Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
32 (2): 207-218, 2026
DOI: 10.9775/kvfd.2025.35390

Research Article

Evaluation of 305-Day Lactation Milk Yield Predictions from Pre-Peak Partial Milk Yields Using Some Data Mining Algorithms

Özcan ŞAHİN [1 (*)], Rabia ALBAYRAK DELİALİOĞLU [2], Gizem ÇİNİ [1], İbrahim AYTEKİN [1], Yasin ALTAY [3,4]

[1] Selcuk University, Faculty of Agriculture, Animal Science Department, TR-42250 Konya - TÜRKİYE
[2] Ankara University, Faculty of Agriculture, Animal Science Department, Biometry and Genetics Unit, TR-06100 Ankara - TÜRKİYE
[3] Eskisehir Osmangazi University, Faculty of Agriculture, Animal Science Department, Biometry and Genetics Unit, TR-26040 Eskişehir - TÜRKİYE
[4] Eskisehir Osmangazi University, Agricultural Studies Practices and Research Center, TR-26040 Eskişehir - TÜRKİYE

[(*)] **Corresponding author:**
Özcan Şahin
Phone: +90 332 223 2815
Cellular phone: +90 532 675 4184
Fax: +90 332 241 0108
E-mail: osahin006@gmail.com

How to cite this article?

Şahin Ö, Albayrak Delialioğlu R, Çini G, Aytekin İ, Altay Y: Evaluation of 305-Day Lactation Milk Yield Predictions from Pre-Peak Partial Milk Yields Using Some Data Mining Algorithms. *Kafkas Univ Vet Fak Derg*, 32 (2): 207-218, 2026
DOI: 10.9775/kvfd.2025.35390

Article ID: KVFD-2025-35390
Received: 30.09.2025
Accepted: 23.02.2026
Published Online: 24.02.2026

Abstract

A total of 75,640 test-day milk yield records of 248 Holstein cows in the first (124 heads), second (75 heads), and third lactation (49 heads) were used as material in the study. All data used in this study were obtained from the database of the Afikim herd management software used on a private dairy farm. To predict 305-day adjusted milk yields (MY305) using some partial milk yield parameters, ALM (Automatic Linear Modeling), C&RT (Classification and Regression Tree), CHAID (Chi-square Automatic Interaction Detector), RF (Random Forest), MARS (Multiple Adaptive Regression Splines), Bagging MARS (Bootstrap Aggregating Multiple Adaptive Regression Splines), and BRNN (Bayesian Regularized Neural Network) data mining algorithms were used with group five-fold cross-validation. When all algorithms are compared in terms of 15 different prediction performance measures, the most successful algorithms are MARS (R^2_{Adj} = 0.844, RRMSE = 6.530 and MAPE = 5.182), Bagging MARS (R^2_{Adj} = 0.840, RRMSE = 6.547 and MAPE = 5.103), while C&RT (R^2_{Adj} = 0.828, RRMSE = 7.028 and MAPE = 5.542) is the most efficient tree-based algorithm. When the model evaluation criteria, including systematic bias and limits of agreement (LoA) among prediction performance measures, were examined together, the prediction success of the data mining algorithms was determined as MARS, Bagging MARS, C&RT, ALM, BRNN, CHAID, and RF, respectively. As a result, it can be stated that 75-day partial milk yield totals before peak milk yield is an important time period and an indirect selection criterion in determining 305-day milk yield. Additionally, it can help producers evaluate the impact of past milk yields on future cow productivity and predict overall herd performance, thereby facilitating timely and informed decision-making.

Keywords: Bagging MARS, BRNN, C&RT, CHAID, Data mining, MARS, Milk yield, Partial milk record, Random forest

Introduction

Most of the total milk production in the world (about 80%) is obtained from cattle. Although the share of cattle in total milk production varies depending on regions and production conditions, it also constitutes a large part of the income sources of enterprises. In Türkiye, most of the milk production is obtained from cattle and approximately 92% of the milk produced is obtained from cattle. The remaining approximately 8% is made up of sheep, goat and buffalo milk [1]. Milk production is the process that starts after calving and continues until the animal is dry. This physiological period is defined as the lactation length and is accepted as 305 days on average in dairy cattle [2]. The curve showing the increasing and decreasing changes in milk yield during this period, which is shaped by the effects of genetic and environmental factors, is called the "lactation curve" [2,3].

With good herd management, there is a rapid increase in milk yield after birth and milk yield reaches its maximum level (peak) within 4-6 weeks. The peak period continues for a certain period of time and then milk production gradually decreases at a lower rate than the postpartum increase. But, peak yield and day to peak can vary

depending on the genetic and herd management factors [3,4]. The success of the enterprises in increasing their milk yields is primarily to keep records and to make optimization in terms of the factors in question thanks to these records. Using the records accurately and in the easiest way facilitates herd management. Keeping these records daily defines the actual lactation milk yield. However, different lactation milk yield prediction methods have been developed so far in order to predict the actual lactation milk yield in terms of time and convenience. In addition, both 305-day milk yields and actual lactation milk yields have been estimated by using partial lactation milk records [5,6]. Another importance of estimating lactation milk yields from partial lactation milk yields is that the animals that will not be used for breeding in the herd are not kept waiting until the first lactation milk yields are determined and the necessary labor force is disabled by eliminating the care and feeding costs to be applied to them. In addition, it is also possible to eliminate the risks of health protection measures (vaccination and disease etc.) of these animals. In other words, estimating the lactation yields of animals based on partial milk yield records enables indirect selection, which can shorten the generation interval by approximately one year. With this application in herd management, selection efficiency is increased. In brief, because milk yield is a quantitative trait, it is easily affected by environmental factors. Therefore, unlike past production approaches, modern dairy producers tend to identify and intervene in the factors affecting milk yield as quickly as possible. This necessitates analyzing and interpreting shorter timeframes and optimizing production processes to ensure sustainability in a competitive market.

Today, alongside the advancement of artificial intelligence technologies, digital tools and systems such as Smaxtech technology [7] that incorporate these algorithms have increasingly found their place in animal husbandry. Therefore, it is important to develop alternative calculation methods to the classical methods used in the calculation of lactation milk yields. Numerous prediction models have taken their place in the literature in terms of determining the traits that have been emphasized until today and evaluating them in animal breeding [2,5,6,8]. Classical regression methods and traditional lactation curve models require certain preconditions, such as linearity, exogeneity, heteroskedasticity, and autocorrelation. However, data mining algorithm applications can model nonlinear and complex interactions in the data structure without any preconditions. In recent years, studies on data mining have been carried out in many livestock breeding areas because they have some advantages due to practical, accurate, successful, and appropriate fit criteria as well as the ability to use larger data sets; prediction of end-of-fattening body weight [9], determination of mastitis with thermal camera [10], body weight prediction in goats [11], body weight prediction in sheep [12], prediction of body measurements in camels [13], prediction of body weight from some body measurements of cattle during growth and development [14], prediction of body weight by biometric measurements [15], prediction of milk yield [16], and honey yield and quality [17,18] were used. Although classical statistical methods are still widely used across various fields, they may fall short when it comes to analyzing large and complex datasets, especially with the rapid advancement of technology. In this context, data mining algorithms have gained prominence, as they are capable of effectively analyzing datasets that traditional statistical methods cannot handle [14,19].

This study aimed to evaluate of 305-day adjusted milk yields (MY_{305}) predictions using the total and mean of pre-peak 15, 30, 45, 60, and 75-day partial milk yields in different Lactation Number (LN) 248 Holstein cows via some data mining algorithms such as Automatic Linear Modeling (ALM), Classification and Regression Tree (C&RT), Chi-squared Automatic Interaction Detector (CHAID), Random Forest (RF), Multiple Adaptive Regression Splines (MARS), Bootstrap Aggregating Multiple Adaptive Regression Splines (Bagging MARS) and Bayesian regularized neural network (BRNN).

