AMS Forage Project Viability Report

Thomas J FitzGerald ([thomas-fitzgerald@outlook.com](mailto:thomas-fitzgerald@outlook.com))

Brief Summary

A machine learning model making predictions about feed composition using existing milking system data shows some promise, but training such a model requires additional feed sample data.

Summary

Using data from Automated Milking Systems (AMS) at a test farm, we attempted to predict changes in feed composition given to the cows. Linear regression models developed as a baseline had very little ability to explain variance in feed composition (R^2 <.003), so we focused on a neural network method.

Daily milking activity would be “polled” by the model, forming a distribution. Ideally, results from a trained model would form a roughly normal distribution with a mean similar to the actual nutrient value of the feed. This estimate could then be used to adjust feed levels.

Unfortunately, we had difficulty obtaining enough feed samples to adequately train a model, but we were able to perform limited tests for viability. The plot below shows a typical example of results from these tests. With more data, a similar method has potential to either give a reasonable prediction of certain nutrient levels, or differentiate between certain cutoffs (Lignin > 3.30 vs. Lignin <3.30, for example).

A diagram of a model prediction

Description automatically generatedA diagram of a model prediction

Description automatically generated

Other nutrient measurements show little promise, like aNDFom on the right.

The data cleaning & preprocessing pipeline is already developed, and should require minimal adjustment to re-use once more feed data is available. If this plan goes forward, I suggest the following, in order of importance:

1. More feed samples. 30+ would be a good starting point, but more is better.
2. Additional input variables (protein/butterfat/energy corrected milk values).
3. Cow weight.

**Project Goals**

Providing sufficient nutrients to dairy cattle is important for productivity, so underfeeding is undesirable. However, feed is often 80-90% of the expense on a dairy farm, so overfeeding is also undesirable. Nutrient composition of feed changes regularly, even from the same provider, so a rapid way to alert dairy farmers to changes would be beneficial.

The original goal of this project was to take data from Automated Milking Systems (AMS) and detect changes in feed quality. Previous research had indicated that changes in fiber content resulted in a measurable change in rumination time.

**Data**

The study was mainly conducted using data from Farm 1 (Hickory Lawn Dairy, WI) between 2023/02/24 and 2023/11/06. Farm 1 used Lely A5 milking robots, and had 950 cows during this period.

Data Dimensionality

|  |  |  |
| --- | --- | --- |
| Description | Rows | Columns |
| Milking | 704984 | 17 |
| Rumination | 173771 | 14 |
| Alert | 49274 | 6 |

Each interaction with the AMS device produced a line of “Milking” data.

This was supplemented with Rumination data, which included daily summary information about each cow’s time spent Ruminating, and the amount of feed dispensed from each robot to that cow.

We were also supplied with Device Indications (Alert) data from the AMS itself, which included information about errors, cleaning cycles, and other issues related to the AMS. Although this data may be useful to control for changes in cow behavior due to system outages (for example, a cow that prefers a particular machine

The above 3 datasets were combined into a single dataset which was used for analysis. A secondary pipeline converted this for use in model training.

Pipelines were also developed for Farm 2 (Lepeska Dairy) and Farm 3 (Hinchley Dairy). Farms 1 & 3 used the same type of system (Lely A5), while Farm 2 used (DeLaval V300). Because Farm 2 dropped out part-way through and Farm 3 lack sufficient feed samples, these are incomplete.

**Methods**

Preprocessing was performed in Python 3.10, primarily using Pandas, and models training was performed using pytorch. EDA and other analysis was conducted using RStudio Build 576, R version 4.2.2.

**Results**

Although the original intent was to use milking events as a time series and use an LSTM model, I was unable to get an LSTM to train properly. Some of this may have been due to memory constraints, since the viability test was performed on a laptop CPU, and there were a number of memory-related crashes when attempting to train an LSTM. Outside of memory issues, there were also exploding gradient problems on other set-ups. However, even when a training cycle completed successfully, the model frequently failed to train, returning a flat, wildly inaccurate result, regardless of input.

To make the LSTM input work with the memory constraints, I had reduced the input to the first 10 milking events each day, or approximately 150 values. I tried training models with a series of linear layers using this input, and had better success than with the LSTM setup. These results make up the bulk of what can be found in model\_testing/.

Lastly, I tried using a daily summary of milking results, instead of the event-based input. These models were unable to learn effectively, and returned close to the mean response value for the dataset regardless of the input given.

**Next Steps**

Obtaining more feed samples is imperative to making this project work. This would help on two fronts:

1. ML models better identify nuanced relationships with more examples.
2. More samples will reduce risk of overtraining.

Additional input variables, in particular macronutrient data of milk (lactose/protein/butterfat/energy corrected milk values) would likely help the model to make more accurate predictions.

Cow bodyweight has a strong effect on milk yield, but is currently not available in our dataset. Even an out-of-date value, like an annual weigh-in, would help the model to control for this somewhat, and may improve accuracy.

The date component of date/time should be removed. The models tested so far were not powerful enough to do so, but there is a direct correlation between sample dates and feed value. The time component should be left in.

On the more unrealistic side, some way to gauge intake of non-PMR food sources for each cow would be helpful. As an example, we’re currently trying to estimate the crude protein (CP) content of the PMR being provided to the cattle. Even in cases where we have CP values for multiple feed sources (PMR, Haylage, corn, etc.), we only know the total intake of PMR. Two cows could have the same PMR intake, but one could eat double the amount of haylage, increasing it’s total CP intake. That said, I have no practical suggestions for accomplishing this.

Although the alarm data did not make it into the final test models, it may be useful in a supplemental model. The current model is based on each cow’s interaction with the AMS, so controlling for changes on the AMS side, like downtime for cleaning, may help with accuracy. This is a low priority, and should only be examined if an otherwise functional model can be developed.