|  |
| --- |
| **Title:** *(Use Normal style (Times New Roman 12). Only capitalise the first letter of the first word. No full stop at the end of the title)* |
| Utilizing machine learning models to predict dry matter intake in dairy cattle via herd and milk production data |
| **Summary:** *(Your summary (Times New Roman 10) must use Body text style and must not be longer than this box)* |
| **Application** Machine learning models based on herd and milk production data can be used to predict dry matter intake in dairy cattle.**Introduction** A key aspect of a sustainable dairy industry is the efficient use of feed resources. Considering the increasing availability of information generated by digitalisation technology, utilising this data with machine learning (ML) algorithms may help stakeholders identify more efficient dairy animals (Hubbart et al., 2023). While dry matter intake (DMI) is a widely adopted metric to evaluate the feed efficiency of cattle, it is challenging to determine the intake of individual cows in large-scale agricultural systems (Seymour et al., 2019). Therefore, there is interest in applying easily obtainable data to ML algorithms to develop models to predict DMI accurately. Potential data that are relevant to DMI include milk yield and composition, and herd data such as body weight, body conditioning score and lactation week (Vazquez & Smith, 2000). This study aims to use ML methods to predict the DMI of Holstein dairy cows by utilizing routine milk production and and herd information available on farms.**Materials and Methods** Data from five studies conducted at the Agri-Food and Biosciences Institute in Hillsborough was used. All studies commenced at calving and involved a feed-to-yield concentrate allocation approach. All cattle were provided with a diet, comprising of grass silage and concentrates. While additional concentrates were given to every cattle through an out-of-parlour feeding sytem. Data collected included milk yield, milk fat, protein and lactose content, milk fat-to-protein ratio, energy corrected milk (ECM) yield, lactation numbers (1, 2, 3, or 4 or more), weeks-in-milk, live weight, body conditioning score (BCS), and DMI. A total of 4403 weekly cow records were collected. This study applied eight machine learning algorithms to predict DMI using the data collected. Linear methods included linear model (LM), generalised linear model with stepwise selection based on evaluating the AIC values (GLM\_StepAIC), non-linear methods using support vector machine (SVM), and k-Nearest Neighbours algorithm (k-NN). Decision trees and ensemble decision methods, classification, and regression tree (CART), bagged classification and regression trees (Bagged CART), random forest (RF) and stochastic gradient boosting (SGB) were also applied. Model performance was then evaluated using mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R2). Fisher’s Least Significant difference (LSD) test of the accuracy of machine learning (ML) algorithms was performed to evaluate the MAE, RMSE and R2 values to determine the difference in performance between the models. All analysis were undertaken using R v4.3.1. **Results** Using various machine learning methods to predict DMI, the random forest algorithm resulted in the most accurate models with the lowest mean RMSE and MAE value, while having high R2 value.**Table 1** Summary statistic (Minimum, Mean, Maximum) and Fisher’s Least Significant difference (LSD) test of the accuracy of machine learning (ML) algorithms to predict DMI in cattle using mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R2).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  Model |   | MAE |   | RMSE |   | R2 |
|  | Min | Mean | Max |  | Min | Mean | Max |  | Min | Mean | Max |
| SVM |  | 1.51 | 1.66a | 1.83 |  | 1.93 | 2.20a | 2.45 |  | 0.68 | 0.73a | 0.78 |
| RF |  | 1.44 | 1.60ab | 1.77 |  | 1.85 | 2.13a | 2.38 |  | 0.70 | 0.75a | 0.79 |
| SGB |   | 1.47 | 1.63b | 1.73 |   | 1.85 | 2.16a | 2.38 |   | 0.70 | 0.74a | 0.78 |
| LM |  | 1.53 | 1.67c | 1.83 |  | 1.93 | 2.21b | 2.48 |  | 0.68 | 0.73b | 0.76 |
| GLM\_StepAIC |  | 1.53 | 1.67c | 1.83 |  | 1.93 | 2.21b | 2.48 |  | 0.68 | 0.73b | 0.76 |
| k-NN |  | 1.61 | 1.73d | 1.97 |  | 2.06 | 2.27c | 2.64 |  | 0.63 | 0.71c | 0.76 |
| Bagged CART |  | 1.60 | 1.79e | 1.95 |  | 2.04 | 2.35d | 2.62 |  | 0.64 | 0.69d | 0.75 |
| CART |  | 1.91 | 2.14f | 2.29 |  | 2.42 | 2.74e | 2.96 |  | 0.51 | 0.58e | 0.66 |

*Min; Minimum, Max; Maximum.\*\*different superscripts in the mean MAE, RMSE, RF columns represent significant (P < 0.05) difference in LSD test between herd regions.***Conclusions** The present study demonstrated that routine records available on farms can be effectively utilised with machine learning algorithms to develop a DMI prediction model. While algorithms varied in performance, the SVM algorithm had the lowest MAE that was significantly different from the other models. SVM, RF and SGB models have the lowest RMSE and highest R2 values that were significantly different between the other models. Therefore, SVM, RF and SGB models has the ability to provide great accuracy when predicting DMI for dairy cows and has the potential to improve nutritional management in the dairy sector.**Acknowledgments** This project was funded by DAERA and AgriSearch.**References**Hubbart, J. A., Blake, N., Holásková, I., Mata Padrino, D., Walker, M., & Wilson, M. (2023). Challenges, 14(1), 14. Seymour, D. J., Cánovas, A., Baes, C. F., Chud, T. C. S., Osborne, V. R., Cant, J. P., Brito, L. F., Gredler-Grandl, B., Finocchiaro, R., Veerkamp, R. F., de Haas, Y., & Miglior, F. (2019). Journal of Dairy Science, 102(9), 7655-7663. Vazquez, O. P., & Smith, T. R. (2000). Journal of Dairy Science, 83(10), 2301-2309.  |