AGRO PRODUCTIVIDAD

Colegio de Postgraduados

Early prediction of lactation milk yield of Holstein-Friesian cows in Querétaro, Mexico

Guevara-Escobar, A.[1]**; Cervantes-Jiménez, M.**[1]**; García-Muñiz, J.G.**[2*]**; Lemus-Ramírez, V.**[3]**; Yabuta-Osorio, A. K.**[3]

[1] Universidad Autónoma de Querétaro, Facultad de Ciencias Naturales. Juriquilla, Querétaro, Querétaro, México. C.P. 76210.

[2] Universidad Autónoma Chapingo, Departamento de Zootecnia, Posgrado en Producción Animal. Chapingo, Texcoco, Estado de México, México. C.P. 56230.

[3] Universidad Nacional Autónoma de México, Centro de Enseñanza, Investigación y Extensión en Producción Animal en Altiplano. Carretera Federal Tequisquiapan a Ezequiel Montes, km. 8.5. Tequisquiapan, Querétaro, México. CP 76790.

* Correspondence: jgarciamppa@hotmail.com

ABSTRACT

Objective: To obtain a machine learning (ML) model to predict the milk yield adjusted to 305 d (MY305) from the same lactation period.

Design/methodology/approach: A database of test days (TD) was used, made up by 11,892 records of daily milk production from cows with more than 150 days in milk (DIM), from 19 farms in Querétaro, Mexico. The milk production was standardized to specific DIMs (5, 10, 20, 30 and 40) and estimations of MY305 were obtained with these, using ML models. The following were also incorporated as explicative variables of the herd: month of birth of the cow, month of start of lactation, number of lactation, number of days for three daily milking events, and the two first linear scores of somatic cells.

Results: The best goodness of fit was achieved with ensemble models, obtaining a deviance of 1503584 in the training with 80% of data chosen randomly, while with 20% of the data reserved to evaluate the deviance model it was 1576776. The relationship between data observed and predictions of MY305 of the ensemble models had a coefficient of determination of $r^2=0.79$ and RMSE of 1256. In the best individual model (deviance of 2281420) of 'deep learning' type, the most important variables were daily milk production at 30, 10, 5 and 20 DIM (19.9, 16.6, 16.2 and 12.8%, respectively).

Limitations on study/implications: The value of RMSE was high. Although TD databases are generated regularly and following systematic measurement procedures but not many farms are represented.

Findings/conclusions: For the database examined, milk production in the early phase of lactation together with a set of automatic learning models resulted in an adequate prediction of MY305.

Keywords: machine learning, somatic cells, lactation curve, test day.

Citation: Guevara-Escobar, A., Cervantes-Jiménez, M., García-Muñiz, J.G., Lemus-Ramírez, V., & Yabuta-Osorio, A.K. (2024). Early prediction of lactation milk yield of Holstein-Friesian cows in Querétaro, Mexico. *Agro Productividad*. https://doi.org/10.32854/agrop.v17i7.2743

Academic Editor: Jorge Cadena Iñiguez
Guest Editor: Juan Francisco Aguirre Medina

Received: Novermber 14, 2023.
Accepted: Julio 14, 2024.
Published on-line: August 08, 2024.

Agro Productividad, 17(7). July. 2024. pp: 93-103.

INTRODUCTION

The predictions of daily cow milk production (MY), including the current lactation, are important to identify critical situations of productive interest and with that to make decisions about nutrition, feed management, reproduction, preventive health, genetic

improvement, among others [1], [2]. The prediction of MY is possible with data-based models of: prior and/or contemporary lactation events of the cow or herd [3]-[5]. Traditionally, MY data in test days (TD) are used to predict the future individual production of cows based on the lactation curve; in these cases, the incomplete gamma function proposed by Wood in 1967 [6], is one of the most frequently used to describe the lactation curve [7]. When the MY is modelled in cow groups, the data of genotype, weight and age of the cow, body condition, frequency of milking and somatic cell count (SCC) in the milk, among others, are also useful.

