

Original Research Article

<https://doi.org/10.20546/ijcmas.2020.904.220>

Prediction of Lifetime Performance in Sahiwal Cattle by Artificial Intelligence Based Machine Learning Models

Namith Chandrashekar^{1*}, Archana Verma¹, Ashok Kumar Gupta¹,
Adesh Kumar Sharma², C. G. Shashank³, Saleem Yousuf¹ and Ravinder Malhotra²

¹Department of Animal Genetics and Breeding, ³Department of Animal Physiology,
ICAR-National Dairy Research Institute, Karnal, Haryana, India

²Department of Dairy Economics, Statistics & Management, ICAR-National Dairy Research
Institute, Karnal, Haryana, India

*Corresponding author

ABSTRACT

Keywords

Lifetime performance, Machine learning algorithms, Prediction, Sahiwal cattle

Article Info

Accepted:
15 March 2020
Available Online:
10 April 2020

Selecting animals for lifetime traits by predicting them based on first lactation traits is a useful method for economic dairy farming. This study was conducted to analyze the prediction performance of various models for prediction of lifetime performance in Sahiwal cattle. The first lactation and lifetime traits were generated from the data on milk production and fertility traits of Sahiwal cattle maintained at ICAR-National Dairy Research Institute, Karnal, India. The adjusted data was used for analysis and using two ML algorithms and also conventional MLR models; lifetime traits were predicted using first lactation traits. Results of this study revealed that most of the lifetime traits were best predicted by SVR model and also RF model. Thus, it is concluded that for prediction of lifetime traits in Sahiwal cattle, one can explore emerging ML models as an alternative to conventional MLR models.

Introduction

Agriculture is the primary occupation in India through which people earn their livelihood. Dairying, which is one of its allied sectors support farmers in performing the agricultural activity and also contribute to their income through the milk and its byproducts.

According to the 19th Livestock census, the total cattle population in India is 190.90 million, among them 151.17 million animals are indigenous cattle, which makes up to 79% of the total cattle population. Further, the annual growth in milk production shows a regular steady increase with 6.4% growth during 2016-17 (BAHS, DADF, GoI, 2017).

With such a large animal population and growth in the sector, it provides a wider scope for the development of industry in much profitable way.

Dairy cattle rearing can be economical and sustainable, only when the animals have better lifetime performance in terms of milk production and when cattle stay in the herd for longer duration with more number of lactations. Higher milk production directly benefits the farmer through increased income, whereas, longer herd life helps in reducing replacement rate of the herd; and, as a result, selection intensity can be increased by practicing voluntary culling (Miglior *et al.*, 2017). So the lifetime production and longevity traits of the animals are crucial in determining the economic worth of the animal. Elite animals should be included in the cattle breeding programs, and selection is the tool which can assist in this process. However the selection of animals needs to be carried out at an early age so that the expenses are reduced and dairy farming will be a successful and profitable venture.

Early selection needs a determination of production capabilities of the animal at an initial stage. Many researchers have found a high positive correlation between first lactation milk yield and lifetime milk yield which ranges from 0.61 to 0.85 (Puri and Sharma, 1965; Hargrove *et al.*, 1969; Dash *et al.*, 2015). So the prediction of lifetime performance of the animals using first lactation production and fertility traits can be carried out to know the worth of the animal and to aid selection at an early stage. Further, for accurate predictions, one has to explore various efficient models to utilize them for the selection purpose.

With the advent of computational technology, intelligent models have been developed for precise predictions. Such recently evolved

Machine learning (ML) models like Random Forests (RF) and Support Vector Regression (SVR) algorithms have better prediction capabilities compared to conventionally used Multiple Linear Regression (MLR) models. The latter conventional models suffer with certain constraints, as the data required for analysis with these MLR models should abide by few assumptions, viz., normal distribution, linear association between dependent and independent variables, and absence of multicollinearity. Whereas, ML models are advantageous over MLR as they overcome nonlinearity and interactions among variables (Shahinfar and Kahn, 2017). Further, these models can learn from training data set and generalize it to unseen test data set. Hence these advanced ML models can be used over MLR models for accurate predictions.

