HOCHSCHULE
RAVENSBURG-WEINGARTEN
UNIVERSITY
OF APPLIED SCIENCES

RAVENSBURG-WEINGARTEN UNIVERSITY

POTENTIALS OF ROUTINELY COLLECTED DAIRY FARM DATA

Ruben Miller

Research Associate at the Institute for Artificial Intelligence

Motivation

Dairy farming routinely produces a lot of data.

Can we use this data?

→ **We aim to find potential and create benefits for farmers using AI**

Project Overview

What are we researching?

Together with experts of different field, different Research Questions were created, these:

- Should create a benefit for the farmer (or the cow)
- Should be able to be answered with the current data

Two we already researched deeper :

- Predicting the milk yield
- Predicting how the calving process will unfold

Our Data
What data do we have?

- Data from 205 farms in Germany, Baden-Württemberg
- Ranging from 2017 to today
- Focus on Holstein, Brown-Swiss and Fleckvieh

We have a lot of different sources:

- MLP data (monthly milk yield and content)
- Breeding values
- Pedigree information
- Clinical diagnoses
- And more …

Milk Yield

Prediction of milk yield

What exactly do we predict, two sub-tasks:

- Predict the total 305 day yield of a cow for the next lactation
- Predict the first 305 day yield of a calf

What data do we use:

- Data of the last lactation of the cow
- Pedigree information, data on mother or older siblings

Not all data is available for all cows.

Milk Yield
Training the models

For training, different models were used:

- XGB
- LGBM
- SVM Regressor (scikit)
- TabNet

These where trained with a different amount of included features

Milk Yield
Key results

Additional features did not compensate for reduced sample size set.

- For example the addition of breeding values

Predictions of the first lactation yield of a calf were bad

- We do not have much early life information
- The time frame is very high (> 2.5 years)

Our prediction of the next lactation reached 0.59 R^2

- The most important feature was the last lactation yield
- This is not precise, but can give farmers a good idea

Milk Yield
Key results

Tree based models were much better than deep learning

- We do have a small dataset
- Tree based models generally perform very good on tabular data
- SVM Regression is similar in performance, but less stable

Calving

Prediction how the calving process will unfold

We want to predict the calving process:

- Bad calving process defined by veterinarian

We predict at time of insemination

- This gives the farmer the option to react to the prediction

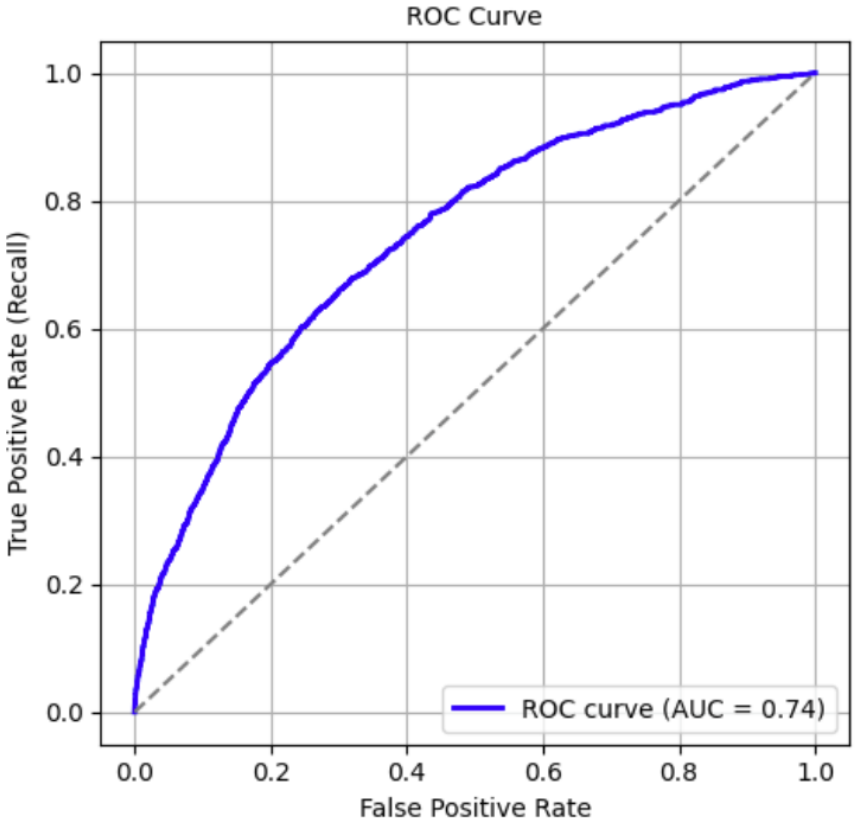
Calving
Key results

We tested LGBM, XGB and random forest

- LGBM best with AUC 0.74

The most important feature are farm stats:

- In a farm with lots of bad calvings, there will be more



Conclusion

There is potential in the already collected data

- The farmer can get an idea of the yield of their cow in the next lactation
- We can give a warning when a calving may have a non optimal process

Tree based models outperform others on our tabular data

Future Work

What will and could be done in the future?