Presently, there are machine milking systems that record the MY in every milking, in addition to other variables. Currently, the integration of this type of database is one of the most important challenges in the dairy industry due to the diversity of technologies available and the overflowing of data that happens in dairy operations [8]. However, the databases of MY in TD continue being essential for the study of prediction models of MY [4]. Modelling of individual lactation curves or groups of cows with empirical models, such as the Wood function [6] or mechanistic ones [9], allow performing punctual or accumulated MY estimations in time. To perform comparisons between cows, there is a well-established metric, where the MY is adjusted to 305 days in milk (DIM) and to a mature equivalent (MY305) [10], [11]. The MY305 can be adjusted according to the milk solids content and by the number of lactations, allowing to establish comparisons between cows.

Currently, the predictive models of MY based on machine learning (ML) algorithms are increasingly more frequent in the literature and their performance has shown greater advantages than the regression models adjusted by least squares [2], [12]. The applications of the ML models in the milking systems are broad [12], [13]. Ensemble models with strategies of classification and/or regression can be obtained with the ML approach, using incomplete data records, crossed-validation routines, selection of variables, exploration of hyper-parameters, selection of models and automatic ensemble models, with the intention of avoiding the model's overfitting [14]. With the ML models, there have been advancements in the interpretability of the model, allowing to identify the importance of the variables included in the model and the contribution to the solution of specific data records or in the set of the database [15].

For this study, the data available from the early lactation phase of Holstein-Friesian cows were used to predict the MY305. With this approach, Pereira *et al.* [16] identified the Bayesian empirical method and only 5 TD, as the most effective in the prediction of MY305 in Holstein livestock. The neural networks have performed better to predict the MY with early lactation data or from the transition period [4], [17]. Although the data from the transition period of the milking cow are important for the health and future milk yield [18], it is a challenge to predict the future MY, since the TD during early lactation tend to be few and also the health state of the cow can be compromised [4]. The use of daily milk production records based on TD data are proposed, together with the data that allow classifying cows according to their month of birth and their calving, number of lactation, herd and somatic cell count. The objective was to obtain a ML model to predict MY305 based on data from early lactation of Holstein-Friesian cows in Querétaro, Mexico.

MATERIALS AND METHODS

Data from lactation periods that started in 2006 were used, generated in Mexico by Mexico's Holstein Association (*Asociación Holstein de México*, AHM), from 19 herds in the state of Querétaro corresponding to 13,935 lactation records of Holstein-Friesian cows. The cows recorded were from lactation one to ten. Each lactation record had up to ten milk weighing moments. Most of the herds were kept in the milk production confinement system.

The Wood function (Equation 1) was used to model the individual lactation curve for the records with more than five TD and DIM higher than 150 (n=11,892). The milk production adjusted to mature equivalent obtained by the AHM (MY305, kg) was predicted through explicative variables: herd, month of birth of the cow, month of start of lactation, number of lactation, number of days in three times daily milking (3x), first two scores of the linear scoring (LS) system based on the recount of somatic cells (LS 1 and LS 2, Equation 2) (adopted by the Dairy Herd Improvement Association, DHIA), and the estimated milk production ($\mathrm{M\widehat{Y}}$, kg) with the Wood function for days: 5, 10, 20, 30 and 40 ($\mathrm{M\widehat{Y}5}$, $\mathrm{M\widehat{Y}10}$, $\mathrm{M\widehat{Y}20}$, $\mathrm{M\widehat{Y}30}$ and $\mathrm{M\widehat{Y}40}$). The $\mathrm{M\widehat{Y}}$s were selected due to the irregularity in the intervals of TD to different DIM. The discreet variables that classify the cows were used based on the report by Grzesiak *et al.* [7] to model the daily production of milk using neural networks.

$$Y_t = at^b \; e^{-ct} \tag{1}$$

where: Y_t is the daily milk production (kg) on day t of lactation, e is the base of the natural logarithms, a, b, c, are adjustment parameters [6].

Calculation of linear scoring based on the somatic cell count (SCC):

$$LS = Log_2\left(SCC/100\right) + 3 \tag{2}$$

where: LS is the linear scale of the somatic cell count, Log_2 is the logarithm base 2, SCC is the somatic cell count (cells ml^{-1} of milk). The conversion of LS to SCC is achieved with the following expression: $SCC = 100 \times 2^{(LS-3)}$ [19].