With these views, the present study was carried out to explore the efficiency of two ML models i.e., RF and SVR in comparison with Conventional MLR models to predict lifetime production and performance traits using first lactation traits in Sahiwal cattle, which is an important indigenous milch breed in India.

Materials and Methods

Data collection and processing

The data on milk production and fertility traits of 820 Sahiwal cows maintained at the Livestock Research Centre, ICAR-National Dairy Research Institute, Karnal, spread over a period of 41 years from 1971 to 2012, were collected from history-cum-pedigree sheets. The first lactation traits such as Age at First Calving (AFC) in days, First Lactation Length (FLL) in days, First Service Period (FSP) in days, First Dry Period (FDP) in days, First Calving Interval (FCI) in days, First Lactation 305 days Milk Yield (FL305MY) in kg, and First Lactation Total Milk Yield (FLTMY) in

kg, which were input variables were generated using the raw data. Further, the lifetime traits such as LMY2, LMY3, LMY4, HL and PL were the target variables.

Data standardization and normalization was carried out and finally 380 Sahiwal animals' data was obtained. Further animals had different number of lactations completed in their lifetime and accordingly different number of records was utilized for each of the lifetime trait prediction. Least squares analysis was utilized to determine the effect of various non- genetic factors on first lactation and lifetime traits. Further, the traits were adjusted for significant non- genetic factors and the adjusted data was used for prediction analysis.

Prediction models

Machine Learning (ML) models such as Random Forests (RF) and Support Vector Regression (SVR) were compared with Multiple Linear Regression (MLR) models for their efficiency in predicting Lifetime performance traits.

These Models were developed using open source R programming packages and in case of ML models each of them were tuned with specific parameters to obtain the better prediction accuracy. Variables for MLR models were selected using backward elimination process whereas for ML models different combination of input variables were examined and one which produced best accuracy was used for prediction.

Multiple linear regression models

These are the simple conventional statistical models, which describes the relationship existing between the input and output variables. The linear model function, *glm()* supported by R programming language has

been employed for the MLR analysis in this paper.

Random forests models

Random forest models are basically decision tree based learners. These decision tree learners enabled with bagging and bootstrapping tools are utilized for prediction purposes. Construction of RF models was carried out using 'random Forest' package of R programming (Liaw and Wiener, 2002), *n*tree parameter was tuned to set number of trees to be grown and *m*try parameter defines number of variables randomly sampled at each split. Node size was kept constant at 5.

Support vector regression models.

The SVR models were developed using 'e1071' package (Kowalczyk, 2014) under R programming environment. Epsilon-regression SVR model was fitted with appropriate tuning parameters such as Kernel function, which helps in modelling non linearity in the input data and cost parameter determines generalization capabilities of the function by defining number of observation outside 'ε' zone, which are tolerated. γ parameter defines as to what extent training of SVR model is influenced by single training example and epsilon value was kept constant at 0.1.

Model evaluation

Metrics package from R programming was used to compute measures of model evaluation. The extent of under-prediction and over prediction was indicated by a statistical measure 'percent bias'. The statistical metrics, viz., R^2 , Root Mean Square Error (RMSE) and Relative Prediction Error (RPE) were employed to evaluate prediction accuracy of models. Finally comparative evaluation of models was based upon their

prediction accuracy and prediction performance.

Results and Discussion

Means of first lactation traits are given in Table 1. The combination of first lactation traits and the hyper-parameter used by each of the different models and their results on predicting five of the lifetime traits have been detailed in Table 2.

Prediction of the lifetime trait LMY2 was carried out using a total of 372 records of Sahiwal cattle and the mean was found to be 3787.16kg. From table 1 it is evident that SVR models exhibited best prediction accuracy with R^2 , RMSE and RPE values as 63.38%, 757.16kg and 58.19% respectively. Further RF model exhibited better accuracy next to SVR, and MLR model showed the least. With respect to bias all the three models had similar values. Hence SVR model was found best for predicting the trait LMY2.

