Machine Learning Approach for Early Prediction of 305-Day Standard Lactation

Jakob Ganitzer, Hermann Schwarzenbacher, Christian Fuerst

ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria

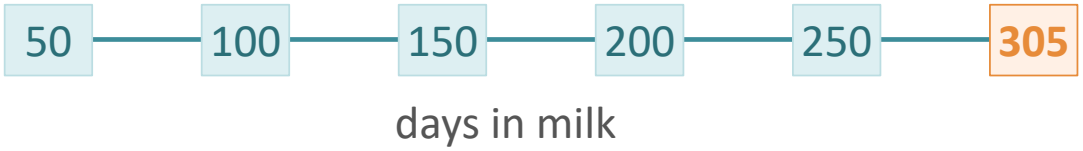
Problem and objective

305-day yield is a key benchmark, but it is only known after lactation is complete.

- Projections from early days in milk are needed
- This study compares a XGBoost model with a regression approach

Practical objective

Estimate the final 305-day standard lactation
from initial test days to support early decisions



Data and validation

169,012	170,598	5	Fleckvieh
lactations for training	lactations in the validation set	selected test days	breed in this validation

Data & Target

- standardized 305-day milk yield in kg as prediction target
- input records represented at daily resolution from test-day and cumulative information

Split

- training: records from 2023
- test/validation: records from 2024

Final comparison at days in milk: 50, 100, 150, 200, 250

Feature set

Feature group	Examples used in the model
Production	current test-day milk yield; cumulative milk yield
Herd context	rolling herd milk-yield average
Genetics	EBV for milk yield; EBV for persistency
Health / status	somatic cell count; lactation number; region
Timing / season	days in milk; calving and control-date sine/cosine encodings
Animal history	previous two milk and SCC measurements; days since last control

Animal and herd identifiers were not supplied as model features.

Model comparison

Regression baseline approach	XGBoost model
Separate equations for each breed, lactation and lactation day.	One gradient-boosted tree model.
Fixed/covariate effects: calving month, cumulative yield, herd average, calving age, milk-kg, EBV, persistency EBV; linear and quadratic terms.	Same core production, herd and genetic information as regression, extended with lag variables and cyclic sine/cosine time encodings.
Requires regular derivation of factors and coefficients.	Trained once, saved, then used for fast inference on new data + continuous training

XGBoost training setup

- learning rate 0.01, maximum depth 10, up to 5,000 trees
- early stopping after 40 rounds; RMSE validation metric
- Nvidia A6000 GPU
- Test time: 4,07 ms per 1000 animals