Machine learning (ML) algorithms were used to obtain a predictive model of MY305 through the H2O package version 3.40.0.1 [14] of the R language [20]. The AutoML function results in a leaderboard of models and ensemble models generated from models of same leaderboard. The deviance was used as a metric of goodness of fit in the training of models and also to order the ensembles and individual models. The model with best adjustment was the one that produced lowest deviance. For the best individual model, the importance of the variables with their SHAP values was determined (SHapley Additive exPlanations). The code was executed without a time limit, with 50 processing threads on a dual Xeon E5-2680 v4 cluster and with a maximum of 360 Gb of RAM to execute the

AutoML function. Memory was reserved with another 152 Gb of RAM for external code routines of XGBoost (Optimized distributed gradient boosting machine) and within the Ubuntu 22.04.2 LTS operating system.

The AutoML function implies the pre-processing of data, normalization, model selection, optimization of hyper-parameters, analysis of predictions, and control of the over-parametrization of the model [21]. The ML algorithms used in this exercise were: DL (Deep learning), GBM (Gradient boosting machine), XGBoost and GLM (Generalized linear model). However, AutoML explores up to 15 types of algorithms to solve problems of classification and regression [14]. Each ML algorithm is executed many times to explore the effects of regularization and crossed validation on the accuracy of the model [22].

The database was divided randomly in data for the training (80%) and 20% for the model testing [13]. During the training phase, the crossed validation was used to verify the stability of the solution; that is, an internal validation of the model defined by the parameter nfolds=5. Thus, the training data were divided randomly into five groups; four groups were used to train the model, and the fifth to test the performance of the previously trained model. In each training run, the training database was again divided into five groups.

The SHAP values [23] are based on the concept of the Shapely values [24] and they quantify the influence that each variable included in the model has for each individual prediction to deviate from the average prediction. However, each variable contributes in a different way in function of the k variables incorporated into the model; that is, 2k combinations. The SHAP values can be evaluated locally (local explanations) where the contribution of each variable on the result from each prediction is examined. Globally, the local explanations were aggregated to understand the impact of specific variables on the entire model. For individual models it is possible to obtain local and global explanations, but it is not possible for the case of ensemble models.

With the best ML model, estimations of values observed for the MY305 from the database reserved for testing were obtained. The relationship between values observed and estimated was explored with a simple linear regression model and the coefficient of determination (r^2), the root of the mean square error (RMSE), and the bias were obtained; the latter through the blandr package according to Bland and Altman [25].

RESULTS AND DISCUSSION

In the management of dairy cows from the herds studied, extended lactation periods and three daily milking moments were identified as important practices. Lactation with less than 150 DIM were not analyzed (Table 1). In the database, the records from the first lactation were the most numerous (40%), and from these lactation events, 31% were with 150 to 305 DIM and with three times daily milking. No first lactation longer than 305 DIM had days to three times daily milking. The average MY305 of the data used was 11,282 kg, although there was variation according to the duration of lactation and the number of lactation (Table 2). The herds with less than 100 lactation records had a lower MY305 than the herds with more than 100 records (9,233 and 11,310 kg, p=0.012).

Table 1. Number of production records used according to their number of lactation, days in milk (DIM), and the practice of three milking moments per day (3X) for Holstein-Friesian cows in Querétaro, Mexico.

DIM	Lactation number				Subtotal	3x
	1	2	3	4+		
>150 to <305	1803	1357	735	741	4636	2969
>=305 to 365	1343	880	466	449	3138	1654
>=365 to 730	1559	1074	613	740	3986	2057
>730	70	27	18	17	132	91
Total	4775	3338	1832	1947	11892	6771

Table 2. Means ($\bar{x}$) and standard error of the mean (se) for the milk production to mature equivalent obtained by the AHM (MY305, kg) of Holstein-Friesian cows in Querétaro, Mexico.