We are currently working on more research questions:

- Predicting ketosis
- Predicting success of insemination

HOCHSCHULE
RAVENSBURG-WEINGARTEN
UNIVERSITY
OF APPLIED SCIENCES

THANK YOU FOR LISTENING!

Doggenriedstraße 70
88250 Weingarten

Postfach 3022
88216 Weingarten

www.rwu.de
info@rwu.de

HOCHSCHULE
RAVENSBURG-WEINGARTEN
UNIVERSITY
OF APPLIED SCIENCES

THANK YOU TO OUR PARTNERS!

Kofinanziert von der Europäischen Union

PARTICIPATING IN eip-agri

Gefördert durch Baden-Württemberg Ministerium für Ernährung, Ländlichen Raum und Verbraucherschutz

LAZ BW Landwirtschaftliches Zentrum Baden-Württemberg
Rinderhaltung • Grünlandwirtschaft • Milchwirtschaft • Wild • Fische-

LGL Geodaten und Landentwicklung - landesweit und amtlich
Unser Auftrag. Unsere Leidenschaft.

FBF

LKV Baden-Württemberg

RBW

UNIVERSITÄT HOHENHEIM

Used Models

Model	XGB	LGBM	SVR	TabNet
Description	• Tree based • Gradient Boosting • Very fast	• Tree Based • Gradient Boosting • Very fast	• Support Vector Machine with Regression • Scales poorly with dataset size	• Deep learning specialized for tabular data • Usually better for bigger datasets

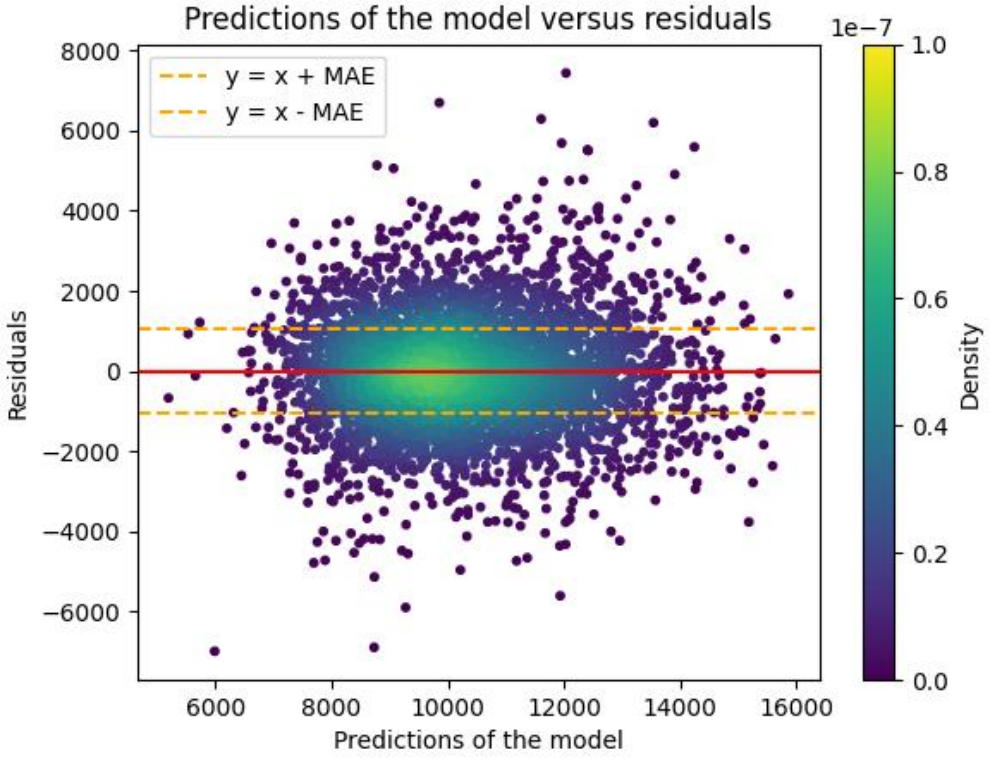
Detailed Results

Details of predicting next lactation

		XGB		LGBM		SVM	
	no. pairs	MAE	R^2	MAE	R^2	MAE	R^2
base case	16 412	1031±9.1	**0.58**±0.006	1029±9.9	**0.59**±0.006	1025±9.7	**0.59**±0.006
+ BVs for SBT	2959	1174±16.4	0.40±0.018	1166±18.5	0.41±0.019	1145±15.7	0.42±0.020
+ BVs for BV	1916	957±26.6	0.57±0.023	956±21.6	0.57±0.017	948±22.6	0.57±0.023
+ BVs for FV	2280	1027±19.1	0.37±0.019	1018±15.9	0.38±0.023	1019±18.0	0.39±0.028
+ fitted curves	2570	1096±18.9	0.55±0.016	1087±21.0	0.56±0.018	1055±21.6	0.58±0.021
base case on reduced data split of…							
…BVs for BV	1916	**940**±78.6	**0.58**±0.023	**937**±22.9	0.58±0.023	**932**±24.5	0.58±0.017
…fitted curves	2570	1103±19.9	0.55±0.032	1092±26.3	0.56±0.017	1062±25.4	0.58±0.021