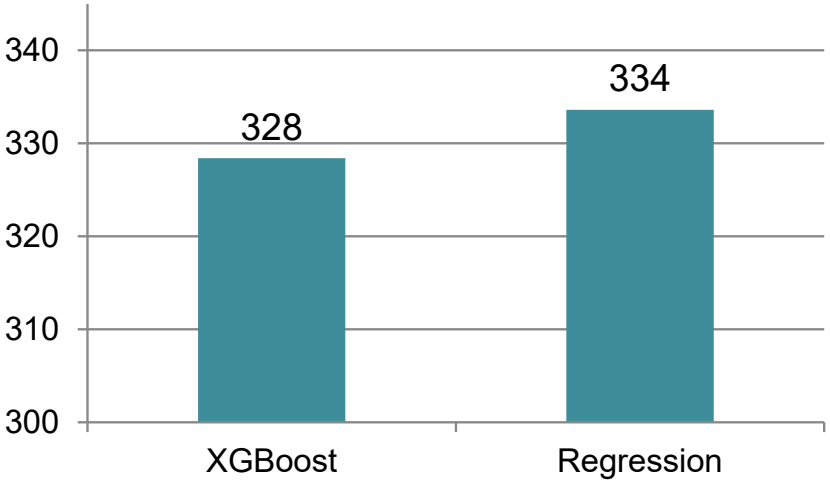
Overall accuracy: small but real gain

328.4 kg
XGBoost Mean Absolute Error (MAE) across days 50–250

333.6 kg
Regression MAE across days 50–250

5.2 kg
absolute MAE advantage

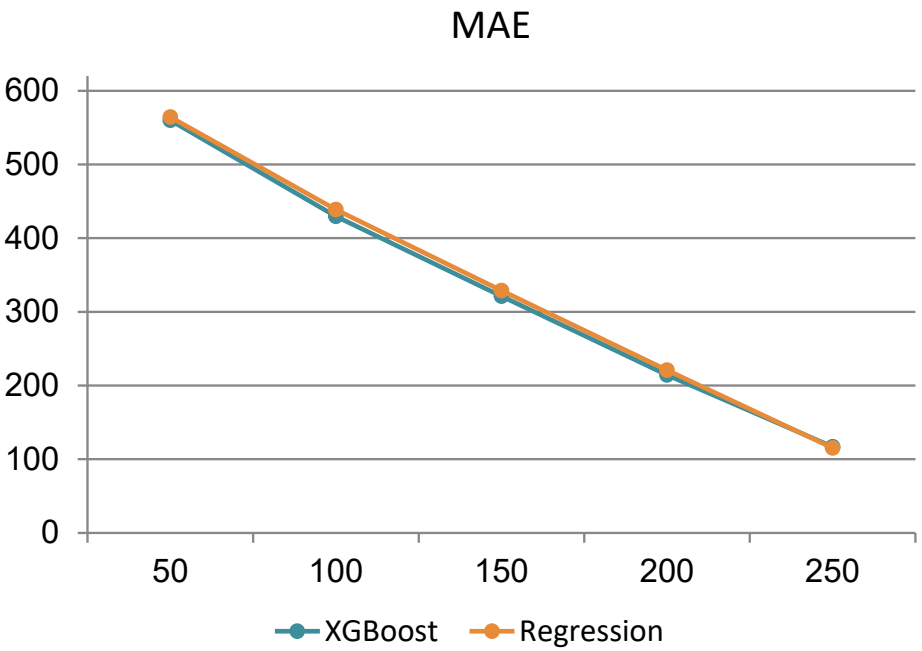
1.6 %
relative MAE advantage

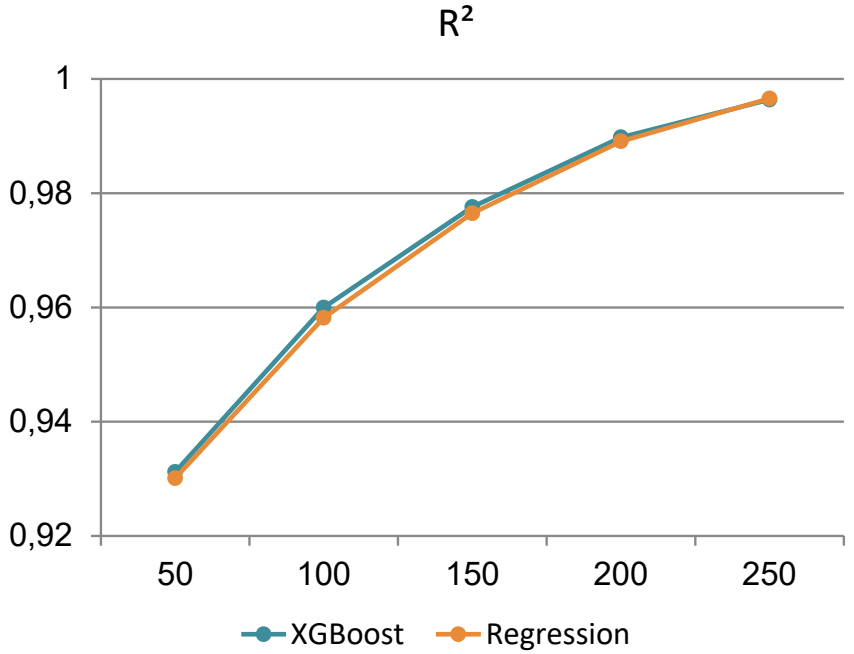


Same general accuracy level

- R^2 = 0.94 for both approaches.
- Pearson r = 0.9711 for XGBoost vs 0.9703 for regression.
- Mean error was similar: 41.5 kg vs 40.8 kg, indicating small average underprediction in both models.

Accuracy improves as lactation information accumulates





XGBoost is better from day 50 to day 200; regression is slightly better at day 250.

Where XGBoost helps most

Days in milk	XGBoost MAE	Regression MAE	MAE advantage
50	559.9 kg	564.3 kg	+4.4 kg
100	429.7 kg	438.8 kg	+9.1 kg
150	321.3 kg	328.8 kg	+7.5 kg
200	214.4 kg	220.9 kg	+6.5 kg
250	116.8 kg	115.2 kg	−1.6 kg

Largest reduction at day 100

–9.1 kg MAE versus regression

- Early-lactation projections are most relevant for management action.
- The numerical gain is not large, so robustness and operational fit matter.
- Late in lactation, both methods approach the final outcome closely.

Feature Importance Interpretation

1	**Current milk yield**	strong direct signal of current lactation level
2	**Rolling herd average**	captures herd environment and management context
3	**Test day / days in milk**	anchors the lactation stage
4	**Milk-yield EBV**	adds genetic expectation
5	**Cumulative yield**	summarizes observed production to date

Main takeaway

The model is not driven by a single input. It combines routine milk-recording data with herd and genetic context.