DIM	Lactation number							
	1		2		3		4+	
	$\bar{x}$	se	$\bar{x}$	se	$\bar{x}$	se	$\bar{x}$	se
>150 a <305	11638	65	11287	73	10875	103	10066	95
>=305 a 365	11697	74	11569	89	11197	130	10789	132

The best result to predict the MY305 was a model ensemble of ML algorithms used. In this case, the deviance was 1503584 and a RMSE of 1226 for the training data, while with the data reserved for the evaluation, the deviance was 1576776 and a RMSE of 1256. These measurements of goodness of fit suggested that the ensemble was a good representation of the database and that there was no over-adjustment. This indicated that the AutoML function solved the possible multicollinearity that could exist between the M$\widehat{Y}$ values. The second best result was an ensemble of the best family of algorithms used, with deviance of 1717832 and RMSE of 1311 for the training data. Four other ensembles followed in the leaderbord (data not shown). Verification of the best ensemble resulted in r^2 of 79.9 between observed and predicted values of MY305 (Figure 1). The bias was −10.42 with a confidence interval at 95% of −63.8 to 43.0; these values were interpreted as the magnitude of underestimation of the model in relation to the MY305 observed. In this case, the confidence interval did not include zero, which indicated that the bias was significant.

The best individual model was the DL type, which remained in sixth place in the leaderboard and their deviance was 2281420 and RMSE of 1510, which resulted in a lower goodness of fit than those of the ensembles. Out of 797 models and ensembles generated, the worst model was GLM type, with deviance 7621002 and RMSE of 2760. This result was explained because the GLM algorithm was highly penalized in the regularization process carried out by the AutoML function to reduce the over-adjustment caused by the multicollinearity in variables. In contrast, with ML algorithms such as neural networks and ensemble methods, this problem can be handled [26], [27]. For the prediction of MY305, the most important variables were the M$\widehat{Y}$s at 5, 10, 20 and 30 DIM, followed by the somatic cell count and lactation number (Table 3). The variables of month of birth

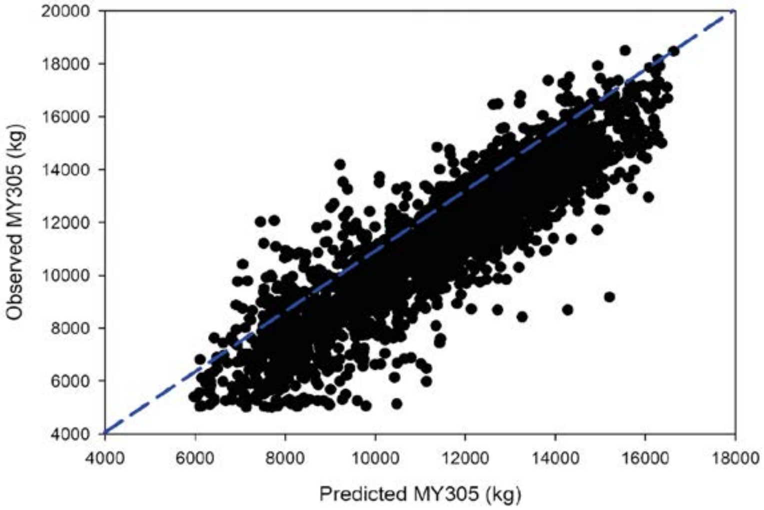

Figure 1. Relationship between values observed of milk production at mature equivalent adjusted to 305 d (MY305) and the corresponding values estimated by ensemble models of machine learning, $y=-951+1.074x$, $r^2=0.79$.

Table 3. Percentage importance of the variables adjusted in individual models of machine learning to predict the production of milk to mature equivalent (MY305) of Holstein-Friesian cows in Querétaro, Mexico.

	Model			
	DL	**GBM**	**GBM**	**XGboost**
$M\hat{Y}30$	19.9	55.1	56.1	34.7
$M\hat{Y}10$	16.6	6.5	6.3	9.6
$M\hat{Y}5$	16.2	7.5	7.5	13.4
$M\hat{Y}20$	12.8	4.4	4.0	10.2
CS 2	6.4	2.3	2.2	3.2
Lactation number	6.2	8.9	9.2	9.1
CS 1	5.2	2.5	2.4	3.3
Herd	4.9	6.8	7.1	7.3
Month of calving	4.3	2.8	2.6	3.4
3X	3.9	1.2	1.0	2.8
Cow's month of birth	3.8	2.0	1.7	2.9
Ranking in the table of models	6°	8°	9°	11°
Deviance	2281420	2672503	2703187	2762434
RMSE	1510	1634	1644	1662

of the cow and calving month were variables of lower importance, just as 3X. Although calving month had importance of 4.2% in the DL model (and lower in other models), seasonal differences were observed for MY305 with milk productions from lactation started in winter, summer and spring being similar between them (11,111, 11,112 and 11,181 kg, $p>0.05$), winter and summer were different from fall (11,349 kg, $p<0.05$), yet at the same time fall and spring were similar. This result suggests that the variables with low importance in the ML models also must be considered in modelling.

