DECISION TREE FOR 305-DAY MILK YIELD IN CROSS-BRED CATTLE

ABSTRACT

Cows' Lactation Milk Yield (LMY) is a crucial factor in animal breeding operations. Investigating the influence of potential environmental factors on lactation milk yieldis of paramount importance andin order to identify the various factors influencing lactation milk yield, dairy cattle records were analysed using the regression tree approach. Age, Parity (P), Lactation Length (LL), and Calving Season (CS) were taken into account as explanatory variables while 305-day Milk Yield (MY) as a dependent variable.Decision tree study revealed that Lactation Length, followed by Parity, Age, and Calving Season, had the greatest impact on the 305-d milk output of cross-bred cows. The regression trees use the tree to represent the recursive part. Each terminal node or leaf of the tree represents a cell of the section and has just addeda simple pattern applied to it in this cell. It was evident from nodes (branches) in regression tree, that cows with parities of 1 and 4 (node 11) produced less milk than cows with parities of 2, 3, 5, 6, 7 and 8 (node 10). More milk was produced by cows older than 4.3 years and whose calving seasons were spring and summer (node 40). With the use of the regression tree method, we were able to extract sub-homogenous groups based on the explanatory variables from records of cross-bred cattle and determine the combinations of environmental conditions that produced the maximum 305-d milk yield.

Key Words: Lactation Length, Parity, Milk Yield,Decision tree,Regression tree.

INTRODUCTION

India being the highest producer of milk with production around 209.9 million ton (MT) and more than recommended level of per-capita availability of 427 grams per day in 2020-21 has great potentiality to tap the emerging global milk marketing opportunities

(Annual report GoI, 2021-22). Lactation milk yield is a crucial indicatorfor economics of a dairy farm. Genetic and environmental factors, such as the year, season, and age of calving and days in milking can be divided into groups that influence milk output. Therefore, it's crucial to comprehend the elements that can be used to alter the cow's habitat and enhance normal milk yield in order to successfully boost trait or performance (TopalandEsenboga, 2001;Hussain*et al.,* 2010). For animal breeding operations to be successful, understanding and estimating the effects of different environmental conditions on economic characteristics is necessary (Kuthu*et al.,* 2007). Understanding the interactions and connections between traits that determine yield is also essential to dairy cattle breeding methodology and these traits are closely related to explanatory variables like age, breed, and physical characteristics (Irshad, 2015). Therefore, it can be said that it is crucial for dairy farms to accurately determine the impact of various influencing factors on milk yield. The correlation coefficient is the most significant statistical indicator of the links between variables, although correlation coefficient alone does not indicate if there is a cause and effect relationship between the variables. It is impossible for correlation coefficients to fully describe the relationships between milk yield and its affecting components. In this regard, it is necessary to present the direct and indirect exposure methods in detail which is possible through decision tree(Isci*et al.,* 2004). Numerous statistical techniques, such as simple and multiple linear regression analysis, factor analysis scores, multiple regression analysis, path analysis, and principal component analysis scores for regression tree analysis, can be used in animal research to identify causal links (Khan *et al*., 2014). The researchers have utilised a variety of mathematical models to predict milk yield and genetic advancement in subsequent lactations. However, the accuracy of such mathematical models is dependent on a variety of biological aspects, therefore they are useless if these effects are not taken into account or used properly (Genc and Mendes, 2021).

To overcome the above said problems, we have used regression tree analysis, a non-parametric analysis technique, that divides the population into associations between independent variables that are significant for homogenous subsets of the same species, and identifies curves with linear and interaction in order to explain the variability in a dependent variable (Kayriand Boysan, 2008).Regression trees are superior to alternative techniques in a number of ways as no intricate computations are required, in addition to being quick, estimate makes it simple to identify the critical variables and also with dirty data (i.e. missing values, lots of variables, nonlinear relationships, outliers, and numerous local effects), CART models can be created.

Although decision tree method have been widely used in the ecommerce and marketing services but it hasrarely been studied in animal sciences till date. In order to effectively identify the correlations between the elements that have an impact on economic traits in animal breeding programmes, the partially new regression tree method can be used. As a result, the current study has been designed to identify and categorise, using the regression tree approach, the variables influencing lactation milk yield in cross-bred dairy cattle.

