Term Project Proposal

Analysis of dairy cow milk composition correlation to fertility

FARM - **F**ertility **A**ssessment **R**elying on **M**ilk data

Bioinformatics - Spring 2022

Team Members:

Chamudi Abeysiriwardana

Graduate Student in Computer Science with focus on Data Science and Machine Learning.

Email: abeysck@uri.edu Phone: (401) 212-7410

Denis Chernov

Undergraduate Student in Pharmaceutical Sciences focusing on laboratorial analyzing and biotechnology.

Email: denis_chernov@uri.edu Phone: (401) 537-8730

Lauren Decastro

Masters student in Cellular and Molecular Biology with a focus on fertility in cows and the impacts of metabolic disease on fertility.

Email: lauren_decastro@uri.edu Phone: (781) 361-1604

Daniel Marasco

Industrial and systems engineering senior with background in statistical process control and process simulation with programming experience in automation and data acquisition. One-time 1st place cow milker, Washington County Fair.

Email: danm95@uri.edu Phone: (401) 300-0598

Adriana Servant

Masters Student in Pharmaceutical Science with a focus in medicinal chemistry and pharmacognosy.

Email: adriana_servant@uri.edu Phone: (401) 871-0263

Abstract

The reproductive health and milk productivity of a dairy cow are of high importance to dairy farmers in order for them to maintain an optimal herd size as well as milk production. Both reproduction and lactation are affected by the overall health of an animal, from environmental factors to aging. Milk composition has been utilized in prior work to identify disease processes in cows, and the link between health and fertility is well known. Our team intends to use data on milk composition to inform forecasting on the future fertility of individual animals.

We hope to use regression analysis, which is a fundamental concept in the field of machine learning, to predict the fertility of a dairy cow using milk composition. Regression analysis falls under supervised learning where the algorithm is trained with both input features and output labels, which helps in establishing a relationship among the variables by estimating how one variable affects the other. From the scientific perspective of this project, we will compare the hormone levels present in the milk, as well as fat and protein content of the milk, with other factors collected from the data, with the hopes of drawing conclusions about how the fertility of the cows could be affected based on any outside variables.

Hypothesis

Milk composition (fat and protein content) can be used as predictive variables for fertility in dairy cows.

Goals

- Identify the relationship between milk composition and fertility.
- Develop a machine learning model using regression analysis

> Develop a machine learning model using regression analysis to predict the fertility of a dairy cow using milk composition.
>
> *Development Environment*
> The development environment used for this project was a Linux based system supporting Keras. Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation.

- Produce a guide (printable or electronic) that can be used by dairy farms to identify animals that are likely to be infertile based on the results of a milk composition test. This could potentially be integrated into a herd management solution like https://bovisync.com/

Background

Fertility in dairy cows is a very complex trait that is impacted by many factors. These factors can include genetics, environment, and nutrition. These variables can be seen when comparing dairy cows at different levels of lactation under different management schemes. Cows in higher production settings may have better fertility due to the better nutrition that they are provided with. This can change, as milk production increases during the postpartum period. Cows are often unable to fully meet the increasing nutritional demands of milk production through diet alone, and will begin to mobilize fat. During this time, the cows are prone to an increased risk for varying health problems such as metabolic disease, decreased immune function, and decreased fertility.

The knowledge of the link between the health and nutritional status of dairy cows has been used in past studies to monitor cow health, in which milk composition data was studied. Milk composition data is often collected routinely using mid-infrared spectroscopy. This data greatly reflects the nutritional status of an individual cow, because when she mobilizes fat, the fat to protein ratio in her milk increases. This ratio has been used to train models that predict health-related issues such as mastitis and lameness. It was shown that cows with a fat-to-protein ratio greater than 2.0 were at a higher risk for general health complications. Knowing the presence of a link between nutrition and fertility that is similar to the link between nutrition and general health status means that it would be feasible to use this same composition data to predict fertility levels in dairy cows. This past study was run using a variety of machine learning methods that we believe may also be useful in the case of this project (Contla Hernandez).

It may also be helpful to establish the impact that other health events may have on fertility. Fertility can be impacted by a range of health events that may happen at a higher frequency, in correlation with the poor or depleting nutrition of a cow. Another potential health conflict that could affect fertility involves any medical treatment provided to a cow. Depending on the severity of a health condition, antibiotics or other pharmaceuticals may be provided for dairy cows, which may have a short or long-lasting impact on their hormone levels, which in turn could affect their fertility. By establishing the magnitude of impact that any of these health events have on fertility, it may provide us with the knowledge to develop a method to train a more accurate and precise model.