Figure 2 shows the SHAP contribution of each variable and each observation (lactation records) in the prediction of MY305 by an individual model of XGBoost type (eleventh on the scoring table). Using the H2O package, it is only possible to obtain the SHAP values for the models XGBoost, GBM and Random Forest. When the SHAP values were graphed, the aggregated trend of the contributions of each observation was appreciated where some of them contributed more than others. In the XGBoost values, the larger magnitude of the SHAP values of $M\widehat{Y}30$ and $M\widehat{Y}5$ indicated a higher predicted value of MY305 (reddish colors with positive SHAP value). In this model, the $M\widehat{Y}30$ variable was more important than the $M\widehat{Y}5$ variable (Table 3), because the SHAP values of $M\widehat{Y}30$ were more useful to build a classification tree. For the $M\widehat{Y}10$ and $M\widehat{Y}20$ variables, the SHAP values of highest magnitude contributed to predicting lower milk yield (reddish colors with negative SHAP values). Likewise, as the SHAP value of number of lactation increased, the contribution was in the sense of lower magnitude of the MY305. For the linear scale (LS) of the somatic cell count, a negative contribution was observed in the model, but with comparatively lower SHAP values than the variables already mentioned.

The distribution of the SHAP values for the $M\widehat{Y}$ variables suggested that there are thresholds for these variables associated with a higher value of MY305. The graphs of partial dependency indicate that these thresholds are different between the ML models generated and in addition that the dependency was not linear between the MY305 and each variable included in the model, for example, for $M\widehat{Y}30$ (Figure 3). The general linear model (GLM) did not show dependency of the MY305 in function of the range of $M\widehat{Y}$ values. For the ensemble models, the partial dependency graph suggested that values of $M\widehat{Y}30$ lower than 20 kg are not important to predict MY305; the same happened with values of $M\widehat{Y}30$ higher than 40 kg when the dependency decreased. The grey bars in Figure 3 represent the frequency distribution of the data of MY305.

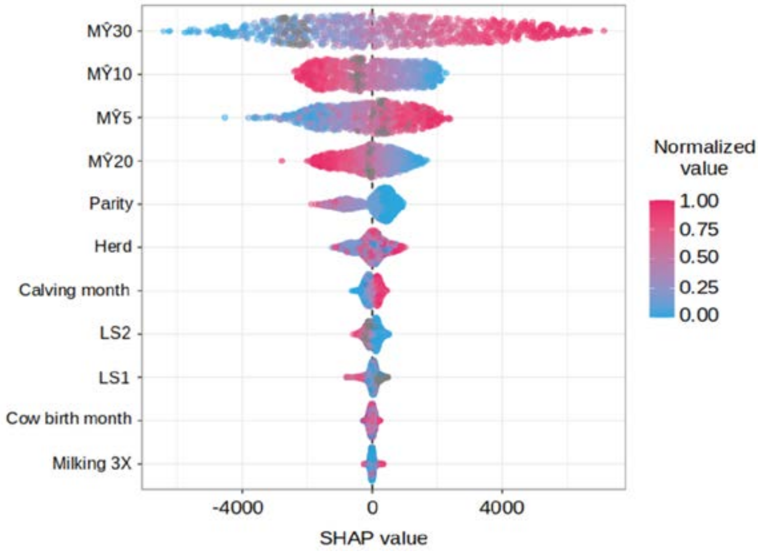

Figure 2. SHAP values for the XGBoost model presented in Table 3.