MATERIALS AND METHODS

Animal Resource and Management: The data for the present investigation were collected from the history cum pedigree sheets of Holstein cross-bred cows maintained at Livestock Farm Complex (LFC), College of Veterinary Sciences & Animal Husbandry, Selesih, Aizawl (Mizoram). The data were collected over a period of 09 years, from 2010 to 2019.The cows were maintained under hygienic condition, with proper supply of feed and adlib clean drinking water and managerial regime under intensive system of management. The animals were kept in well-ventilated sheds with brick flooring. They were routinely examined for general healthconditions.In addition to this, feeding of concentrates, cultivated fodder and dry fodder was regular practice.

Data Records: The data (100 records of lactating animals) were recorded from cross-bred cows of Livestock Farm Complex (LFC), College of Veterinary Sciences & Animal Husbandry, Selesih, Aizawl (Mizoram) from period 2010-2019. The four seasons were spring (March, April and May), summer (June, July and August), autumn (September, October and November) and winter (December, January and February).

Parameters of interest: Calving season, parity, lactation length & age were considered explanatory variables and 305-d milk yield was considered dependent variable.

Regression tree method: To represent recursive partition prediction trees were used where a cell of the partition attached to a simple model which applies in that cell only is represented by each of the terminal nodes or leaves of the tree.The data was divided into segments by the Regression Tree that was as homogeneous as possible with regard to the dependent variable (Lahmann and Kottner 2011).

A regression tree can be seen as a kind of additive model of the form:

$$\mathrm{m}\,(x) = \sum_{i=1}^{n} k_i \times I(x \in D_i)$$

Where k_i are constants; I (.) is an indicator function returning '1' if its argument is true and '0' otherwise; Di are disjoint partitions of the training data D such that –

$$\cup_{i=1}^{n} D_i = D \;\; and \;\; \cap_{i=1}^{n} D_i = \phi$$

(Hastie and Tibshirani 1990)

A homogeneous, or "pure," terminal node is one in which all cases have the same value for the dependent variable. Regression trees (RT) is a technique that uses a tree to predict the value of a continuous target variable. Each branch leading from the tree's base to its leaves represents a region, and each node inside the tree represents a logical test of a predictor variable.

Statistical analysis: The data were analysed by Software R-4.0.3 by using rpart and rpart.plot package to compute the results.

RESULTS & DISCUSSION

In our study, records of 100 lactating animals were recorded from period 2010-2019 with average age of 5.14 years (minimum 3 years, maximum 10 years) and average parity of 3.11 (shown in Table 1). The average lactation length of the population was 315.60 daysand the average total milk yield of the population was 1898.273 litres.

First of all, we did the correlation analysis for various traits in our study. The result showed highly significant positive relationship between lactation length & milk yield and age & parity. However,age and parity had negative correlation with milk yield although they were not significant. Calving season was also found to be positively correlated with age and parity of animal while no significant association was found with milk yield.

The parameters that were anticipated to have an impact on cross-bred cows' 305-d milk yield were demonstrated using regression tree diagram (as shown in Fig. 1). The overall descriptive data of the 305-day milk yield were displayed at node 1 (the top of the regression tree diagram). The lactation length (LL) variable separated the principal node into two nodes, indicating that lactation length is the factor that has the greatest impact on the 305 days milk yield.

According to the lactation length variable yet again, Node 2 was further divided into 2 child nodes (Node 4 & Node 5). The parity variable further divided Node 5 into Nodes 10 and 11. The regression tree diagram shows that the secondary variable affecting cross-bred dairy cows' 305-d milk yield is parity. The age variable caused Node 10 to be further divided into 2 child nodes (Node 20 & Node 21), indicating that it is the next significant variable determining 305-d milk yield. Based on the calving season variable, Node 20 was further divided into 2 child nodes (Node 40 and Node 41). Nodes could no longer be divided into smaller groups. As a result, these nodes are referred to as terminal nodes. Finally, we obtained six terminal nodes with adequate homogeneity (Nodes 3, 4, 11, 21, 40, and 41).