The mean LMY3 obtained using 291 Sahiwal animal records was found to be 6256.27kg. For the prediction of the trait SVR model with 49.04%, 1399.40kg and 70.80% as R^2 , RMSE and RPE values respectively were found to be the best compared to RF and MLR models (Table 1). Further bias was found to be the least for the SVR model. So it is observed that SVR model had the best accuracy in predicting the trait LMY3.

The trait LMY4 had a mean value as 8685.69kg from a total of 216 animal records. Though RF model had lower values of R^2 , i.e., 22.47%, the other two metrics RMSE and RPE were 2874.89kg and 82.90%, which was better compared to MLR and SVR models (Table 1).

Even in terms of biasness RF model had the least values. Hence it is observed that RF

model best suited for predicting the trait LMY4.

Mean HL according to 355 lactation records of the Sahiwal cattle was 2888.67 ± 49.71 days. Clearly, SVR model (Table 1) attained best accuracy as 2.41% (R^2), 663.27 days (RMSE) and 92.64% (RPE). Further, RF and SVR models exhibited lowest biasness compared to that of the MLR model. Thus, SVR and MLR models showed similar performance with SVR being somewhat better, in predicting HL trait.

PL was calculated using 355 lactation records and the mean PL of Sahiwal cattle was found to be 1781.58 ± 49.82 days. The SVR model (Table 1) seemed to have relatively better prediction performance with accuracy as 8.31% (R^2), 424.42 days (RMSE) and 91.06% (RPE). Also, SVR model showed the least prediction bias as against the MLR model. Thus, the SVR model performed best in predicting the PL trait.

Prediction of lifetime traits from first lactation production and fertility traits showed relatively lower prediction accuracies in terms of R^2 for different models. Among the production traits, only for LMY2, the R^2 values were higher than 50%, and the rest of the traits had less than 50%. These findings were in accordance with the prediction carried out in Karan Fries cattle using MLR model (Singh et al., 2013, Dash et al., 2015).

The seven predictors (first lactation traits) investigated in this study, seemed to be inadequate to satisfactorily explain the variability inherent in the target variables (lifetime traits). The limited availability of explanatory variables might have influenced the lower prediction performance of the models.

Table.1 Means of first lactation production and fertility traits of Sahiwal cattle

Trait	Mean	SE
AFC (days)	1107.08	7.18
FCI (days)	432.93	5.13
FSP (days)	146.57	5.15
FDP (days)	121.50	4.59
FLL (days)	311.43	4.42
F305MY(kg)	1850.74	35.10
FLTMY(kg)	2031.68	45.13

Table.2 Comparative analysis of machine learning models vs. MLR models to predict various lifetime performance traits in Sahiwal cattle

Lifetime Trait	Model	Variables	Hyper-parameters	R ² (%)	RMSE (kg)	RPE%	Bias (%)
LMY2	MLR	FL305MY, FLL	-	61.63	775.05	58.87	-0.04
	RF	FL305MY, FDP	mtry=1, ntree=500	62.89	754.78	59.47	-0.04
	SVR	AFC, FL305MY, FSP	C=1, kernel=radial, gamma=0.5	63.38	757.16	58.19	-0.04
LMY3	MLR	FLTMY, FSP	-	48.29	1421.49	71.04	-0.15
	RF	FL305MY, FSP, FDP	mtry=1, ntree=100	49.08	1439.15	73.15	-0.17
	SVR	FLTMY, AFC, FSP, FDP	C=1, kernel=radial, gamma=0.5	49.04	1399.40	70.80	-0.12
LMY4	MLR	FLTMY, AFC, FLL, FDP	-	23.80	3025.64	93.44	-0.30
	RF	FLTMY, AFC, FLL	mtry=1, ntree=5000	22.47	2874.89	82.90	-0.25
	SVR	FLTMY, FSP,	C=0.1, kernel=radial, gamma=2	23.58	3127.85	96.37	-0.33
HL	MLR	FSP	-	2.36	653.40	96.49	-0.012
	RF	FSP	mtry=1, ntree=100	0.96	693.03	105.56	-0.004
	SVR	AFC, FSP, FDP	C=10, kernel=linear	2.41	663.27	92.64	0.009
PL	MLR	FL305MY, FSP	-	6.29	431.24	103.40	-0.065
	RF	FL305MY, FSP	mtry=1, ntree=1000	2.4	451.60	111.81	-0.04
	SVR	FLTMY, AFC, FCI	C=100, kernel= linear	8.31	424.42	91.06	-0.005