Material And Methods

Experimental Animals

The animal material of the study consisted of a total of 75,640 test-day milk yield records of 248 Holstein cows (mean LMY305=8,765 /. and mean days to peak milk yield=68 days) in the first (124 heads), second (75 heads), and third lactation (49 heads) of a private dairy farm in the Ilgın district of Konya province. All data used in the study were obtained from the database of the Afikim herd management software (AfiFarm v. 3.0) used at the farm. The farm used an 8x2 herringbone milking system and milking was done twice a day. In order to standardize milk yields to a 305-day lactation basis for cows with days in milk shorter or longer than 305 days, the correction factors recommended by Akman and Eliçin [20] were applied.

Prediction Methods

In this study, three regression tree-based algorithms (C&RT, CHAID, and RF) and four different data mining algorithms (ALM, MARS, Bagging MARS, and BRNN) were employed.

ALM Algorithm

The ALM algorithm, implemented as an Automated Linear Modeling procedure, is used in multiple regression analysis and data mining to determine the most suitable linear

model by automatically selecting a subset of predictors when a large number of independent variables are available. By automating variable selection and model optimization, the ALM algorithm facilitates the model-building process and can improve predictive performance [21].

C&RT Algorithm

C&RT algorithm, developed by Breiman et al.[22], is one of the most widely used decision tree algorithms. This algorithm utilizes a binary tree structure, where each node results in only two branches. The C&RT algorithm builds a classification model when the dependent variable is categorical and a regression model when the dependent variable is continuous. It generally uses the Gini index as the branching criterion [23]. The Gini index measures how effectively a node in a decision tree separates individuals into different classes. This metric ranges from 0 to 1, where 0 indicates perfect separation, representing a completely homogeneous (pure) group, and 1 represents complete heterogeneity. The lower the Gini index, the more homogeneous the node is [22,24,25].

CHAID Algorithm

Chi-squared Automatic Interaction Detector (CHAID) algorithm, introduced to the literature by Kass [26], is a tree-based method that constructs decision trees by employing the chi-square statistic to achieve optimal splits. The split decisions are made based on Bonferroni-adjusted p-values. One of the key characteristics of the CHAID algorithm is its ability to handle categorical and ordinal variables, as well as variables with missing data. Furthermore, as a non-parametric method that does not rely on linear assumptions, CHAID is particularly effective in multidimensional datasets [26], Unlike C&RT, this algorithm allows for more than two branches to emerge from a single node [27,28].

RF Algorithm

RF algorithm, developed by Breiman [29], is widely used for both classification and regression problems. It combines the predictions of multiple decision trees to obtain a single aggregated result. The fundamental difference between the RF algorithm and a single decision tree algorithm lies in the randomization of the processes involved in bootstrap sampling and in selecting a random subset of predictors at each split [30].

MARS Algorithm

The MARS algorithm, developed by Friedman [31], is a non-parametric regression method that identifies the relationship between dependent and independent variables. Through a stepwise process, it generates basis functions by considering candidate knots and potential interaction terms [15]. These basis functions are typically defined as piecewise linear functions and are chosen in such a way as to minimize the error variance at each step. Moreover, MARS is capable of modeling interactions among predictors by including interaction terms in the model [32,33].

Bagging MARS Algorithm

The Bagging MARS algorithm is a method that combines Multivariate Adaptive Regression Splines (MARS) with the Bootstrap Aggregating (Bagging) technique. The primary objective of this method is to improve the accuracy of the final model by modeling nonlinear relationships in regression analysis, while also addressing the issue of overfitting. Bagging MARS generates multiple MARS models, each trained on different bootstrap samples, and aggregates their predictions. This approach, particularly effective with high-dimensional and complex data, reduces overfitting while simultaneously enhancing the model's generalization ability, thus leading to more reliable results [34,35].

BRNN Algorithm

The Bayesian Regularized Neural Network (BRNN) algorithm is a method used to model complex relationships within data, combining artificial neural networks (ANNs) with Bayesian regularization techniques. This approach regulates the learning process by assigning prior probability distributions to the network weights, thereby providing better generalization in both regression and classification tasks. Additionally, it reduces overfitting and delivers effective results, especially with noisy datasets. Compared to other methods, BRNN has proven highly reliable in improving neural network performance, particularly with complex and nonlinear datasets [36].

Prediction Performance Evaluation Criteria of Data Mining Algorithms

The prediction performance criteria of the algorithms used in the study are presented in *Table 1*.

In evaluating the predictive performance of the algorithms, smaller values are expected for Akaike's Information Criterion (AIC), corrected Akaike's Information Criterion (AICC), mean error (ME), mean absolute percentage error (MAPE), mean relative approximation error (MRAE), mean absolute deviation (MAD), global relative approximation error (RAE), standard deviation ratio (SDratio), root mean square error (RMSE), and relative root mean square error (RRMSE). In contrast, the coefficient of determination (R^2) and the adjusted coefficient of determination (Adj. R^2) should take values close to 1 [37-39]. Also, Bland-Altman analysis was used to determine the systematic bias and limits of agreement (LoA) between the predictions of data mining algorithms and the 305-day adjusted milk yield.

Table 1. *Performance criteria and formulae of the algorithms*

Performance Criteria	Experission
Root mean square error	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2}$
Relative root mean square error	$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2}}{\overline{y}} x100$
Standard deviation ratio	$SD_{ratio} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y_i})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{ip} - \overline{y_{ip}})^2}}$
Coefficient of variation	$CV(\%) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2}}{\overline{Y}} * 100$
Pearson's correlation coefficients	$PC = \frac{cov(y_i, y_{ip})}{S_{y_i} S_{y_{ip}}}$
Performance index	$P = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2}}{(1+r)\frac{1}{n}\sum_{i=1}^{n} y_i} * 100$
Mean error	$ME = \frac{1}{n}\sum_{i=1}^{n} y_i - y_{ip} \vee$
Relative approximation error	$RAE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_{ip})^2}{\sum_{i=1}^{n} y_i^2}}$
Mean relative approximation error	$MRAE = \sqrt{\frac{1}{n}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} y_i^2}}$
Mean absolute percentage error	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - y_{ip}}{y_i}\right\rvert x100$
Mean absolute deviation	$MAD = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - y_{ip}\rvert$
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_{ip})^2}{\sum_{i=1}^{n}(y_{ip} - \overline{y_{ip}})^2}$
Adjusted coefficient of determination	$AdjR^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_{ip})^2/(n-1)}{\sum_{i=1}^{n}(y_{ip} - \overline{y_{ip}})^2/(n-p-1)}$
Akaike's information Criterion	$AIC = n.\in\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2\right] + 2k$
Corrected Akaike's information criterion	if n/k>40, or $AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$

Statistical Analysis

In the study, a total of seven different data mining algorithms were used, three of which were regression tree-based. Differences among the means of the lactation number (LN) levels for the traits considered in the study were tested using one-way ANOVA. Additionally, the Tukey HSD test was used to identify which LN levels differed significantly. Also, the differences between the 305-day adjusted milk yield and the predictions of each data mining algorithm were compared using a paired t-test. All analyses followed a group 5-fold cross-validation procedure using an individual animal-based strategy rather than an observation-based one. The ALM, C&RT and CHAID algorithms were performed using IBM SPSS Statistics (v23) [40]. In the C&RT and CHAID regression tree algorithms, the minimum parent node size was set to 10 and the minimum child node size to 5. In the RF algorithm, the number of trees (ntree) and mtry were set to 100 and 5, respectively, using the 'randomForest (v4.7.1.1)' package in R [41,42]. For the MARS and Bagging MARS algorithms, the 'caret (v6.0.94)', 'earth (v5.3.4)', and 'mda (v0.5-5)' R packages were used, while the 'brnn (v0.9.0)' package was used to predict MY305 using the BRNN algorithm with three neurons and a tangent hyperbolic activation function [43-45]. Finally, the R package 'ehaGoF (v0.1.1)' was used to calculate the goodness-of-fit criteria [46]. Furthermore, a Bland-Altman analysis was conducted at 95% confidence intervals to evaluate the statistical agreement between data mining algorithms and 305-day adjusted milk yield predictions.

Results

In the present study, some descriptive statistics of partial total ($PART_{15}$, $PART_{30}$, $PART_{45}$, $PART_{60}$, $PART_{75}$) and mean traits of Holstein cows in the first (124 heads), second (75 heads), and third lactation (49 heads) before peak milk yield are presented in *Table 2*.

