

Title: Data integration and processing platform using data science and machine learning techniques for real time prediction of dairy cattle health and welfare changes

Team: ATB: T. Amon T. Kabelitz , O. Kashongwe
UOS: M. Atzmueller , T. Römer ,
HS: H. Tapken

Abstract:

Dairy farms collect increasingly large amounts of data to improve animal performance and to prevent diseases. Various data science and/or machine-learning techniques have been used to predict infections and performance of dairy with more success in theory (offline datasets) than in practice. The status on disease detection on farms is that, despite the abundance of devices used, too many cases of sickness or stress are diagnosed too late or not at all and with a large number of false alarms. This leads to waste of time and money for farmers. At the center of this detection problem is the data recording system and the integration strategy. The data recording systems by design and specific purpose present several challenges for effective use due to varying data granularity to make decisions. Major challenges include the variety of data types/ formats and the different time dimensionality ranging from 10 minutes to one-month intervals. Recorded are milk production, disease alarms, barn climate and image-based sensor data monitoring cow behavior and welfare. The aim of the project is to develop and/or apply relevant methodologies to combine data recorded by various sensors into a hub for data analysis, modeling and support for decision-making. Current data science techniques will be employed to provide timely and high quality data, and subject them to best prediction models for (near to) real-time prediction of disease occurrence for individual animals, towards supporting farmers' decision-making through digital solutions. The project is expected to produce a PhD thesis in data science with focus on farm animal data integration and modeling.

Desired skills of the applicant:

Applicants for this position should have good knowledge of data engineering and experience working with large datasets in the agricultural sector, especially livestock. Skills in database management, data pipeline architectures and machine learning modelling are preferred. A computer / data science degree or a dairy science degree with strong interest and skills in data science would be an advantage.

References:

- Kashongwe B.O., Kabelitz, T., Ammon, C. Doherr, M., Amon, T. & Amon B (2023). Influence of automated milking systems data preprocessing methods on the prediction of mastitis occurrence with machine learning models. *Journal of Big Data* (under review).
- Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, 100367-100377.
- Ryan, C., Guéret, C., Berry, D., Corcoran, M., Keane, M. T., & Mac Namee, B. (2021). Predicting illness for a sustainable dairy agriculture: predicting and explaining the onset of mastitis in dairy cows. arXiv preprint arXiv:2101.02188.
- Van der Voort, M., Jensen, D., Kamphuis, C., Athanasiadis, I. N., De Vries, A., & Hogeveen, H. (2021). Invited review: Toward a common language in data-driven mastitis detection research. *Journal of Dairy Science*, 104(10), 10449-10461.