Here are the abstracts for each project proposal submitted. We encourage you to use these to decide your votes for first, second, and third place. Place your vote using the voting platform on the Hackathon website.

1. **Making better cheese: the hidden information in milk infrared spectra**

This proposal explores the potential of Fourier-transformed infrared spectroscopy, commonly used for basic milk assessments, to provide opportunities for cost-saving process improvements and quality control innovations. We show how the unique molecular fingerprints of different regional milks can be identified using artificial intelligence. We demonstrate this approach, both in general and also for specific provinces, such as those involved in the production of Grana Padano, one of the great cheeses of Italy. Our model, trained on 70,000 spectra and tested on 30,000, achieved 88% accuracy in differentiating milk from Grana Padano-producing provinces versus others. A second model reached 98% accuracy in identifying milk from the ten largest Italian milk-producing regions. Overall, this research highlights the utility of artificial intelligence in finding ways to save costs in cheese production, potentially eliminating the need for multiple tests when an existing easily sampled biofluid can offer much more information than currently used, and opening new avenues for quality improvement for Italy's cheese producers.

2. **Spatial patterns in milk quality profiles leveraging Principal Component Analysis and Multiple Imputation**

The Italian dairy sector is economically significant, supported by the production of several renowned protected designation of origin (PDO) cheeses. However, milk production conditions vary considerably across the country, influenced by both environmental and herd-related factors. This study leverages publicly available data on livestock environment in Italy to analyze bulk milk compositional data and coagulation properties. The data, collected across a large number of herds across most provinces, were considered representative of the respective provinces for each month. The study aimed to map the regional variability in milk composition and coagulation properties across Italy. Principal component analysis (PCA) was used to identify potential clustering of regions in terms of overall milk quality, reducing data dimensionality to enable comparative analysis. Missing data were imputed using the multiple imputation by chained equations (MICE) algorithm. Mapping the obtained principal component (PC) scores revealed consistent clustering of provinces across the months of analysis. The scores on the first three PCs were used to represent a unique RGB color, allowing for the mapping of these primary PCs in a single map, confirming the observed differences. The mapping of a single representation incorporating all months of analysis led to the clear identification of milk quality clusters, specifically in the Foggia, Matera, and Lucca provinces, as well as larger clusters in the eastern part of the country, the Campania

region, and the Sicily island. The analysis of PCA loadings showed that the major milk components (protein, fat, casein, and total solids) contributed similarly to the PC scores, while coagulation properties and cell counts had varying effects. Further studies are required to understand these trends and assess the statistical significance of the differences and the legitimacy of the identified clustering, which are an aggregate consequence of herd and environmental factors affecting milk production.

3. **Levering explainable artificial intelligence for efficient data-driven decisions to improve welfare in livestock species**

Here, we successfully employed Random Forest models for the prediction of monthly measures for five welfare indicators in dairy cattle (longevity, mastitis, subclinical ketosis, subclinical acidosis, and reproduction), applying as features the earliest previous monthly mean values of acetone, fat, lactose, protein, average conductivity, somatic cell counts, and urea in the milk (minimum Pearson correlation between predicted and observed values equal 0.974). The use of the predicted values to classify the observations in the good, intermediary, and risk overall welfare categories resulted in a precision of 100%, 89%, and 99%, respectively, when compared with the original classification. Additionally, explainable artificial intelligence (XAI) tools were employed to lever the interpretability of the obtained models in order to provide potential management decisions that could improve the welfare score of each animal in the next evaluation.

4. **Smart Weather Health Prediction for Livestock: AI-Enhanced Milk Protein Forecasting Using Gradient Boosted Models**

The dairy industry is under increasing pressure from both governments and society to provide dairy products at competitive prices, while also ensuring stringent environmental stewardship. Weather conditions significantly affect milk characteristics, but there is limited information available on how it impacts milk protein fractions. Hence, there is a need to understand how weather conditions affect changes in milk proteins as this will help in adjusting and developing strategies to mitigate the negative effect of weather conditions on milk quality. This project aims to combine weather conditions and milk performance recording data for milk protein prediction. A Gradient Boosting Regressor was used in developing the model and achieved a Mean Squared Error of 0.021 and an R-squared of 0.91, highlighting its strong predictive capability. This project showcased the value of AI in Agric monitoring.

5. **Impact of Thermal Stress on the Reproductive Performance of italian dairy cattle: An Analysis with Machine Learning Techniques**

Heat stress is one of the main factors in the dairy industry that can negatively impact milk production, health, and reproductive performance in dairy cows. This study aims to assess the influence of environmental stress on the reproductive performance of Holstein cows. Two key parameters were examined: the Temperature-Humidity Index and Days Open, which is the average number of days from calving to conception, for the period from January 2017 to December 2023 in Italy. To analyse the effect of climate

conditions on Days Open and, consequently, on conception probability, this study employs Machine Learning techniques based on data related to reproductive activity and climate conditions. We developed a Machine Learning-based pipeline designed to analyse the relationship between climatic conditions and Days Open. The Machine Learning approach provides insights into the role of Temperature-Humidity Index as a predictor of reproductive outcomes, suggesting that medium to long-term Temperature-Humidity Index measures may have a greater impact on predicting Days Open compared to short-term values. This finding highlights the importance of considering extended periods of heat stress when developing strategies for managing fertility in dairy cattle, thereby contributing to improved reproductive management in environments with varying climatic conditions.

6. **A Reporting System for Small-Scale Farms: Leveraging Open Data for Grigio Alpina Cattle in Italy**

The digital transformation of agriculture promises sustainability but often excludes small-scale farms due to corporate monopolies and data lock-ins. These barriers limit access to digital tools and reinforce dependence on commercial models. To address these challenges, we developed a digital reporting system for Grigio Alpina cattle in Italy, using open data from the Livestock Environment Opendata (LEO). The system analyzes key lifecycle events, milk composition and cheesemaking properties, health risks, and reproduction traits. Unsupervised learning techniques are used to assess each animal's overall productivity compared to the broader population, offering farmers an intuitive way to explore how their animals perform relative to others. This interactive tool offers visual insights and comparative statistics, empowering farmers to monitor animal health and improve their selection for further generations. Our work takes a step towards breaking data lock-ins among large corporations, promoting transparency and accessibility of information, and fostering sustainable and resilient farming systems.