According to *Table 2*, it is evident that as the number of lactation increases, both total milk yield and partial period milk yields rise significantly (P<0.01). It is an anticipated trend in dairy production that, with increasing LN, the milk yield potential of cows improves until adulthood. Average milk yields follow the same trend, reaching their highest levels particularly in the third lactation. This increase can be attributed to the improvement of cows' physiological capacity with age and the enhanced efficiency of mammary tissue function. Furthermore, the lower coefficient of variation (CV) observed in the third lactation indicates that milk yields are more homogeneous at this stage. Therefore, cows in their third lactation not only achieve higher but also more stable milk production.

C&RT Algorithm

The regression tree diagram of the C&RT algorithm with group 5-fold cross-validation and MY_{305} prediction is given in *Fig. 1*.

Table 2. Descriptive statistics of partial total and average milk yields of Holstein cows in different lactations

Variables	LN	N	Min	Max	$\overline{X} \pm S_{\overline{x}}$	S_x	CV	P-value
MY_{305}	1	124	5345.00	11260.00	7552.00 ± 100.00^{C}	1118.00	14.81	<0.000**
	2	75	5917.00	11577.00	9115.00 ± 138.00^{B}	1199.00	13.15	
	3	49	7891.00	12360.00	9627.00 ± 146.00^{A}	1021.00	10.61	
$PART_{15}$	1	124	165.90	557.40	313.76 ± 5.75^{C}	64.02	20.40	<0.000**
	2	75	172.40	586.70	423.27 ± 9.58^{B}	83.00	19.61	
	3	49	291.10	611.00	462.20 ± 10.30^{A}	72.20	15.61	
$PARTM_{15}$	1	124	11.06	37.16	20.92 ± 0.38^{C}	4.27	20.40	<0.000**
	2	75	11.49	39.11	28.22 ± 0.64^{B}	5.53	19.61	
	3	49	19.41	40.73	30.81 ± 0.69^{A}	4.81	15.61	
$PART_{30}$	1	124	396.90	1186.00	686.20 ± 11.70^{C}	129.90	18.92	<0.000**
	2	75	451.20	1299.30	948.80 ± 19.10^{B}	165.00	17.39	
	3	49	667.50	1370.30	1037.40 ± 21.50^{A}	150.70	14.53	
$PARTM_{30}$	1	124	13.23	39.53	22.87 ± 0.39^{C}	4.33	18.92	<0.000**
	2	75	15.04	43.31	31.63 ± 0.64^{B}	5.50	17.39	
	3	49	22.25	45.68	34.58 ± 0.72^{A}	5.02	14.53	
$PART_{45}$	1	124	661.90	1803.00	1092.30 ± 17.20^{C}	192.10	17.58	<0.000**
	2	75	770.50	2039.50	1502.00 ± 27.50^{B}	238.50	15.88	
	3	49	1114.80	2172.30	1643.00 ± 32.40^{A}	226.70	13.80	
$PARTM_{45}$	1	124	14.71	40.07	24.27 ± 0.38^{C}	4.27	17.58	<0.000**
	2	75	17.12	45.32	33.38 ± 0.61^{B}	5.30	15.88	
	3	49	24.77	48.27	36.51 ± 0.72^{A}	5.04	13.80	
$PART_{60}$	1	124	946.40	2457.00	1514.70 ± 22.60^{C}	251.70	16.62	<0.000**
	2	75	1096.20	2680.00	2049.50 ± 35.50^{B}	307.50	15.00	
	3	49	1627.20	2894.80	2243.90 ± 41.40^{A}	289.70	12.91	
$PARTM_{60}$	1	124	15.77	40.95	25.25 ± 0.38^{C}	4.20	16.62	<0.000**
	2	75	18.27	44.67	34.16 ± 0.59^{B}	5.13	15.00	
	3	49	27.12	48.25	37.40 ± 0.69^{A}	4.83	12.91	
$PART_{75}$	1	124	1257.00	3091.60	1935.70 ± 27.50^{C}	305.90	15.80	<0.000**
	2	75	1438.10	3311.20	2583.40 ± 43.20^{B}	374.30	14.49	
	3	49	2185.10	3588.90	2830.90 ± 48.50^{A}	339.40	11.99	
$PARTM_{75}$	1	124	16.76	41.22	25.81 ± 0.37^{C}	4.08	15.80	<0.000**
	2	75	19.18	44.15	34.45 ± 0.58^{B}	4.99	14.49	
	3	49	29.14	47.85	37.75 ± 0.65^{A}	4.53	11.99	

***(P<0.01; A, B, C; LN: Lactation Number; MY_{305}: 305-day milk yield; PART: Partial milk total; PARTM: Partial milk mean*

The C&RT algorithm generated a total of 24 homogeneous subsets and 5 main branches for the MY_{305} prediction. The most important independent variables are the partial sum $PART_{75}$, $PART_{15}$, $PART_{60}$ and $PART_{45}$, respectively. In the prediction model, partial milk yield means and $PART_{30}$ variables were eliminated by the algorithm because they were not effective in MY_{305} prediction. When the regression tree diagram is carefully analyzed, it is determined that the most important independent variable in MY_{305} prediction is $PART_{75}$. Because 7 of the 12 branches in the tree diagram belong to the $PART_{75}$ feature. In other words, the independent variable $PART_{75}$ contributes more to the MY_{305} prediction than other variables.

The root node (Node 0) is branched into less (Node 1) and more (Node 2) than 2367.75 *l* of $PART_{75}$ feature. It was estimated that if the total milk yield ($PART_{75}$) of lactating animals in the first 75 days was less than 2367.75 *l*, 7436.818 *l*, and if it was more than 2367.75 *l*, the animals