The primary tools for this project will be the R programming environment and Python. R was designed for statistical analysis, and will be appropriate for visualizing the data, analyzing the correlation, and testing the significance of the hypothesis. R may also be used as a data

and literature search tool through the R-Entrez API. Python may be used as a tool to implement machine learning to perform some of the analysis and graphics. Python with Keras will be used for machine learning tasks that will be specified further in the future, depending on the data that becomes available.

Project Plan

Literature review

- Ongoing

Internal Reports - Sharing subject knowledge

- Milk Composition
- Factors in Fertility
- Machine learning regression analysis

Data collection/aggregation and preparation

Analysis - machine learning and statistical analysis

Documentation

Documentation will include continuous work throughout all parts of the project.

Final Report

Timeline

February

- Complete proposal - All members
- Literature reviews - All members

March

- Internal reports
    - Milk composition and dairy overview - Lauren
    - Fertility - Lauren, Denis, Adriana
        - Each can present on a different viewpoint of fertility

        - Ex: Lauren discusses the cow aspect of fertility, Adriana discusses the links to humans / similarities and differences, Denis discusses the pharmaceutical effects that could impact fertility
    - Statistical analysis - Dan
    - Machine Learning - Chamudi

- Compiling and cleaning data - cursory analysis
- Search for additional data sources - All members

April

- Develop scripts for analysis - Chamudi and Dan
- Review results - All members
- Document for final report - All members

May

- May 4 - Final Report Due & Presentation Draft
- May 11 - Presentation (during final exam period)

Team member administrative functions:

- Lauren - Team Leader : Guides the high-level direction of the project. Resident cow expert.
- Daniel - Meeting organizer: plans meetings and keeps the team on track with deliverables. Will do the statistical analysis with R programming, verify significance of results.
- Chamudi - will do the machine learning using python and update on github
- Denis - pharmaceutical science - can provide general science background as well as specific information regarding fertility, hormones, potential factors that affect fertility
- Adriana - able to write quickly and efficiently and will edit documents and assignments before submission; pharmaceutical science - can provide general science background as well as specific information regarding fertility, hormones, potential factors that affect fertility

Outside Means of Research:

- For Data:
    - Farms (x2) - Lauren will be in touch through program personnel
- Contacts:
    - Dr. Natallia Katenka (statistics professor at URI)
    - Dr. Maria Hoffman (contact for dairy farms)

Team Goal: *To produce a printable or electronic guide that dairy farms can use to identify animals that are likely to be fertile based on the results of a milk composition test.*

Weekly Meeting Time:

- Thursdays: 9:00 - 10:00am

Necessary Resources for Project Completion:

- Farm Data
- Python and Keras
- Literature
- Mendeley: reference manager

Biology section of paper:

- Primarily written and edited by Adriana, Denis, and Lauren

Computer Science section of paper:

- Primarily written and edited by Chamudi and Daniel

Bioinformatics integrated section of paper:

- Written and edited by all group members

Aranciaga, N., J. D. Morton, D. K. Berg, and J. L. Gathercole. 2020. Proteomics and metabolomics in cow fertility: a systematic review. Reproduction. 160:639–658. doi:10.1530/REP-20-0047. Available from: https://rep.bioscientifica.com/view/journals/rep/160/5/REP-20-0047.xml

Beever, D. E. 2006. The impact of controlled nutrition during the dry period on dairy cow health, fertility and performance. Anim. Reprod. Sci. 96:212–226. doi:10.1016/J.ANIREPROSCI.2006.08.002.

Contla Hernández B, Lopez-Villalobos N, Vignes M. Identifying Health Status in Grazing Dairy Cows from Milk Mid-Infrared Spectroscopy by Using Machine Learning Methods. *Animals (Basel)*. 2021;11(8):2154. Published 2021 Jul 21. doi:10.3390/ani11082154

Ouweltjes, W., B. Beerda, J. J. Windig, M. P. L. Calus, and R. F. Veerkamp. 2007. Effects of Management and Genetics on Udder Health and Milk Composition in Dairy Cows. J. Dairy Sci. 90:229–238. doi:10.3168/JDS.S0022-0302(07)72624-3.

Walsh, S. W., E. J. Williams, and A. C. O. Evans. 2011. A review of the causes of poor fertility in high milk producing dairy cows. Anim. Reprod. Sci. 123:127–138. doi:10.1016/J.ANIREPROSCI.2010.12.001.