Predicting somatic cell count in milk samples using machine learning

Tarr Bence[a]**, Dégen László**[b]**, Monostori Attila**[b]**, Szabó István**[a]**, Tőzsér János**[c]

[a]Institute of Technical Sciences, Hungarian University of Agriculture and Life Sciences
tarr.bence@uni-mate.hu

[b]Animal Breeding Performance Testing Ltd.
degen.laszlo@atkft.hu,monostori.attila@atkft.hu

[c]Institute of Animal Husbandry, Hungarian University of Agriculture and Life Sciences
Tozser.Janos@uni-mate.hu

Abstract

Milk quality is an important factor both for the farmers to be able to sell their product and for the milk industry to be able to plan its production based on quantity and quality. Milk quality has a direct link with cow health, more specifically with utter health. One of the most common utter disease is mastitis. It always captures a lot of interest based on its frequency and cost as a dairy disease which eventually leads to an involuntary and premature culling of milking cows and decreased milk yield. The genetic evaluation of mastitis is very difficult as it is a low heritable trait and categorical in nature [2]. That is why it is necessary to find other markers that could predict the occurrence of mastitis. One of the widely used such marker is the somatic cell count (SCC) [3] which is considered to be the most suitable indicator trait for mastitis resistance in view of its medium to high genetic correlation with mastitis and its greater heritability than mastitis. The SCC is also easy to record in the practice. The selection for lower SCC in milk has a positive effect against the incidence of mastitis. The selection against high SCC also does not deteriorate the immune system of cattle and decreases a risk of infection at the same time. The genetic evaluation[1] of this trait is mostly based on somatic cell score (SCS), a logarithmic transformation of SCC to achieve normality of distribution. In our study we used the milk database of Holstein cows from 5 different farms. From each farm we had altogether 8000 samples tested.

The samples were analyzed using chemical methods every months for a year. 11 different types of data were recorded from each sample. Our aim was to find the best mixture of recorded data that would predict the value of linearized somatic cell count. After the logarithmic linearization the SCC results were divided into 3 main groups (based on the probability of mastitis). Thus our prediction problem turned into a classification problem. We used machine learning to train our algorithm. We were experimenting with different types of classification methods and found good results for prediction of SCC in milk samples. We compared the results using Naive Bayes, Linear SVC and K-Neighbours Classifier methods. We also changed the input variables as not all the 9 measured input variable will be necessary for good prediction results. Our preliminary results show that using machine learning it is possible to build a model that can be used to predict mastitis in diary cows based on variables generally analyzed during milk quality checking test.

References

[1] M. ALAM, C. CHO, T. CHOI, B. PARK, J. CHOI, Y. CHOY, S. LEE, K. CHO: *Estimation of Genetic Parameters for Somatic Cell Scores of Holsteins Using Multi-trait Lactation Models in Korea*, Asian-Australas J Anim Sci 28.3 (2015), pp. 303–310, DOI: https://doi.org/10.5713/ajas.13.0627.

[2] A. A. E. AMIN: *Estimates of Heritability for Somatic Cell Count, Test-Day Milk Yield and Some Udder-Teat Characteristics in Saudi Dairy Goats using Random Regression Animal Model*, Adv. Anim. Vet. Sci 6.3 (2018), pp. 128–134, DOI: http://dx.doi.org/10.17582/journal.aavs/2018/6.3.128.134.

[3] K. C. HALASA TARIQ: *Differential Somatic Cell Count: Value for Udder Health Management*, Frontiers in Veterinary Science 7 (2020), DOI: https://doi.org/10.3389/fvets.2020.609055.