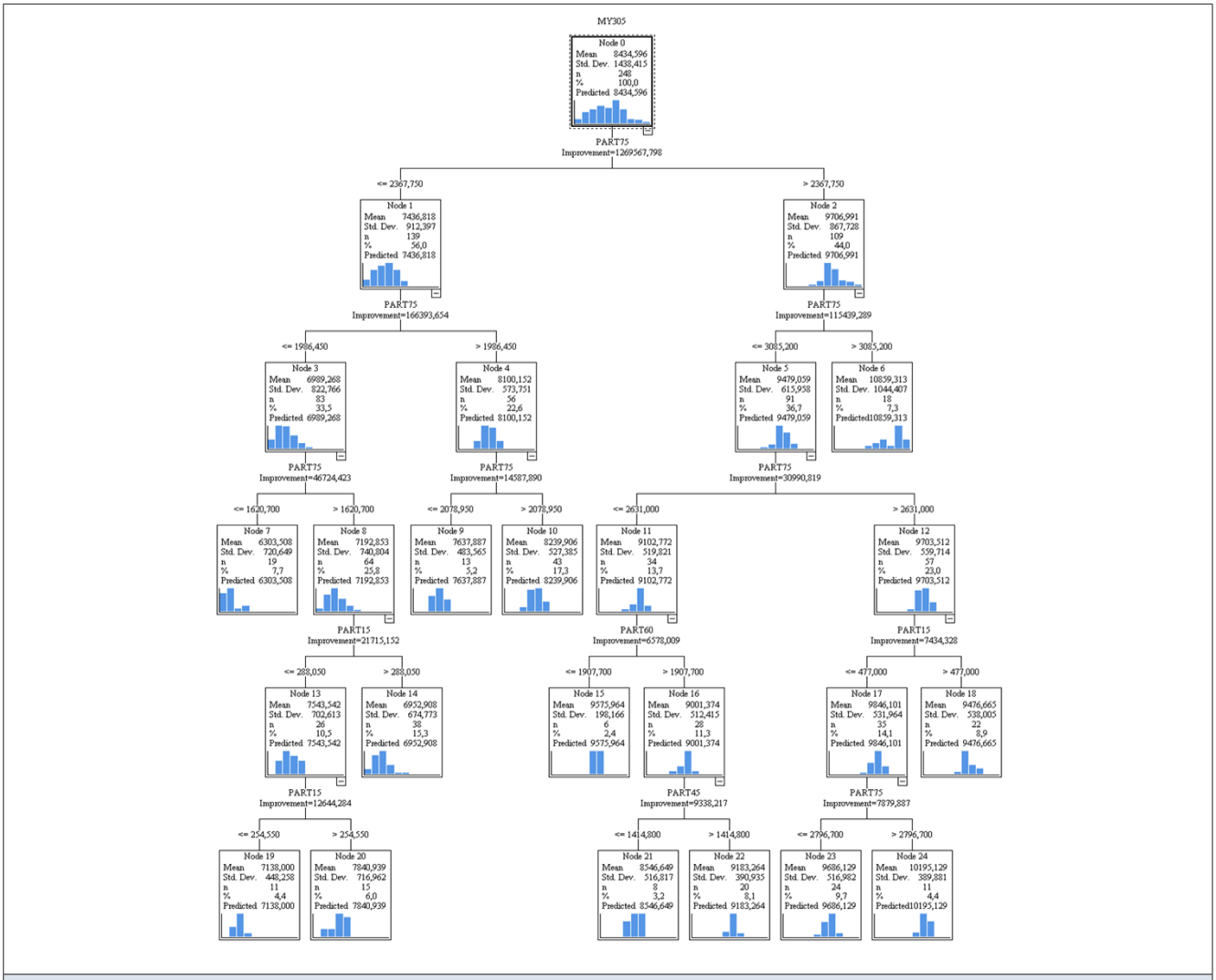

Fig 1. Classification tree diagrams constructed by C&RT for MY_{305} (Group cross-validation 5 parent 10 child node 5)

would reach a milk yield of 9706.991 *l*. In the regression diagram, the highest MY_{305} prediction occurred at node 24. If the total milk yield in the first 15 days of lactation ($PART_{15}$) was lower than 477 *l* and the total milk yield in the first 75 days ($PART_{75}$) was higher than 2796.70 *l*, it was predicted to reach 10195.129 *l* milk yield after 305 days (Node 24). The lowest MY_{305} was estimated as 6303.508 *l* when total milk yield in the first 75 days ($PART_{75}$) was lower than 1620.70 *l* (Node 7).

MARS Algorithm

MY_{305}, the parameter results of the model obtained using the MARS algorithm are given in *Table 3*.

In the MY_{305} MARS algorithm prediction equation, $PART_{75}$, LN_3, $PART_{15}$, $PARTM_{15}$, $PARTM_{15}$, $PARTM_{75}$, $PARTM_{45}$, $PART_{60}$ and $PARTM_{60}$ independent variables and 12 basis functions were used. The MARS algorithm does not use a single variable as the main variable but uses binary and triple interactions. In the MARS prediction equation, the basis functions BF_5, BF_7, BF_8, BF_{10} and

Table 3. *MARS algorithm prediction equation of MY_{305}*

Functions	Terms	Coefficients
BF_1	Intercept	8930.00
BF_2	max(0, $PART_{75}$ - 2536) x LN_3	-1.180
BF_3	max(0, 546 - $PART_{15}$) x max(0, $PARTM_{15}$ - 31.3)	-5.650
BF_4	max(0, 546 - $PART_{15}$) x max(0, 25.9 - $PARTM_{75}$)	-4.270
BF_5	max(0, 25.1 - $PARTM_{45}$) x max(0, 2536 - $PART_{75}$)	+0.425
BF_6	max(0, $PART_{60}$ - 1219) x max(0, $PARTM_{60}$ - 31.8)	-0.314
BF_7	max(0, $PART_{60}$ - 1219) x max(0, 31.8 - $PARTM_{60}$)	+1.250
BF_8	max(0, $PART_{60}$ -1219) x max(0, $PART_{75}$ - 2404)	+0.00646
BF_9	max(0, $PART_{60}$ - 1219) x max(0, 2404 - $PART_{75}$)	-0.0262
BF_{10}	max(0, 546 - $PART_{15}$) x max(0, 31.3 - $PARTM_{15}$) x max(0, 1522 - $PART_{60}$)	+0.00127
BF_{11}	max(0, 25.1- $PARTM_{45}$) x max(0, $PART_{60}$ - 1327) x max(0, 2536 - $PART_{75}$)	+0.00757
BF_{12}	max(0, 25.1- $PARTM_{45}$) x max(0, 2536 - $PART_{75}$) x max(0, $PARTM_{75}$ - 22.8)	-0.493

BF_{11} contributed positively to MY_{305}, while the remaining basis functions had a negative effect. For the MARS algorithm to work, the basis functions must take values within the range of the specified constraints. Otherwise, it does not affect the dependent variable (MY_{305}) and can be considered as a zero contribution. As can be seen from *Table 3*, the MARS algorithm, in the BF1 function (Intercept), predicted the MY_{305} milk yield to be 8930 *l* if other functions were not working. In BF_2, it was predicted that the MY_{305} milk yield would decrease if the total milk yield of animals in their third lactation was less than 2536 *l* in the first 75 days. A total milk yields greater than 2536 *l* in the first 75 days of animals in their third lactation can be considered an indirect selection criterion for the MY_{305} trait. Since the regression coefficient in BF_3 is negative (-5.650), regardless of the number of lactation, it is desired that the total milk yield of animals in the first 15 days be greater than 546 *l*, while the average milk yield in the first 15 days is expected to be greater than 31.3 *l*. The cutoff points determined for total and average milk yields in the first 15 days for dairy cows can be used as practical information in herd management and feeding. In short, the MARS algorithm prediction equation of MY_{305} (*Table 3*) was given that the product of the coefficient and the related variables indicates the contribution of that term to the estimated milk yield. To avoid the complexity of the MARS algorithm model and to make it more understandable, the MY_{305} estimate is made in *Table 4* using the independent variables of a random dairy cow in the dataset. The independent variables of the randomly selected 3rd lactation (LN_3) dairy cattle were $PART_{15}$=442.70 *l*, $PART_{30}$=979.60 *l*, $PART_{45}$=1514.70 *l*, $PART_{60}$=2086.50 *l*, $PART_{75}$=2696.50 *l* and $PARTM_{15}$=29.51 *l*, $PARTM_{30}$=32.65 *l*, $PARTM_{45}$=33.66 *l*, $PARTM_{60}$=34.78 *l*, and $PARTM_{75}$=35.95 *l*, respectively. Accordingly, the MARS algorithm MY_{305} prediction calculation process details are given in *Table 4*.

While the MY_{305} of the dairy cattle randomly selected from the data set was 9799.92 *l* in reality, it was estimated as 9568.05 liters as a result of the MARS algorithm. In making this prediction, it was found that the MARS algorithm made predictions using only BF_1, BF_2, BF_6 and BF_8 basis functions.

Variable Importance Results of Data Mining Algorithms

Table 5 shows the sensitivity analysis results for each

Table 4. *MARS algorithm prediction of MY_{305} of a random dairy cattle selected from dataset*