Cows with an average age of 4.3 years, parity of 2 to 8 except 4, and lactation length of 255 to 405 days produced highest milk yield. A significant ($p<0.01$) and favourable (positive) correlation between lactation length and milk yield was discovered (Table 2).

In contrast to our study, the effect of parity on milk yield was found to bestatistically significant by Ozçelik and Arpacik (2000), Şeker *et al*. (2009). Correlations between lactation length and milk yield have been reported as non-significant and negative by Mundan *et al*. (2009) which is contradictory to our findings however El-Saied*et al.*(1998), Malhado et al. (2013) reported correlation between them as positive.

Calving season had a substantial impact on 305-days milk yield and milk yield was lower in autumn and winter seasons which were in contrast with other researchers (Sehar, 2005; Erdem *et al*., 2007; Akcay *et al.*, 2007; and Cilek, 2009). Also, according to Barash *et al*. (2001), Jersey cows that calved in months other than December, January and February of the year had low milk production and they concluded this might be a result of high humidity and environmental temperature in summer season which is in line with our study.In line with our study, Bakir *et al.* (2010) also reported that the age had a significant impact on milk yield.Bayril and Yilmaz (2010) reportedsignificant impact on the 305 day milk yield in Holstein cows by the calving year, parity, age and body weight. Likewise, Gorgulu (2011) noted that the most significant influences on the components of milk yield were age and the number of lactations.

Our research demonstrated that lactation length is the main determinant of 305-day milk yield in cross-bred dairy cows, with summer and spring calving cow producing higher yields than winter and autumn calving cows. Mirtagioglu *et al.* (2008) also used the regression tree method to identify the factors that affected the 305-days milk yield and indicated that for 305-day milk yield; age was the most important effective factor, followed by lactation length and calving season. According to Chaudhary *et al.* (2020), the change in milk yield gets smaller over time. The milk yield change rate decreases by 0.0012 units for every year of age increase which is similar to our finding as we also got negative relationship between milk yield and age.

CONCLUSIONS

The average age, parity, lactation length and milk yield of population were 5.14 years, 3.11, 315.60 days and 1898.27 litres respectively. Significant positive relationship between lactation length & milk yield; age & parity as well ascalving season & age and parity of animal both were found. Age and parity had negative correlation with milk yield although they were not significant. The most important determinant for 305-day milk yield, according to regression tree analysis, is lactation length, which is followed by parity, age and calving season. Milk yield of cows older than 4.3 years only, was affected by the calving season. Cows exclusively with milk production and lactations length lasting from 255 to 405 dayswere influenced by parity. In contrast to earlier studies, the regression tree method offers the chance to gather unique information and this paper offers an insight to the applicability of decision tree in animal sciences.

Disclaimer

This paper is an extended version of a preprint document of the same author.

The preprint document is available in this link: https://www.researchsquare.com/article/rs-2985653/v1

[As per journal policy, preprint article can be published as a journal article, provided it is not published in any other journal]

REFERENCES

Akcay, H., Ilaslan, M. and Koc, A.(2007). Effect of calving season on milk yield of Holstein cows raised at Dalaman State farm Turkey.*Journal of Faculty of Agriculture*,**4**: 59-61.

Bakir, G., Keskin, S. and Mirtagioglu, H.(2010). Determination of the effective factors for 305 days milk yield by regression tree (RT) method. *Journal of Animal and Veterinary Advances*,**9**: 55-59.

Barash., H., Silanikove, N., Shamay, A. and Ezra, E. (2001). Interrelationships among ambient temperature, day length and milk yield in dairy cows under a mediterranean climate. *Journal of Dairy Science*, **84**: 2314-2320.

Bayril, T. and Yilmaz, O. (2010). Milk yield traits of Holstein cows raised in KazovaVasfiDiren Agriculture Farm. *Journal of the Faculty of Veterinary Medicine, YüzüncüYıl University*,**21**: 115-119.

Chaudhary, J.K., Verma, M.R. and Kshandakar, Shashank.(2020). A longitudinal study on the impact of species and age of animals on milk production in dairy animals. *Indian Journal of Dairy Science*,**73**: 94-96.