Further among different models used for prediction, in most cases, SVR model found to exhibit better prediction accuracy compared to RF and MLR models. Other researchers (Gonzalez-Sanchez *et al.*, 2014; Shine *et al.*, 2018) have reported similar prediction abilities of SVR model obtained through supervised learning compared to other ML models. The RF model, which was found to be best for predicting LMY4, is also found better compared to MLR models. Shahinfar and Khan, 2018, also reported better accuracies obtained through RF models compared to conventional methods. The computational intelligence and ability to improve through the experiences might have contributed to the better prediction accuracy of these ML models. In conclusion, the present paper investigated the efficiency of machine learning models in comparison with conventional multiple linear regression models for predicting lifetime performance traits using first lactation production and fertility traits in Sahiwal cattle.

The results from the study revealed that among three models, i.e. SVR, RF and MLR, used in the present study for prediction analysis, SVR was found to be best in accuracy for the traits LMY2, LMY3, HL and PL. Further RF model produced the best performance in predicting the trait LMY4. Thus in general machine learning models performed better compared to conventional MLR models. Hence one can use emerging machine learning models as an alternative to conventional multiple linear regression models in predicting lifetime performance and select animals at an early age for practicing profitable dairying.

Acknowledgements

The authors are immensely grateful to Head, Animal Genetics and Breeding Division and Director, ICAR-NDRI, Karnal for providing necessary facilities during this research work.

References

- Dash, S.K., Gupta, A.K., Singh, A., Mohanty, T.K., Prakash, V., Ahmad, T. and Sahoo, S.K. 2015. Multiple linear correlation and prediction of lifetime performance on the basis of first lactation fertility and production traits in Karan Fries (Holstein Friesian crossbred) cattle. *Indian Journal of Dairy Science* 68(6): 603-606.
- Gonzalez-Sanchez, A., Frausto-Solis, J. and Ojeda-Bustamante, W. 2014. Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research* 12(2): 313-328.
- Hargrove, G.L., Salazar, J.J. and Legates, J.E. 1969. Relationships among First-Lactation and Lifetime Measurements in a Dairy Population. *Journal of Dairy Science* 52(5): 651-656.
- Kowalczyk, A. 2014. Support Vector Regression with R. <https://www.svm-tutorial.com/2014/10/support-vector-regression-r/> (accessed 11 April 2019).
- Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. *R News* 2: 18-22.
- Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P., Baes, C.F., 2017. A 100-year review: Identification and genetic selection of economically important traits in dairy cattle. *Journal of Dairy Science* 100: 10251-10271.
- Puri, T.R. and Sharma, K.N.S. 1965. Prediction of lifetime production on basis of first lactation yield and age at first calving for selection of dairy cattle. *Journal of Dairy Science* 48(4): 462-467.
- Shahinfar, S. and Kahn, L. 2018. Machine learning approaches for early prediction of adult wool growth and quality in Australian Merino sheep. *Computers and Electronics in Agriculture* 148: 72-81.

Shine, P., Murphy, M.D., Upton, J. and Scully, T. 2018. Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. *Computers and Electronics in Agriculture* 150: 74-87.

Singh, S., Rana, Z.S., Pander, B.L., Dhaka, S.S. and Kumar, R. 2013. Prediction of phenotypic value of lifetime performance from early lactation traits in Haryana cattle. *Indian Journal of Animal Research* 47(4): 352-355.

How to cite this article:

Namith Chandrashekar, Archana Verma, Ashok Kumar Gupta, Adesh Kumar Sharma, C. G. Shashank, Saleem Yousuf and Ravinder Malhotra. 2020. Prediction of Lifetime Performance in Sahiwal Cattle by Artificial Intelligence Based Machine Learning Models. *Int.J.Curr.Microbiol.App.Sci.* 9(04): 1867-1873. doi: <https://doi.org/10.20546/ijcmas.2020.904.220>