Functions	Algorithm Terms and Calculations	Coefficients	Contribution to MY_{305}
BF_1*	Intercept	**8930.00**	**8930.00**
BF_2*	(0, $PART_{75}$ - 2536) * LN_3 (0, 2696.50 - 2536) * 1 = 160.50*1 = **160.50**	**- 1.180**	- 1.180*160.50= - **189.39**
BF_3	(0, 546 - $PART_{15}$) * (0, $PARTM_{15}$ - 31.3) (0, 546 - 442.70) * (**0**, 25.84 – 29.51) = 103.30*0 = 0	- 5.650	- 5.650*0=0
BF_4	(0, 546 - $PART_{15}$) * (0, 25.9 - $PARTM_{75}$) (0, 546 - 442.70) * (**0**, 25.9 – 35.95) = 103.30*0 = 0	- 4.270	- 4.270*0=0
BF_5	(0, 25.1 - $PARTM_{45}$) * (0, 2536 - $PART_{75}$) (**0**, 25.1 – 33.66) * (**0**, 2536 - 2696.50) = 0*0 = 0	+ 0.425	+ 0.425*0=0
BF_6*	(0, $PART_{60}$ - 1219) * (0, $PARTM_{60}$ - 31.8) (0, 2086.50 - 1219) * (0, 34.78 - 31.8) = 867.50*2.98 = **2585.15**	**- 0.314**	- 0.314*2585.15 = - **811.74**
BF_7	(0, $PART_{60}$ - 1219) * (0, 31.8 - $PARTM_{60}$) (0, 2086.50 - 1219) * (**0**, 31.8 – 34.78) = 867.50*0 = 0	+ 1.250	+ 1.250*0 = 0
BF_8*	(0, $PART_{60}$ -1219) * (0, $PART_{75}$ - 2404) (0, 2086.50 - 1219) * (0, 2696.50 - 2404) = 867.50*292.50 = **253743.75**	**+ 0.00646**	+ 0.00646*253743.75= + **1639.18**
BF_9	(0, $PART_{60}$ - 1219) * (0, 2404 - $PART_{75}$) (0, 2086.50 - 1219) * (**0**, 2404 - 2696.50) = 867.50 * 0 = 0	- 0.0262	- 0.0262*0 = 0
BF_{10}	(0, 546 - $PART_{15}$) * (0, 31.3 - $PARTM_{15}$) * (0, 1522 - $PART_{60}$) (0, 546 - 442.70) * (0, 31.3 - 29.51) * (**0**, 1522 - 2086.50) = 103.30*1.79*0 = 0	+ 0.00127	+ 0.00127*0 = 0
BF_{11}	(0, 25.1- $PARTM_{45}$) * (0, $PART_{60}$ - 1327) * (0, 2536 - $PART_{75}$) (**0**, 25.1- 33.66) * (0, 2086.50 - 1327) * (**0**, 2536 – 2696.50) = 0*759.50*0 = 0	+ 0.00757	+ 0.00757*0 = 0
BF_{12}	(0, 25.1- $PARTM_{45}$) * (0, 2536 - $PART_{75}$) * (0, $PARTM_{75}$ - 22.8) (**0**, 25.1- 33.66) * (**0**, 2536 - 2696.50) * (0, 35.95 - 22.8) = 0*0*13.15 = 0	- 0.493	- 0.493*0 = 0
	Prediction	**= 8930.00 - 189.39 - 811.74 + 1639.18 = 9568.05**	

** The basic functions and operations used in MY_{305} prediction calculation were given in bold*

data mining algorithm in order to predict the variable importance values of the independent variables of MY_{305}.

In MY_{305}, $PART_{75}$ independent variable had the highest relative importance value in ALM (61.90%) algorithm, followed by MARS (41.49%) and BRNN (12.34%) algorithms. In $PARTM_{30}$ independent variable, Bagging had the highest relative importance value with MARS (28.87%), while RF (14.21%) algorithm had the highest relative importance value in $PARTM_{75}$. In addition, the highest and lowest percentages of the independent variables included in the model of ALM, RF, MARS, Bagging MARS and BRNN algorithms used in the current study were 61.90-5.30%, 14.21-5.21%, 41.49-3.19%, 28.87-0.01% and 12.34-6.79%, respectively. ALM ($PART_{15}$, $PARTM_{15}$, $PART_{30}$, $PARTM_{30}$, $PART_{45}$, $PARTM_{60}$ and $PARTM_{75}$), MARS ($PART_{30}$, $PARTM_{30}$, $PART_{45}$ and $PARTM_{45}$), and BRNN algorithm (LN) independent variables were included in the model, while all independent variables were included in the model in RF and Bagging MARS algorithms.

Table 5. *Variable importance of data mining algorithms*

Variables	ALM	RF	MARS	Bagging MARS	BRNN
$PART_{15}$	0.00%	10.35%	3.19%	3.86%	6.79%
$PARTM_{15}$	0.00%	7.62%	4.63%	0.26%	6.79%
$PART_{30}$	0.00%	5.21%	0.00%	8.66%	9.13%
$PARTM_{30}$	0.00%	6.24%	0.00%	28.87%	9.13%
$PART_{45}$	0.00%	6.82%	0.00%	13.98%	10.31%
$PARTM_{45}$	0.00%	6.87%	0.00%	0.31%	10.31%
$PART_{60}$	32.80%	10.91%	32.34%	19.19%	11.43%
$PARTM_{60}$	0.00%	11.50%	9.46%	0.32%	11.43%
$PART_{75}$	61.90%	14.16%	41.49%	24.22%	12.34%
$PARTM_{75}$	0.00%	14.21%	4.24%	0.32%	12.34%
LN	5.30%	6.13%	4.65%	0.01%	0.00%

Prediction Performances of Data Mining Algorithms

The performance criteria values of data mining algorithms are presented in *Table 6*.

The difference between the 305-day adjusted milk yield and the predictions of the data mining algorithms was found to be statistically nonsignificant (P>0.05). According to *Table 6*, we found that MARS, Bagging MARS, C&RT, ALM, BRNN, CHAID, and RF were the data mining algorithms with the best prediction performance when we combined the model evaluation criteria, which include systematic bias and limits of agreement (LoA) among prediction performance indicators. When all algorithms are compared, the results obtained from the MARS algorithm of the algorithm with the best criteria in terms

Table 6. *Predictive performance of data mining algorithms*

Model Performance Criteria	ALM	C&RT	CHAID	RF	MARS	Bagging MARS	BRNN
RMSE	593.408	592.790	675.514	734.114	550.754	552.227	595.225
RRMSE	7.035	7.028	8.009	8.704	6.530	6.547	7.057
SDR	0.413	0.413	0.471	0.511	0.384	0.385	0.415
CV	7.050	7.040	8.030	8.720	6.540	6.560	7.070
r	0.911	0.911	0.882	0.860	0.923	0.923	0.910
PI	3.682	3.678	4.255	4.680	3.395	3.404	3.695
Bias (ME)	-1.782	-0.001	0.000	4.150	0.000	2.832	-1.645
RAE	0.005	0.005	0.006	0.007	0.004	0.004	0.005
MRAE	0.004	0.004	0.005	0.005	0.004	0.004	0.004
MAPE	5.680	5.542	6.530	6.924	5.182	5.103	5.730
MAD	469.352	453.917	537.991	570.093	427.008	421.737	472.576
R^2	0.829	0.829	0.779	0.738	0.853	0.852	0.828
R^2_{adj}	0.828	0.828	0.777	0.738	0.844	0.840	0.828
AIC	3171.398	3170.881	3235.675	3272.937	3158.399	3167.724	3168.914
AIC_c	3171.447	3170.930	3235.724	3272.937	3160.202	3170.711	3168.914
LoA	±1163.08	±1161.87	±1324.01	±1438.86	±1079.48	±1082.36	±1166.64
P-value	0.962	0.999	1.000	0.929	1.000	0.936	0.965

RMSE; Root mean square error, RRMSE; Relative root mean square error, SDR; Standard deviation ratio, CV; Coefficient of variation, r; Pearson's correlation coefficients, PI; Performance index, ME; Mean error; RAE, Relative approximation error, MRAE; Mean relative approximation error, MAPE; Mean absolute percentage error; MAD; Mean absolute deviation, R^2; Coefficient of determination, R^2_{adj}; Adjusted coefficient of determination, AICc; Akaike's information criterion, CAI_C; Corrected Akaike's information criterion, LoA; Lower-Upper limits of agreement

