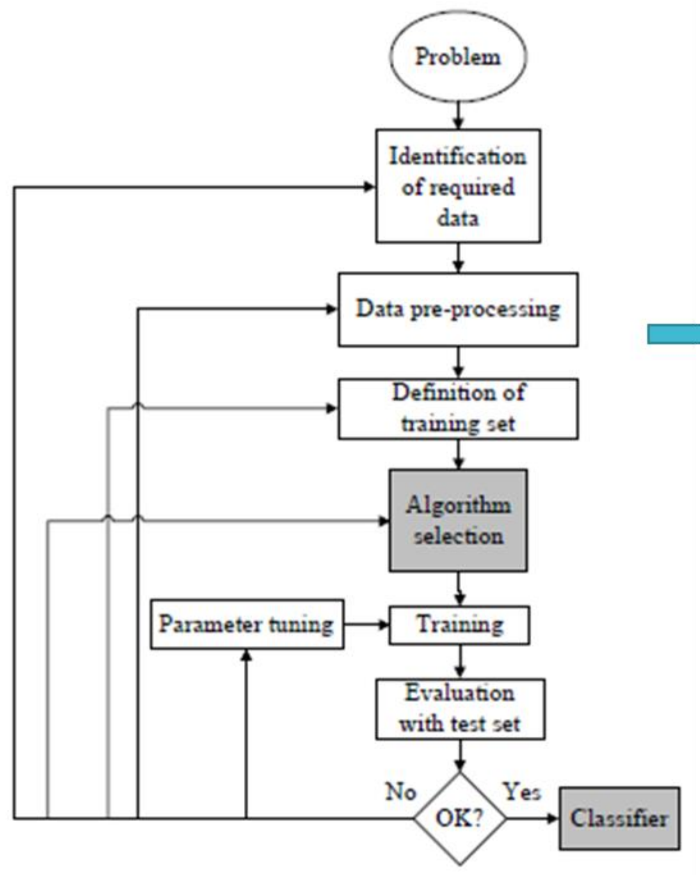


Figure 1. The process of supervised ML

Historical data