

between 20 and 30 DIM to minimize the MY variation among cows. Whether the cow had CM in the previous lactation (for multiparous) or during the first 30 DIM were included in the model. First month DHI test somatic cell score (SCS) was also included in the model. Mastitis diagnosis DIM and NCM cow DIM were broken into 10-d intervals and only the 10-d MY prior diagnosis was used. R (3.4.1) and Python (2.7, with the scikit-learn package) were used for data analysis and fitting a logistic regression model. Due to a highly-skewed class distribution, a cost-sensitive classification approach was used with different values of misclassification costs. Five CM prediction windows (1 to 5d before diagnosis) were compared. History of CM in the previous lactation and during the first 30 DIM of the current lactation were significantly associated with 30–60 CM ($P < 0.01$). First DHI SCS was significantly higher in the 30–60 CM cows than the NCM cows (2.56 ± 2.17 vs. 1.79 ± 1.60). The logistic regression model was tested on a held-out test set, which contained 40% of the original data. Precision values (actual sick cows over all predicted sick cows) at a sensitivity level of 0.7 ± 0.01 showed an increasing trend from 0.60 to 0.76 when shortening the predicting window from 5d to 1d before diagnosis. Multiple-step prediction with different algorithms will be tested in the future to improve the accuracy of the model.

Key Words: logistic regression, machine learning, early lactation

327 Predicting clinical mastitis at 30 to 60 DIM using an integrated real-time data warehouse. D. Liang^{*1}, A. Golechha², V. Cabrera¹, and J. Patel^{2,1}, ¹*Department of Dairy Science, University of Wisconsin-Madison, Madison, WI*, ²*Department of Computer Science, University of Wisconsin-Madison, Madison, WI*.

This project aims to permanently predict the onset of cow-level clinical mastitis (CM) using daily milk production (MY) and health records from an integrated data warehouse in the University of Wisconsin-Madison. Milking data and the management records are being transferred from farm computers to our server on a daily-basis since May 2017. With 976,921 daily cow-level MY and 35,169 health-related management records, 125 CM cases between 30 to 60 DIM (30–60 CM) and 4,626 MY records before the diagnosis were found from 118 cows. In the same data set, 21,214 daily MY records between 30 and 60 DIM were found for 3,265 cows (5,844 lactations) that were diagnosed 30–60 CM (NCM). Daily MY was adjusted to the difference from the average MY