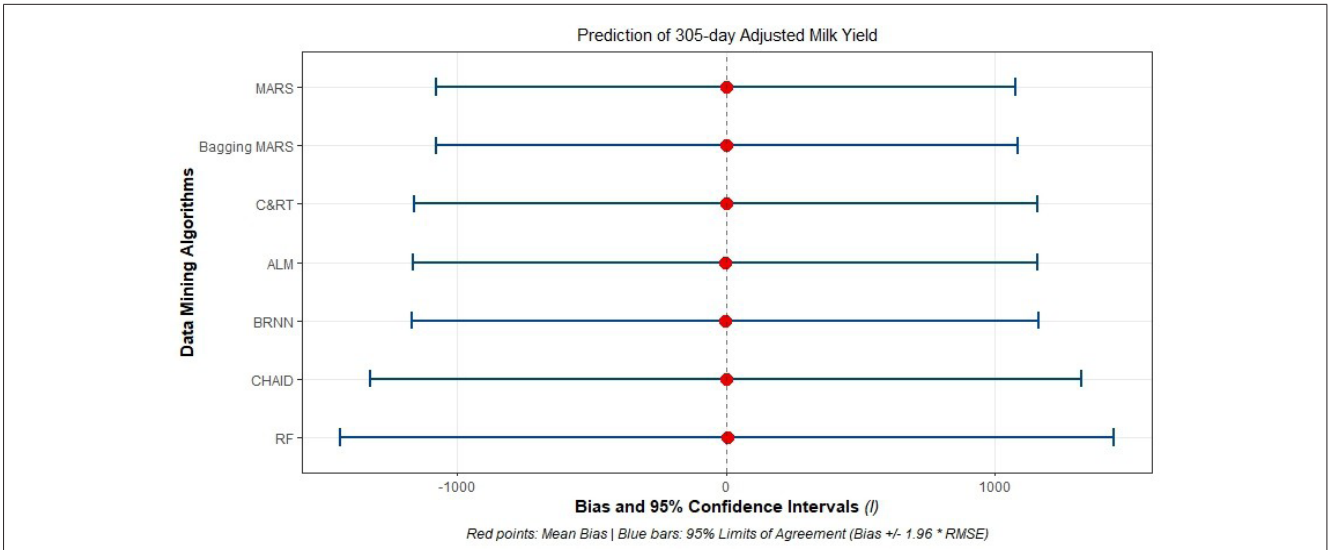

Fig 2. Bland-Altman statistical limits of agreement (LoA) for 305-day adjusted milk yield prediction by data mining algorithms

of performance criteria; high Pearson correlation coefficient (0.92), R^2 (85.3%), R^2_{adj} (84.4%), low MAPE (5.18%), SD ratio (0.38), RMSE (550.75), RAE (0. 004), CV (6.54%) and AIC (3158.39), while the Bagging MARS algorithm with the second best performance criteria was found as high Pearson correlation coefficient (0.92), R^2 (85.2%), R^2_{adj} (84.00%), low MAPE (5.10%), SD ratio (0.38), RMSE (552.22), RAE (0.004), CV (6.56%) and AIC (3167.72). In addition, the best performing tree algorithm is the C&RT algorithm with performance criteria of 0.91, 82.9%, 82.8%, 5.54%, 0.41, 592.79, 0.005, 7.04% and 3170.88 in the same order. These algorithms were followed by ALM, BRNN, CHAID and RF algorithms in terms of performance criteria.

The Bland-Altman statistical (LoA) for 305-day adjusted milk yield estimation using data mining algorithms are given in *Fig. 2*.

In predicting adjusted milk yield over 305 days, the MARS, CHAID, and C&RT algorithms were found to provide unbiased predictions as they were quite close to the zero point. The Bagging MARS and RF algorithms exhibited low positive bias, while the ALM and BRNN algorithms showed negative bias. When data mining algorithms were evaluated based on their prediction precision, the 95% confidence interval was found to be as follows: MARS, Bagging MARS, C&RT, ALM, BRNN, CHAID, and RF.

Discussion

Traditional methods for predicting milk yield are widely used in the literature [5,6]. Nowadays, studies are being focused on different algorithms other than traditional milk yield methods, but they are still limited. Studies evaluating the efficacy of novel methodologies in animal husbandry are currently insufficient. Traditional lactation curve models are defined by mathematical functions such as constant, exponential or gamma. These models are successful in plotting the average curve at the population level. However, they lack flexibility in capturing individual variations at the onset of lactation, abrupt peak deviations, and irregularities due to environmental factors. High variance and stochastic variations at the onset of lactation, which classical regression models cannot explain, can be predicted with a lower margin of error (RMSE) using machine learning algorithms [47]. However, these innovative approaches, which can replace complex traditional models, may play a crucial role in preventing anomalies in animal care and feeding management. For instance, continuous monitoring of milk yield at the individual or herd level not only provides detailed insights into health and nutritional requirements but also establishes a robust data foundation that supports decisions regarding culling or selection. Researchers are working intensively on algorithms with high predictive power that will help breeders make early and rapid decisions in herd management.

Salehi et al.[48] employed conventional artificial neural networks (ANNs) to classify monthly milk yield records into two categories such as high yield (≥9000 kg) and low yield (<9000 kg). The study compared two distinct classifier models one trained on records spanning the entire yield range, and another trained exclusively on records concentrated around the 9000 kg threshold. The results demonstrated that the classifier trained on threshold-specific records achieved markedly higher accuracy (99.7%) compared with the general classifier (92.3%). Grzesiak et al.[49], in the study of MY305 prediction using partial lactation records, examined the artificial neural networks (ANN) and multiple linear regression (MLR) model for cows in their 1st, 2nd or 3rd lactation and lactation

duration not less than 200 days. They found that $R^2 = 0.87$ for MLR and $R^2 = 0.85$ for the test set, and the partial regression coefficients were significant at $P \leq 0.01$. The mean prediction error (MEP) for 305-day lactation yield was determined as MLR (7.4%) and ANN (6.9%), and the mean milk yield for 305-day lactation predicted by ANN was 13.12 *l* lower than the mean actual yield of 49 control cows, while the mean difference in MLR was -91.33 *l*. The same researchers reported that the predictions made by neural network or multiple regression model were not different from the values predicted by the current traditional evaluation system ($P>0.05$). Sharma et al.[50] used an artificial neural network (ANN) to predict first lactation 305-day milk yield using partial lactation records of Karan Fries crossbred dairy cattle and reported that the best training algorithm was the 'Levenberg-Marquardt algorithm with Bayesian adjustment', which achieved more than 92% prediction accuracy, which was comparatively higher than the traditional MLR model. Eyduran et al.[11] evaluated the relationship between 305-day milk yield and various environmental factors (calving season, calving year, number of births, calving interval, and dry period) of 645 head Brown Swiss cattle from 1884 lactation records using the Regression Tree method. They found that year of calving was statistically the most effective factor on 305-day milk yield of Brown Swiss cattle, followed by parity and calving interval in the regression tree diagram ($P<0.01$). As a result, they reported that among the cows, those with more than 352 calving intervals had higher yields with an average of 3629.345 *l* and cows with more than 2 parities had higher yields than cows with 1 and 2 parities. Akıllı and Atıl [51] in their study titled evaluation of normalization techniques on artificial neural networks for prediction of 305-day milk yield in Holstein Friesen cows, comparatively tested eight normalization techniques and five different back propagation algorithms. They reported that the Bayesian Regularization algorithm and the Decimal Scaling normalization technique had the best performance ($R^2_{Adj} = 0.8181$, RMSE = 0.0068, MAPE = 160.42 for the test set; $R^2_{Adj} = 0.8141$, RMSE = 0.0067, MAPE = 114.12 for the validation set). Usman et al. [52] compared the performance of three different artificial neural network algorithms for the prediction of 305-day milk yield in the first lactation of 1092 head of Vrindavani crossbred cattle and used two different input sets in the analysis to predict milk yield. For the first input set, the R^2 values of Bayesian Regularization (BR), Levenberg Marquardt (LM) and Scaled Conjugate Gradient (SCG) algorithms were found to be 79.89%, 73.65% and 74.65% respectively, while the lowest RMSE values were 16.89, 20.52 and 20.45 respectively. For the second input set-2, the R^2 values of BR, LM and SCG algorithms were found to be 82.67%, 74.22% and 76.69% respectively, while the lowest RMSE values were reported as 14.45%, 17.45% and 16.56% respectively. As a result, they emphasized that BR algorithm can be used to predict 305-day milk yield in the first lactation in crossbred cattle since it shows higher accuracy than LM and SCG algorithms. Genç and Mendeş [21] reported that the most influential factors affecting 305-day milk yield were breed, lactation period, province, and parity. They found that the Automated Linear Model (ALM) achieved an accuracy rate of 64.2%. Furthermore, they emphasized that the ALM approach was particularly effective in identifying the determinants of the outcome variable, especially when dealing with large and complex datasets containing numerous predictors. Boğa [53] used MARS and Bagging MARS algorithms to create a lactation model for 305-day milk yield in dairy cattle. The prediction performance values for the Mars algorithm were r (0.988), R^2 (96.8%), Adj- R^2 (96.8%), low MAPE (1.374%), SD ratio (0.178), RMSE (10.204) and AIC (143. 073), while Bagging MARS algorithm had r (0.762), R^2 (43.6%), Adj- R^2 (43.5%), low MAPE (4.515%), SD ratio (0.751), RMSE (7.364) and AIC (735.927). The researcher reported that the MARS algorithm gave better results in modeling milk yield for 305-day lactation. Liseune et al.[54] suggested a model to predict the entire lactation curve of dairy cows by leveraging historical lactation milk yield information observed in the preceding cycle by using deep learning. They concluded that the ANN model, in addition to herd statistics, can predict the entire lactation curve of a cow in the next cycle by using the cow's past milk yield sequence and reproductive and health events from the previous cycle. This advantage of these algorithms also highlights their current applicability in herd management programs. Boğa et al.[16] investigated seasonal milk yield predictions using MARS algorithm in 157 head Holstein cross breeds. The three threshold values determined in the model were number of days milked (159 days), age (39.6 months), and peak yield (37.1 kg). They also determined that the number of days milked, average seven-day milk yield and lactation number are the most important variables in determining the prediction equation, and the most appropriate value for the prediction equation of the dependent variables is the winter milk yield variable.