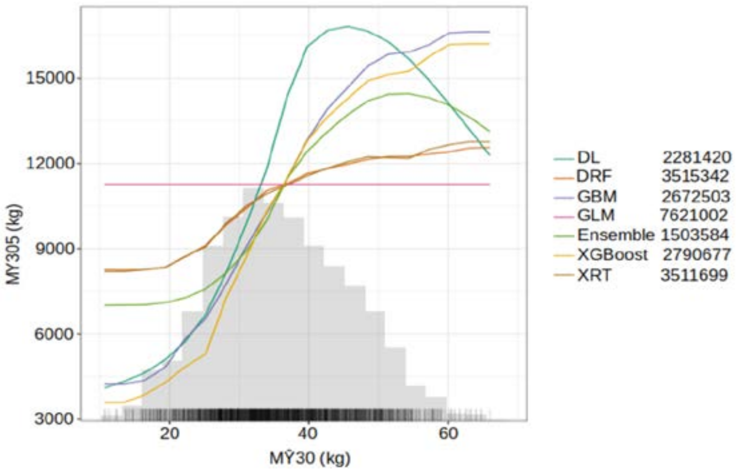

Figure 3. Partial dependency of MY305 on M$\hat{Y}$30 for different ML models. The deviance for each model is presented. The frequency histogram of M$\hat{Y}$30 is shown in grey.

The average values and standard deviations of MY305 of the first and second lactation events in the database examined (11,586±2,760 kg and 11,473±2,701 kg) were found within the range reported by Toledo-Alvarado *et al.* [28] in Holstein cows in Mexico, where the MY305 for the first lactation was 11,473±2,443 kg and for the second one 11,612±2,652 kg; other lactation events showed slightly lower total yield than those reported in this study. Since the heifers are genetically superior, they generally have higher MY305 than the mature cows of lower genetic quality [29]. Likewise, the values of MY305 reported here were numerically lower than the average of 12,662 kg for Holstein in the United States [30].

We used estimated values of daily milk production (M$\hat{Y}$) during the early lactation based on modelling with the Wood function. These data fulfilled the function of representing measurements of early lactation; in many farms these data could come directly from the registry of the automatized milking system. Based on our result, the daily registry of milk production can be explored during early lactation to obtain estimates of the MY305 or daily MY. This would be similar to the approach of [2], who used the machine learning algorithm XGBoost. In that study, with climate data of the 60 previous days, the identification, age and weight of buffalo, their feed consumption and their days in milk, the daily frequency of milking, milk production, and fat and protein composition in the next 28 days ($r^2=0.90$) were predicted.

The variance and auto-correlation of deviations of the daily milk production during early lactation could be related to clinical mastitis [1].

The variables of somatic cell count were included in the models, but they were not the most important. Of the 13,935 records, the LS did not have a value in any of the two dates examined (4,661 and 4,361 records) and 4,206 records did not have a value in any of these dates. It is possible that the low importance of LS 1 and LS 2 was a reflection of a disperse database.

It is also possible that the management of some herds was satisfactory and the incidence of mastitis low. The herds with less than 100 records of lactation had higher counts than the herds with more than 100 records (LS 1 of 29.56 and 22.52 and LS 2 of 27.40 and 18.31, $p<0.001$). However, this result can be an effect from the dilution of the LS, since the herds with more than 100 cows had higher MY305. Correcting the LS in function of the MY would eliminate the effect of dilution [31].

The proposal by Singh *et al.* [17] to predict the MY305 ($r^2=0.82$) of buffalo in the first lactation is the one most similar in the literature to this study's approach. They considered the milk production in test days at 6, 36, 66 and 96 DIM, the daily milk production in the peak of lactation, and the age at first calving. Cook *et al.* [29] considers the first test day (around 20 DIM) as an important predictor of yield at 305 d, because many of the health problems during the transition period would explain a low daily milk production at first DIM.

In this study, the decision to use $M\widehat{Y}$ at specific lactation days was made, with the aim of avoiding a disperse database. Although the ML algorithms can deal with disperse databases, their efficiency in training the model is low because the processing time increases and the complexity of the models also increases. In this sense, the use of the daily milk production records during early lactation would still have to be explored, even when the records of the number of test days and the days in milk of the measurement do not agree, as is the case with disperse databases. One possibility would be to group the measurements of test day by week of lactation, and thus to reduce the number of lactation records without data for a given week.

CONCLUSION

This study showed that ensemble models of machine learning allow estimating the milk production adjusted to mature equivalent and that the deep learning algorithm is capable of generating the best individual model. Among the ten best models, six ensembles had the lowest deviance, followed by a deep learning model and the last three were gradient boosting machine. The daily milk production variables in the early phase of lactation, from 5 to 30 days, were important in the models; however, this was not the case for the variables that classify the livestock, such as month birth, month of calving, the herd and others. No model included the daily milk production at 40 days of lactation.

