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

Ganitzer Jakob*[1], Schwarzenbacher Hermann[1], Fürst Christian[1]

[1] ZuchtData EDV Dienstleistungen GmbH

Accurate prediction of standardized 305-day lactation yield at various days in lactation enables early performance assessments. This study develops and validates an XGBoost regression model to predict standardized total lactation yield using partial lactation data in Fleckvieh dual purpose cattle. The objective was to develop a machine learning model for 305-day standard lactation milk yield prediction from intermediate records and to compare the approach against a linear regression method.

We engineered a feature set combining production measurements (latest test day milk yield, cumulative yield to date, herd average yield), genetic predictors (estimated breeding values for milk yield and persistency), somatic cell counts, and cyclical features (sine/cosine transformations for test day). Additionally, lag features representing prior milk yields and somatic cell counts were generated to incorporate animal-specific historical trends. A year/lactation level data splitting prevented leakage between training and validation sets. Model validation included metrics at multiple lactation stages and overall animal-level performance stratification.

Over the full test-set (overall performance) the XGBoost model achieved R^2 = 0.94 and a mean absolute error of 330 kg (305-day yield), demonstrating competitive performance with traditional regression baselines (R^2 = 0.94, MAE = 334 kg). Sensitivity testing showed cumulative yield, current milk production and herd average yield, combined with the specific day of lactation as dominant predictors. Error analysis stratified by lactation number highlighted that the model maintains stability across different animal ages. This XGBoost model provides accurate, interpretable 305-day lactation yield predictions from intermediate measurements, providing a robust alternative to traditional lactational yield prediction methods. Future work will focus on daily milk yield prediction and lactation curve forecasting. Keywords: Dairy cattle, machine learning, lactation yield prediction, XGBoost