In conclusion, in the studies conducted, the differences in the results of calculating or estimating real milk yields in both data mining and traditional methods vary depending on the size of the data used and the methods. However, in model comparisons, the criteria for the goodness of fit of RMSE, RRMSE, RAE, CV, MRAE, MAPE, SDR, and MAD should generally be very close to 0 and the closeness of r, R^2, and Adj- R^2 values to 1 should be taken into account. In this respect, comparisons should be made and the most appropriate algorithm and model should be selected. Furthermore, the bias and precision of algorithm predictions must also be considered. Among

the algorithms considered in the study, it can be said that the MARS and Bagging MARS algorithms were more preferable than the others. MARS and Bagging MARS algorithms can be integrated into existing herd management systems to overcome their complexities. Some herd management systems already have simple regression-based predictions in their software. Integrating these models into decision support tools within herd management software used in farms can support early decision-making mechanisms in herd management programs such as maintenance and feeding, thereby increasing dairy farm profitability. To the best of our knowledge, there are only a limited number of studies in the literature that employ algorithms based on partial lactation records, and this study represents the first to use such algorithms with partial lactation data. These findings may help producers assess the impact of historical milk yields on future cow productivity and predict overall herd performance, thereby facilitating timely and informed decision-making. Briefly, by using these algorithms in herd management, early culling of cows from the herd could facilitate the rapid integration of genetically superior individuals, thereby supporting increased productivity and offering the potential to enhance the herd's average performance. Furthermore, this practice accelerates genetic progress, too.

Declarations

Availability of Data and Materials: The datasets and analyzed during the current study available from the corresponding author (ÖŞ) on reasonable request.

Ethical Statement: As this study was conducted within the scope of non-experimental agricultural practices, ethical committee approval was not required. Also, no animals were used in laboratory procedures, and therefore, no violation of animal rights is involved.

Funding Support: No funding was received for this study.

Conflict of Interest: The authors declare that they have no conflicts of interest. All authors have given their consent that this work is valid and represent their views of the study and have given their consent for this work to be published.

Declaration of Generative Artificial Intelligence (AI): The authors declare that the article, tables and figures were not written/created by AI and AI-assisted Technologies.

Authors Contribution: ÖŞ, YA and IA contributed to the conceptualization, design and writing of the study. RAD contributed to the writing of the methodology part of the study and the evaluation of the statistical analysis results. GÇ contributed to the collection of literature and the correction of the study. All authors have written, read and agreed to the published version of the manuscript.

References

1. **FAO:** FAOSTAT. http://www.fao.org/faostat/en/#data/QL; *Accessed:* 01.11.2024.
2. **Gök Y, Şahin M, Yavuz E:** Comparatıve analysıs of indıvıdual lactatıon curve models in some cattle breeds. *KSU J Agric Nat,* 24 (5): 1118-1125, 2021. DOI: 10.18016/ksutarimdoga.vi.845660
3. **Orhan H, Kaygısız A:** Comparison of different lactation curve models for Holstein cattle. *Anim Prod,* 43 (1): 94-99, 2002.
4. **Boztepe S, Aytekin İ, Zülkadir U:** Dairy Cattle. Selçuk University Press, Konya, Türkiye, 2015.
5. **Keskin İ, Boztepe S:** Prediction of 305 days milk yield using partial milk yield in Holstein cattle. *JOTAF,* 8 (1): 1-7, 2011.
6. **Altay Y:** Prediction of (Actual) 305 days milk yield using different lactation milk yield prediction methods and partial milk yield in Holstein cattle. **In**, 4th *International Conference on Agriculture Animal Science and Rural Development,* June 12-14, 47-55, Ankara, Türkiye, 2020.
7. **Girdauskaitė A, Arlauskaitė S, Rutkauskas A, Džermeikaitė K, Krištolaitytė J, Televičius M, Malašauskienė D, Anskienė L, Japertas S, Antanaitis R:** In-line monitoring of milk lactose for evaluating metabolic and physiological status in early-lactation dairy cows. *Life,* 15 (8): 1204, 2025. DOI: 10.3390/life15081204
8. **Kellogg DW, Urquhart NS, Ortega AJ:** Estimating Holstein lactation curves with a gamma curve. *J Dairy Sci,* 60 (8): 1308-1315, 1977. DOI: 10.3168/jds.S0022-0302(77)84028-9
9. **Coşkun G, Şahin Ö, Altay Y, Aytekin İ:** Final fattening live weight prediction in Anatolian Merinos lambs from some body characteristics at the initial of fattening by using some data mining algorithms. *BSJ Agri,* 6 (1): 47-53, 2023. DOI: 10.47115/bsagriculture.1181444
10. **Coşkun G, Aytekin İ:** Early detection of mastitis by using infrared thermography in Holstein-Friesian dairy cows via classification and regression tree (CART) analysis. *Selcuk J Agr Food Sci,* 35 (2): 115-124, 2021. DOI: 10.15316/SJAFS.2021.237
11. **Eyduran E, Zaborski D, Waheed A, Celik S, Karadas K, Grzesiak W:** Comparison of the predictive capabilities of several data mining algorithms and multiple linear regression in the prediction of body weight by means of body measurements in the indigenous Beetal goat of Pakistan. *Pak J Zool,* 49, 257-265, 2017. DOI: 10.17582/journal.pjz/2017.49.1.257.265
12. **Ali M, Eyduran E, Tariq MM, Tirink C, Abbas F, Bajwa MA, Jan S:** Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. *Pak J Zool,* 47 (6): 1579-1585, 2015.
13. **Tırınk C, Eyduran E, Faraz A, Waheed A, Tauqir NA, Nabeel MS, Sheikh IS:** Use of multivariate adaptive regression splines for prediction of body weight from body measurements in Marecha *(Camelus dromedaries)* camels in Pakistan. *Trop Anim Health Prod,* 53: 339, 2021. DOI: 10.1007/s11250-021-02788-y
14. **Coşkun G, Şahin Ö, Ozkan İA, Aytekin İ:** Comparison of data mining algorithms used in predictive of live weight from body measurements in Holstein cattle at different growth and development periods. *Agric Eng,* 375, 37-46, 2022. DOI: 10.33724/zm.1092837
15. **Tırınk C:** Comparison of bayesian regularized neural network, random forest regression, support vector regression and multivariate adaptive regression splines algorithms to predict body weight from biometrical measurements in Thalli sheep. *Kafkas Univ Vet Fak Derg,* 28 (3): 411-419, 2022. DOI: 10.9775/kvfd.2022.27164
16. **Boğa DÇ, Boğa M, Bulut M:** Forecasting seasonal milk production using MARS algorithm for multiple continuous responses in Holstein dairy cattles. *BSJ Agri,* 7 (2): 25-26, 2024. DOI: 10.47115/bsagriculture.1383832
17. **Kibar M:** A cheaply non-destructive technique to estimate honey quality: thermal imaging and machine learning. *Uludağ Bee J,* 24, 79-92, 2024. DOI: 10.31467/uluaricilik.1429971
18. **Kibar M, Altay Y, Aytekin İ:** Exploring the integration of thermal imaging technology with the data mining algorithms for precise prediction of honey and beeswax yield. *Anim Sci J,* 95:e70015, 2024. DOI: 10.1111/asj.70015
19. **Şahin Ö:** Evaluation of some factors on birth and weaning weights in