AKNOWLEDGEMENTS

The authors wish to thank Mexico's Holstein Association for the information provided.

REFERENCES

[1] Kok, A.; Tsousis, G.; Niozas, G.; Kemp, B.; Kaske, M.; Van Knegsel, A.T.M. 2021. Short communication: Variance and autocorrelation of deviations in daily milk yield are related with clinical mastitis in dairy cows. *Anim. 15*(10). 100363. doi: 10.1016/j.animal.2021.100363.

[2] Ji, B.; Banhazi, T.; Phillips, C.J.C.; Wang, C.; Li, B. 2022. A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. *Biosyst. Eng.* 216. 186-197. doi: 10.1016/j.biosystemseng.2022.02.013.

[3] Liseune, A.; Salamone, M.; Van Den Poel, D.; Van Ranst, B.; Hostens, M. 2021. Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning. *Comput. Electron. Agric.* 180. 105904. doi: 10.1016/j.compag.2020.105904.

[4] Salamone, M.; Adriaens, I.; Vervaet, A.; Opsomer, G.; Atashi, H.; Fievez, V.; Hostens, M.M. 2022. Prediction of first test day milk yield using historical records in dairy cows. *Anim. 16*(11). 100658. doi: 10.1016/j.animal.2022.100658.

[5] Zhang, F.; Weigel, K.A.; Cabrera, V.E. 2022. Predicting daily milk yield for primiparous cows using data of within-herd relatives to capture genotype-by-environment interactions. *J. Dairy Sci. 105*(8). 6739-6748. doi: 10.3168/jds.2021-21559.

[6] Wood, P.D.P. 1967. Algebraic Model of the Lactation Curve in Cattle. *Nature*. 216. 164-165. doi: 10.1038/216164a0.

[7] Grzesiak, W.; Błaszczyk, P.; Lacroix, R. 2006. Methods of predicting milk yield in dairy cows -Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Comput. Electron. Agric. 54*(2). 69-83. doi: 10.1016/j.compag.2006.08.004.

[8] Cabrera, V.E.; Fadul-Pacheco, L. 2021. Future of dairy farming from the Dairy Brain perspective: Data integration, analytics, and applications. *Int. Dairy J*. 121. 105069. doi: 10.1016/j.idairyj.2021.105069.

[9] Pollott, G.E. 2004. Deconstructing milk yield and composition during lactation using biologically based lactation models. *J. Dairy Sci. 87*(8). 2375-2387. doi: 10.3168/jds.S0022-0302(04)73359-7.

[10] Cole, J.B.; Null, D.J.; VanRaden, P.M. 2009. Best prediction of yields for long lactations. *J. Dairy Sci. 92*(4). 1796-1810.doi: 10.3168/jds.2007-0976.

[11] Cole, J.B.; Ehrlich, J.L.; Null, D.J. 2012. Short communication: Projecting milk yield using best prediction and the MilkBot lactation model. *J. Dairy Sci. 95*(7). 4041-4044. doi: 10.3168/jds.2011-4905.

[12] Shine, P.; Murphy, M.D. 2012. Over 20 years of machine learning applications on dairy farms: A Comprehensive mapping study. *Sensors. 22*(1) 52. doi: 10.3390/s22010052.

[13] De Vries, A.; Bliznyuk, N.; Pinedo, P. 2023. Invited Review: Examples and opportunities for artificial intelligence (AI) in dairy farms. *Appl. Anim. Sci. 39*(1). 14-22. doi: 10.15232/aas.2022-02345.

[14] Hall, P.; Gill, N.; Kurka, M.; Phan, W.; Bartz, A. 2023. Machine learning interpretability with H2O driverless AI. H2O.ai Inc, Mountain View California. https://docs.h2o.ai/driverless-ai/latest-stable/docs/booklets/MLIBooklet.pdf

[15] Rudin, C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell. 1*(5). 206-215. doi: 10.1038/s42256-019-0048-x.