Cilek, S. (2009). Milk yield traits of Holstein cows raised at Polatli state farm in Turkey. *Journal of Animal and Veterinary Advances*, **8**: 6-10.

El-Saied, U.M., Carriedo, J.A., Baro, J.A., De Le Fuente, L.F. and San Primitivo, F. (1998). Genetic correlations and heritabilities for milk yield and lactation length of dairy sheep. *Small Ruminant Research*, **27**(3): 217-221.

Erdem, H., Atasever, S. and Kul, E. (2007). Milk yield and fertility traits of holstein cows raised at gokhoyuk state farm 2. Fertility traits. *Anadolu Journal of Agricultural Sciences (Turkey)*, **22**: 47-54.

Genç, S. and Mendes, M. (2021). Determining the Factors Affecting 305-Day Milk Yield of Dairy Cows with Regression Tree. *Turkish Journal of Agriculture-Food Science and Technology,* **9**(6): 1154-1158.

Gorgulu, O. (2011). Path analysis for milk yield characteristics in jersey dairy cows. *Asian Journal of Animal and Veterinary Advances*, **6**: 182-188.

Hastie, T. and Tibshirani, R. (1990). Generalized Additive Models. Chapman Hall & CRC. Monographs on Statistics & Applied Probability. Chapman and Hall/CRC 1.

Hussain, M., Ghafoor, A. and Saboor, A. (2010). Factors affecting milk production in buffaloes: a case study. *Pakistan Veterinary Journal,* **30(2)**: 115–117.

Irshad, A. (2015). Factors affecting quality and quantity of milk. Department of M.Sc. and Tech. VCRI, Namakkal.

Isci, O., Takma, C. and Akbas, Y. (2004). Study on factors affecting 305 day milk production of Holstein Friesian using path analysis. In *Natural Animal Science meeting,* 1-3.

Kayri, M. and Boysan, M. (2008). Investigation of the relationship between cognitive susceptibility and depression levels by classification and regression tree analysis. *Hacettepe University Journal of Education*, **34** (34): 168-177.

Khan, M.A., Tariq, M.M., Eyduran, E., Tatlıyer, A., Rafeeq, M., Abbas, F. and Javed, K.(2014). Estimating body weight from several body measurements in Harnai sheepwithout multicollinearity problem.*The Journal of Animal & Plant Sciences,* **24**: 120-126.

Kuthu, Z.H., Javed, K. and Ahmad, N. (2007). Reproductive performance of indigenous cows of Azad Kashmir. *Journal of Animal and Plant Sciences,* **17**: 47-51.

Lahmann, N.A. and Kottner, J. (2011). Relation between pressure, friction and pressure ulcer categories: a secondary data analysis of hospital patients using CHAID methods. *International Journal of Nursing Studies,* **48**(12): 1487-1494.

Malhado, C.H.M., Malhado, A.C.M., Ramos, A.D.A., Carneiro, P.L.S., Souza, J.C.D.and Pala, A. (2013). Genetic parameters for milk yield, lactation length and calving intervals of Murrah buffaloes from Brazil. *Revista Brasileira de Zootecnia*, **42**: 565-569.

Mirtagioglu, H., Keskin, S. and Bakir, G. (2008). Regression Tree Analysis for 305 day milk yield in Holstein cows. *Indian Veterinary Journal*, **85**: 943-945.

Sehar, O. (2005). Some production traits of Holstein cows at a state farm in middle Anatolian conditions. *Journal of Lalahan Livestock Research Institute (Turkey)*, **45**: 9-19.

Topal, M. and Esenbuğa,N. (2001). A study on direct and indirect effects of some factors on weaning weight of Awassi lambs. *Turkish Journal of Veterinary and Animal Sci*ences,**25(3)**: 377–382.