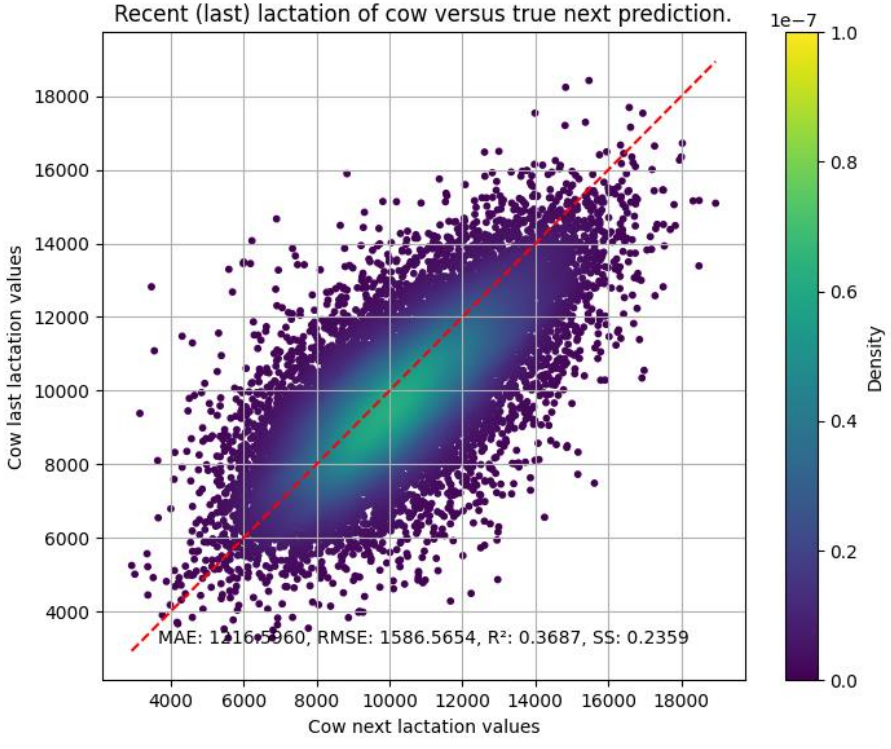
Detailed Results

Last lactation task



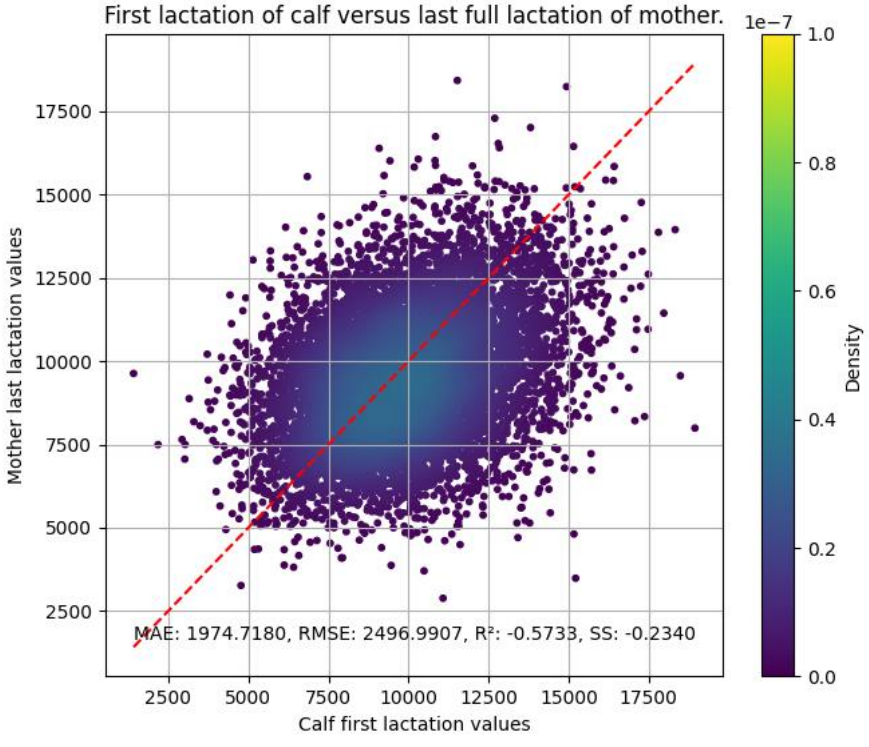
Detailed Results

Correlation last and next lactation



Detailed Results

Mothers lactation and fist lactation of calf



Training history

Of train run of XGB for next lactation task