[16] Pereira, J.A.C.; Suzuki, M.; Hagiya, K. 2001. Prediction of 305 days milk production from early records in dairy cattle using an empirical Bayes method. *Asian Australas. J. Anim. Sci. 14*(11). 1511-1515. doi: 10.5713/ajas.2001.1511.

[17] Singh, N.P.; Dutt, T.; Usman, S.M.; Baqir, M.; Tiwari, R.; Kumar, A. 2022. Prediction of first lactation 305 days milk yield using artificial neural network in Murrah buffalo. *Indian J. Anim. Sci. 92*(9). doi: 10.56093/ijans.v92i9.117570.

[18] Pulina, G.; Tondo, A.; Danieli, P.P.; Primi, R.; Crovetto, G.M.; Fantini, A.; Macciotta, N.P.P.; Atzori, A.S. 2020. How to manage cows yielding 20,000 kg of milk: technical challenges and environmental implications. *Ital. J. Anim. Sci. 19*(1). 865-879. doi: 10.1080/1828051X.2020.1805370.

[19] Schukken, Y.H.; Wilson, D.J.; Welcome, F.; Garrison-Tikofsky, L.; Gonzalez, R.N. 2003. Monitoring udder health and milk quality using somatic cell counts. *Vet. Res. 34*(5). 579-596. doi: 10.1051/vetres:2003028.

[20] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation: Vienna, Austria. 2673 p.

[21] Truong, A.; Walters, A.; Goodsitt, J.; Hines, K.; Bruss, C.B.; Farivar, R. 2019. Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools. doi: 10.48550/ARXIV.1908.05557.

[22] Mitchell, R.; Frank, E. 2017. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput. Sci*. 3. e127. doi: 10.7717/peerj-cs.127.

[23] Fryer, D.; Strümke, I.; Nguyen, H. 2021. Shapley values for feature selection: The good, the bad, and the axioms. doi: 10.48550/ARXIV.2102.10936.

[24] Shapley, L.S. 1951. Notes on the n-person game II: the value of an n-person game. U.S. Air Force Project RAND», Rand Corporation, Santa Monica California, Research Memorandum 670. https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf

[25] Bland, J.M.; Altman, D.G. 2010. Statistical methods for assessing agreement between two methods of clinical measurement. *Int. J. Nurs. Stud. 47*(8). 931-936. doi: 10.1016/j.ijnurstu.2009.10.001.

[26] Chan, J.Y.L.; Leow, S.M.H.; Bea, K.T.; Cheng, W.K.; Phoong, S.W.; Hong, Z.W.; Chen, Y.L. 2022. Mitigating the multicollinearity problem and its machine learning approach: a review. *Math. 10*(8). 1283. doi: 10.3390/math10081283.

[27] Fdez-Díaz, L.; Quevedo, J.R.; Montañés, E. 2023. Regularized boosting with an increasing coefficient magnitude stop criterion as meta-learner in hyperparameter optimization stacking en-semble. *Neurocomputing*. 551. 126516. doi: 10.1016/j.neucom.2023.126516.

[28] Toledo-Alvarado, H.O.; Ruíz-López, F.J.; Vásquez-Peláez, C.G.; Berruecos-Villalobos, J.M.; Elzo, M.A. 2014. Parámetros genéticos para producción de leche de ganado Holstein en dos modalidades de control de producción. *Rev. Mex. Cienc. Pecu. 5*(4). 443-457. https://www.scielo.org.mx/pdf/rmcp/v5n4/v5n4a5.pdf

[29] Cook, N.; Oetzel, G.; Nordlund, K. 2006. Modern techniques for monitoring high-producing dairy cows 1. Principles of herd-level diagnoses. *In Pract. 28*(9). 510-515. doi: 10.1136/inpract.28.9.510.

[30] García-Ruiz, A.; Cole, J.B.; VanRaden, P.M.; Wiggans, G.R.; Ruiz-López, F.J.; Van Tassell, C.P. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U.S.A. 113*(28). doi: 10.1073/pnas.1519061113.

[31] Green, L.E.; Schukken, Y.H.; Green, M.J. 2006. On distinguishing cause and consequence: Do high somatic cell counts lead to lower milk yield or does high milk yield lead to lower somatic cell count? *Prev. Vet. Med. 76*(1-2). 74-89. doi: 10.1016/j.prevetmed.2006.04.012.