

Management Summary

The goal of this study was to evaluate an intuitive, user-friendly tool to help farmers make reproduction management decisions by using herd- and cow-level parameters to forecast the chance of conception success for a particular insemination. Machine learning was used to predict the insemination outcomes of individual cows based on phenotypic data using the Modulos AutoML platform.

Machine learning techniques are ideally suited for investigating dairy cattle reproduction performance due to their capacity to handle complicated correlations in data for explanatory variables. ML is a well-established and widely used methodology in both agriculture and academia, but scientific research on its use in breeding is typically conducted without the use of a standard process model to increase the performance and efficiency of machine learning applications. In our master's thesis we developed an ML application for predicting the reproductive management of dairy cows following the principles of the CRISP-ML(Q).

The study examined data from 2,500 Swiss farms during a five-year period from 2015 to 2020. Data on health, reproduction, production, and breeding values were taken from Qualitas AG's ArgusQ database (a subsidiary of swissherdbook) and evaluated using the Modulos AutoML platform. There were 210'000 breeding records from cows of various breeds in the prepared data set (including Holstein, Swiss Fleckvieh and Simmental). Each data point in the final data set included 65 explanatory variables and 1 binary label. The model provided a reasonable prediction of the likelihood of conception success. In the external validation (blind-test), based on 15% of the available data (38'046 records) the model predicted 19'982 successful and 18'064 unsuccessful inseminations, the number of false positives are 5'645, the number of false negatives are 4'667, while true positives are 14,337 and true negatives are 13'397. The F1 Score binary was 73.5%, the recall score was 75.4% and the accuracy was 72.8%. This also results in a positive ML effect of 6.57 CHF per successful insemination.

We can therefore summarize that the developed model has a positive economic and statistical performance. However, we consider the number and percentage of false negatives problematic. The consequence of false negative results could be the culling of still functional cows by the breeders, this poses an ethnic risk. Further research and optimization is necessary before swissherdbook could release such a product for its customers.

Key words: reproductive management, dairy cattle, prediction, conception, algorithm, machine learning, quality assurance methodology, guidelines

Table of contents

Acknowledgments	Error! Bookmark not defined.
Management Summary	Error! Bookmark not defined.
Table of Contents	Error! Bookmark not defined.
List of Figures.....	Error! Bookmark not defined.
List of Tables.....	Error! Bookmark not defined.
Abbreviations	Error! Bookmark not defined.
1 Introduction	Error! Bookmark not defined.
2 Literature Review	Error! Bookmark not defined.
2.1 swissherdbook and the tasks of a breeder association	Error! Bookmark not defined.
2.2 Global dairy sector	Error! Bookmark not defined.
2.3 Swiss dairy sector	Error! Bookmark not defined.
2.4 Smart Farming and Digitalization in Dairy Farming	Error! Bookmark not defined.
2.5 Collect and use dairy data for dairy breeding and management	Error! Bookmark not defined.
2.5.1 Evolution of data recording and use since the early years	Error! Bookmark not defined.
2.6 Definition of AI.....	Error! Bookmark not defined.
2.6.1 Machine Learning.....	Error! Bookmark not defined.
2.6.2 The Role of Big Data.....	Error! Bookmark not defined.
2.7 Opportunities and Risks of AI in dairy farming.	Error! Bookmark not defined.
2.7.1 Genomic evaluations using AI.....	Error! Bookmark not defined.
2.7.2 Management Decision Tool using AI.....	Error! Bookmark not defined.
2.7.3 Risk of overfitting and the black box.....	Error! Bookmark not defined.
3 Methods	Error! Bookmark not defined.
3.1 The Value Proposition Canvas	Error! Bookmark not defined.
3.2 Value Proposition Canvas for swissherdbook..	Error! Bookmark not defined.
3.2.1 Customer Profile	Error! Bookmark not defined.
3.2.2 The Value Map.....	Error! Bookmark not defined.

3.3	The Machine Learning Canvas.....	Error! Bookmark not defined.
3.4	Introduction of CRISP-ML(Q) Methodology	Error! Bookmark not defined.
3.5	CRISP-ML(Q) Process Phases	Error! Bookmark not defined.
3.5.1	Business and Data Understanding	Error! Bookmark not defined.
3.5.2	Data Preparation	Error! Bookmark not defined.
3.5.3	Modeling	Error! Bookmark not defined.
3.5.4	Evaluation	Error! Bookmark not defined.
3.5.5	Deployment.....	Error! Bookmark not defined.
3.5.6	Monitoring and Maintenance	Error! Bookmark not defined.
3.6	CRISP-ML Checklist.....	Error! Bookmark not defined.
4	Results and Discussion	Error! Bookmark not defined.
4.1	Business and Data Understanding	Error! Bookmark not defined.
4.1.1	Define the Scope of the Machine Learning Application.....	Error! Bookmark not defined.
4.1.2	Success Criteria	Error! Bookmark not defined.
4.1.3	Feasibility	Error! Bookmark not defined.
4.1.4	Data Collection.....	Error! Bookmark not defined.
4.1.5	Data Quality Verification.....	Error! Bookmark not defined.
4.1.6	Review of Output Documents.....	Error! Bookmark not defined.
4.2	Data Preparation	Error! Bookmark not defined.
4.2.1	Select Data	Error! Bookmark not defined.
4.2.2	Clean Data	Error! Bookmark not defined.
4.2.3	Construct Data	Error! Bookmark not defined.
4.2.4	Standardize Data	Error! Bookmark not defined.
4.2.5	Backtracking circles	Error! Bookmark not defined.
4.3	Modeling	Error! Bookmark not defined.
4.3.1	Literature research on similar problems ...	Error! Bookmark not defined.
4.3.2	Define quality measures of the model	Error! Bookmark not defined.
4.3.3	Model selection	Error! Bookmark not defined.
4.3.4	Incorporate domain selection	Error! Bookmark not defined.
4.3.5	Model training	Error! Bookmark not defined.



4.3.6	Using unlabeled data and pre-trained models	Error! Bookmark not defined.
4.3.7	Model Compression	Error! Bookmark not defined.
4.3.8	Ensemble methods	Error! Bookmark not defined.
4.3.9	Assure reproducibility and Documentation	Error! Bookmark not defined.
4.4	Model evaluation	Error! Bookmark not defined.
4.4.1	Validate performance	Error! Bookmark not defined.
4.4.2	Determine robustness	Error! Bookmark not defined.
4.4.3	Increase explainability for ML practitioner & end user	Error! Bookmark not defined.
4.4.4	Compare results with defined success criteria	Error! Bookmark not defined.
4.5	Deployment	Error! Bookmark not defined.
4.6	Monitoring and Maintenance	Error! Bookmark not defined.
5	Conclusions and Outlook	Error! Bookmark not defined.
5.1	Conclusions	Error! Bookmark not defined.
5.2	Outlook	Error! Bookmark not defined.
	Bibliography	Error! Bookmark not defined.
	Appendix 1: R Scripts for data selection	Error! Bookmark not defined.