TABLE

Table 1: Description of population under study

	N	Minimum	Maximum	Mean	Variance
Age	100	3	10	5.14±0.199	3.965
P	100	1	8	3.11±0.198	3.938
LL	100	90	480	315.60±6.403	4099.636
MY	100	179.0	4004.5	1898.273±69.043	476696.033

Abbreviations: Age, Lactation Length (LL), Parity (P), and Milk yield (MY)

Table 2: Correlation among Parameters [Age, Lactation Length (LL), Parity (P), Calving Season (CS) and Milk yield (MY)]

Correlation Among Parameters						
		Age	**LL**	**P**	**CS**	**MY**
Age	Pearson Correlation	1	0.097	0.980**	0.309**	-0.106
	Sig. (2-tailed)		0.339	0.000	0.002	0.295
	N	100	100	100	100	100
LL	Pearson Correlation	0.097	1	0.100	0.002	0.687**
	Sig. (2-tailed)	0.339		0.321	0.986	0.000
	N	100	100	100	100	100
P	Pearson Correlation	0.980**	0.100	1	0.280**	-0.100
	Sig. (2-tailed)	0.000	0.321		0.005	0.320
	N	100	100	100	100	100
CS	Pearson Correlation	0.309**	0.002	0.280**	1	0.010
	Sig. (2-tailed)	0.002	0.986	0.005		0.919
	N	100	100	100	100	100
MY	Pearson Correlation	-0.106	0.687**	-0.100	0.10	1
	Sig. (2-tailed)	0.295	0.000	0.320	0.919	
	N	100	100	100	100	100
**. Correlation is significant at the 0.01 level (2-tailed)						

FIGURE

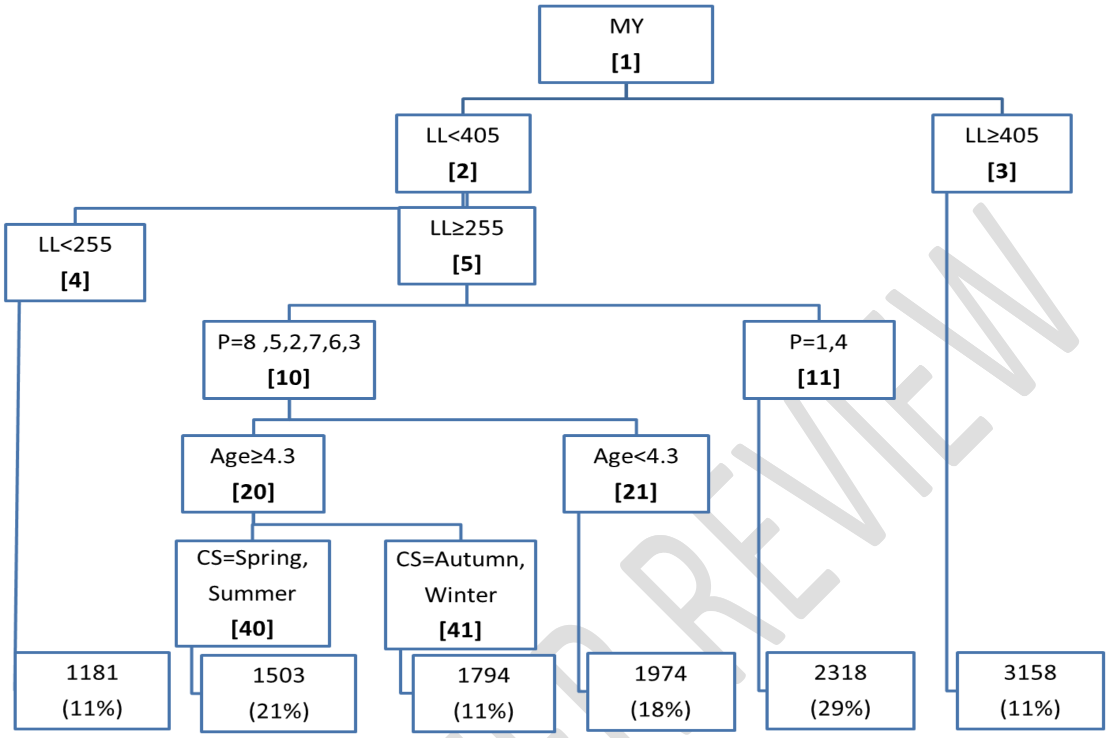

Figure 1: Regression Tree Diagram