Machine Learning in poultry can we use machine learning with poultry data sets?

By: Karlinton Rafael Flores

Nutrition and animal data analyst.

flores@redroosterdata.com

A surge in data influx

With an increase in technology solutions, there has been a rise in data availability. The main examples are found in data collected over the internet or big retail chains. Society is producing enormous amounts of different types of data. About 2.5 quintillion bytes of data were produced per day in 2017 (DOMO, 2017). Every minute 69,444 LinkedIn users applied for jobs every minute of the day in 2020 (DOMO, 2020). This data is often analyzed with solutions such as machine learning.

A similar trend in data generation is happening in the poultry industry. New technology solutions such as sensors produce a constant number of values that are not being analyzed and stored. In research, new technologies such as the omics sciences (proteomics, genomics, and transcriptomics) are opening the real of molecular data sets. Like other industries do animal production and research data sets can be analyzed, and then data mined with machine learning. Therefore, the objective of this proceeding is to summarize the work done to answer the following question: What questions and data sets can be used with machine learning?

The data sets

A variety of data sources can be added and merged to be analyzed with machine learning. With machine learning the more data the better. In other words, the more variables that are measured and are added to the data set the better. However, as in statistical models the data points need to be reliable, and complete. Thus, filtering and managing the data set is the most important step in the machine learning analysis process. Some data points or variables might not make the cut. Missing values are great examples of data points variables that might be filtered out. The decision on what to add is made by the user and the question that wants to be answered.

The questions

A specific question must be identified by the research team or company. The questions that are currently asked through statistical models are different from the ones asked through machine learning models. Examples of questions in the animal industry and animal research make are: What are the effects of one or two treatments (and their interactions), management practices, or nutritional factors on body weight? With machine learning as the analysis tool, the questions would be formulated as:

How are these variables affecting my body weight at market age?

What variables are important for body weight at market age in my production system?

Can I predict body weight at market age with these variables in the next flock?

The process

After all the data has been cleaned the data is inputted into machine learning software. In our case, we used the software WEKA. This software was created and developed by the University of Waikato I New Zealand. WEKA offers multiple algorithms that can be tested with the data set at the same time. These algorithms do not yield a p-value, but the yield correlation and error indicators such as the root mean square error. The user is responsible to match the data with the correct algorithm and corroborating procedures to make sure the machine did learn. The procedures discussed in the presentation are some corroborations that output is reliable, but there is no replacement to validation!

The outputs

The outputs the software give to the user are two. A prediction model is the first one. This prediction would have correlation and error indicators as their output. The interpretability of the model would depend highly on the algorithm chosen by the user. Some algorithms do not show a visible model to the user, so interaction with it is limited. When a model is not visible (i.e., an equation) it is harder to explain to the end-user and should be considered as a factor for the model's usability. Prediction allows us to answer questions about future flocks. However, they do not answer the question of what the most important variables for my production system are, to find those answers we need data mining.

The second outcome of the data analysis is a data mining analysis. This analysis yields useful information as to what are the variable that driving the model and that have an inherent value. These variables are extracted by the user and should be considered for future research or to increase efficiency in production systems.

Conclusion

We always have questions that need an answer in animal research and production. These questions are being solved by tools as statistics. However, machine learning is proving to be a tool that can take more data, more variables to predict and extract important information. Machine learning is being used to solves another industry problem, and it has been proven that can be used in the animal industry. We should take advantage of it and apply it to boost our research, develop new products, and increase production efficiency.

References

DOMO 2017. Data never sleeps 5.0. https://www.domo.com/learn/data-never-sleeps-5?aid=ogsm072517_1&sf100871281=1. Accessed 02/02/2021 2021.

DOMO 2020. Data never sleeps 8.0. https://www.domo.com/learn/data-never-sleeps-8. Accessed 02/02/2021 2021.