Operational use, limits and next steps

Why XGBoost is attractive operationally

- One flexible model instead of maintaining many regression equations.
- Can incorporate nonlinear effects, seasonal encodings and recent animal history.
- Fast inference; test run: 1,000 animals in 4.07 ms.

Current limits

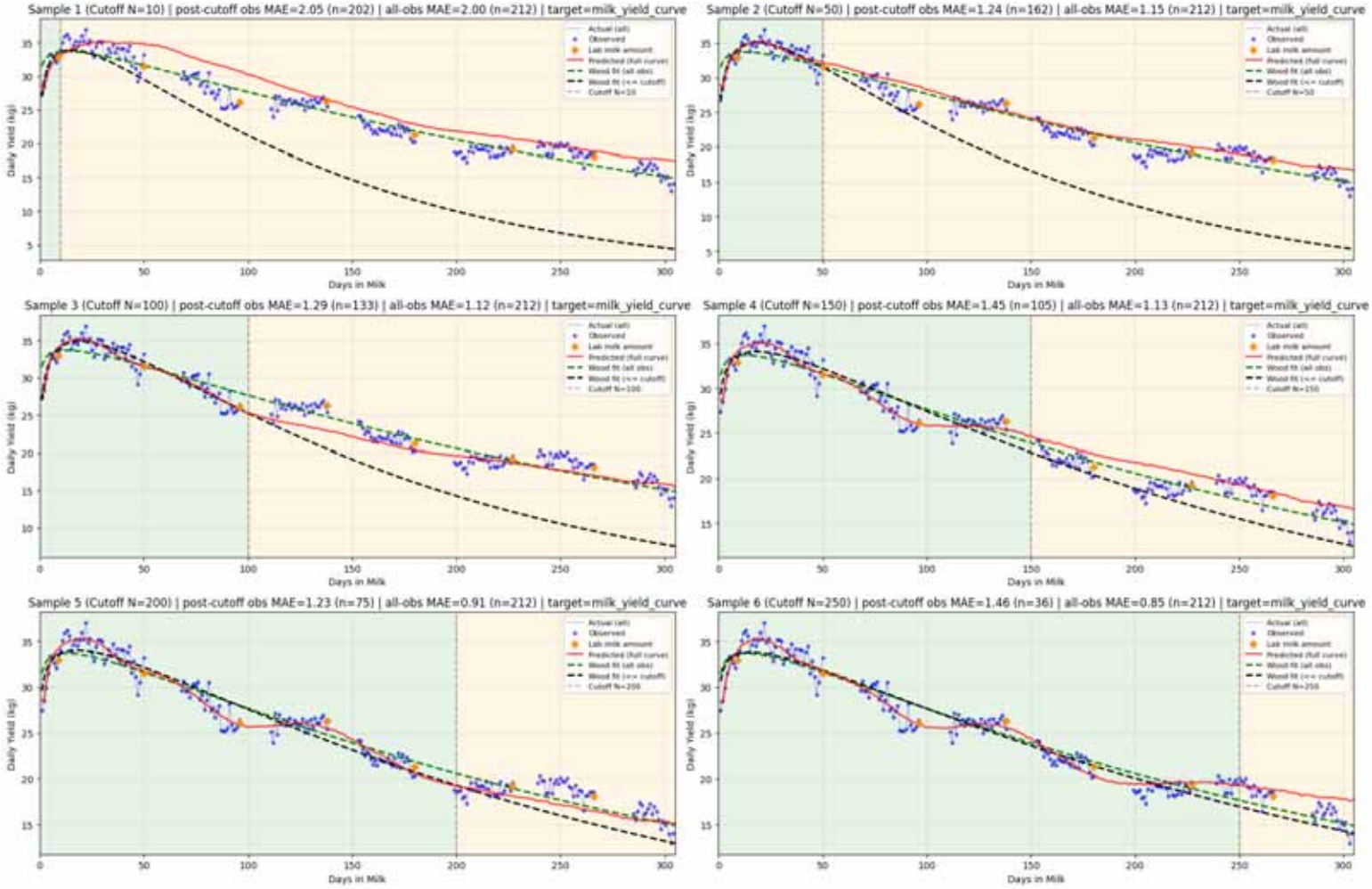
- Validation restricted to Fleckvieh in Austria.
- Final comparison was at five validation days.
- Needs monitoring for extreme yields and unusual lactation curves.

Next work

- additional breeds and later data releases
- continuous or at least annual retraining
- daily milk yield prediction and full lactation-curve forecasting
- production monitoring and calibration checks

RINDERZUCHT AUSTRIA Innovation

ZUCHT DATA



Thank You

Funding: This research was funded by the **Qplus-Kuh** project, 77-02-BML Zusammenarbeit, Project “Qplus Kuh“, reference no. LE-77-02-BML-2024-21859