Awassi sheep by using GLM and CART analysis. *Trop Anim Health Prod,* 54:400, 2022. DOI: 10.1007/s11250-022-03405-2

20. **Akman N, Eliçin A:** Record Keeping and Evaluation in Dairy Cattle Farming. Ministry of Agriculture, Forestry and Rural Affairs, General Directorate of Agricultural Enterprises, Advanced Techniques in Livestock Seminar, pp 304-327, Tahirova-Gönen. Ankara University, Faculty of Agriculture, Offset Unit, Ankara, Türkiye, 1984.
21. **Genç S, Mendeş M :** Linear modeling analysis using for determining the factors affecting 305-day milk yield. *Arq Bras Med Vet Zootec*, 73, 949-954, 2021. DOI: 10.1590/1678-4162-12346
22. **Breiman L, Friedman JH, Olshen RA, Stone CJ:** Classification and Regression Trees. Chapman and Hall/CRC. 1984.
23. **Altay Y:** Phenotypic characterization of hair and honamli goats using classification tree algorithms and multivariate adaptive regression spline (MARS). *Kafkas Univ Vet Fak Derg,* 28 (3): 401-410, 2022. DOI: 10.9775/kvfd.2022.27163
24. **Loh WY, Shih YS:** Split selection methods for classification trees. *Statistica Sinica,* 7, 815-840. 1997.
25. **Hastie T, Tibshirani R, Friedman J:** The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009.
26. **Kass GV:** An exploratory technique for investigating large quantities of categorical data. *J R Stat Soc Ser C-Appl Stat,* 29 (2): 119-127, 1980. DOI: 10.2307/2986296
27. **Loh WY:** Fifty years of classification and regression trees. *Inter Stat Rev,* 82 (3): 329-348, 2014. DOI: 10.1111/insr.12016
28. **Olfaz M, Tirink C, Önder H:** Use of CART and CHAID algorithms in Karayaka sheep breeding. *Kafkas Univ Vet Fak Derg,* 25 (1): 105-110, 2019. DOI: 10.9775/kvfd.2018.20388
29. **Breiman L:** Random forests. *Mach Learn,* 45, 5-32, 2001. DOI: 10.1023/A:1010933404324
30. **Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ:** Random forests for classification in ecology. *Ecology,* 88 (11): 2783-2792, 2007. DOI: 10.1890/07-0539.1
31. **Friedman J:** Multivariate adaptive regression splines. *Ann Stat,* 19 (1): 1-67, 1991. DOI: 10.1214/aos/1176347963
32. **Zhang W, Goh AT, Zhang Y:** Multivariate adaptive regression splines application for multivariate geotechnical problems with big data. *Geotech Geol Eng,* 34, 193-204, 2016. DOI: 10.1007/s10706-015-9938-9
33. **Jekabsons G:** VariReg: A Software Tool for Regression Modeling Using Various Modeling Methods. Riga Technical University, Latvia, 2010.
34. **Akın M, Eyduran SP, Eyduran E:** MARS Algorithm in Solving Regression and Classification Type Problems in Agricultural Sciences with R Software. Ankara Nobel Yayın. 2020.
35. **Arleina O, Otok B:** Bootstrap aggregating multivariate adaptive regression splines (Bagging MARS) for poor households classification in region of Jombang. *SSRN Electronic J,* 2014:1-6, 2014. DOI: 10.2139/ssrn.2489898
36. **Burden F, Winkler D:** Bayesian regularization of neural networks. Artificial neural networks: Methods and applications. Humana Press, Totowa, NJ. 23-42. 2009.
37. **Zhang W, Goh ATC:** Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci Front,* 7 (1): 45-52, 2016. DOI: 10.1016/j.gsf.2014.10.003
38. **Zaborski D, Ali M, Eyduran E, Grzesiak W, Tariq MM, Abbas F, Waheed A, Tirink C:** Prediction of selected reproductive traits of indigenous Harnai sheep under the farm management system via various data mining algorithms. *Pak J Zool,* 51 (2): 421-431, 2019. DOI: 10.17582/journal.pjz/2019.51.2.421.431
39. **Altay Y:** Prediction of the live weight at breeding age from morphological measurements taken at weaning in indigenous Honamli kids using data mining algorithms. *Trop Anim Health Prod,* 54 (3):172, 2022. DOI: 10.1007/s11250-022-03174-y
40. **IBM Corp:** IBM SPSS statistics for Windows, version 23.0. Armonk, NY: IBM Corp. 2015.
41. **Liaw A, Wiener M:** Classification and regression by random forest. *R News,* 2 (3): 18-22, 2002.
42. **R Core Team:** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/; *Accessed:* 15.03.2025.
43. **Kuhn M:** caret: Classification and regression training. Version 6.0-93 [R package]. https://CRAN.R-project.org/package=caret; *Accessed:* 20.12.2024.
44. **Milborrow S:** "earth: Multivariate adaptive regression splines". Version 5.3.4 [R package]. https://CRAN.R-project.org/package=earth; *Accessed:* 05.01.2025.
45. **Hastie T, Tibshirani R:** "mda: Mixture and flexible discriminant analysis". Version 0.5-5 [R package]. https://CRAN.R-project.org/package=mda; *Accessed:* 10.12.2024.
46. **Eyduran E:** ehaGoF: Calculates goodness of fit statistics. Version 0.1.1 [R package]. https://CRAN.R-project.org/package=ehaGoF; *Accessed:* 12.01.2025.
47. **Guevara L, Castro-Espinoza F, Fernandes AM, Benaouda M, Muñoz-Benítez AL, del Razo-Rodríguez OE, Peláez-Acero A, Angeles-Hernandez JC:** Application of machine learning algorithms to describe the characteristics of dairy sheep lactation curves. *Animals,* 13 (17):2772, 2023. DOI: 10.3390/ani13172772
48. **Salehi F, Lacroix R, Wade KM:** Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Comput Electron Agric,* 20 (3): 199-213, 1998. DOI: 10.1016/S0168-1699(98)00018-0
49. **Grzesiak W, Lacroix R, Wójcik J, Blaszczyk P:** A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Can J Anim Sci,* 83 (2): 307-310, 2003. DOI: 10.4141/A02-00
50. **Sharma AK, Sharma RK, Kasana HS:** Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Appl Soft Comput,* 7 (3): 1112-1120, 2007. DOI: 10.1016/j.asoc.2006.07.002
51. **Akıllı A, Atıl H:** Evaluation of normalization techniques on neural networks for the prediction of 305-day milk yield. *Turk J Agr Eng Res,* 1 (2): 354-367, 2020. DOI: 10.46592/turkager.2020.v01i02.011
52. **Usman SM, Singh NP, Dutt T, Tiwari R, Kumar A:** Comparative study of artificial neural network algorithms performance for prediction of FL305DMY in crossbred cattle. *J Entomol Zool Stud,* 8 (5): 516-520, 2020.
53. **Boğa DÇ:** Creating a lactation model for 305-day milk yield with different resampling techniques (Bagging Mars) in Mars modeling. *The Black Sea J Sci,* 14 (2): 522-539, 2024. DOI: 10.31466/kfbd.1383458
54. **Liseune A, Salamone M, Van den Poel, D, Van Ranst, B, Hostens M:** Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning. *Comput Electron Agric,* 180:105904, 2021. DOI: 10.1016/j.compag.2020.105904