RESEARCHREPORT
ANIMAL SCIENCE

Title: Adaptation of Machine Learning Technologies to Predict Swine Production Outcomes to Assist in Disease Detection – NPB **#: 19-109**

Investigator: James F. Lowe, DVM, MS, DABVP (Food Animal)

Date Submitted: October 5, 2020

Industry Summary:

By its nature, swine production is highly variable, resulting in inaccurate predictions of future performance. One of the critical areas where a lack of predictability has a significant impact on decision-making is animal health and well-being. Other industries have improved their decision-making by increasing the precision of the predictions about future outcomes through machine learning techniques. This project's objective was to adapt advanced analytical methods, proven in other industries, to make predictions about outcomes in swine production with existing data.

The first step in improving health decision making is the accurate prediction of future production outcomes. Machine learning models could generate accurate predictions of gestation length, the probability that an individual sow completes gestation, and the number of weaned pigs for an individual sow at the time of breeding. The predicted results from the model allowed for calculating the number of pigs to be weaned on a given day, facilitating the prediction of pig flow through the system, optimizing both the sow farm's operational efficiency and the resulting

downstream flow. The ability to predict outcomes with a high degree of precision can also serve as the foundation of a novel, passive disease detection system for swine breeding herds.

This project is the first step in precisely predicting the outputs of pork production systems months in the future. Machine learning allows us to make what is complex simple and what appears to be unpredictable, predictable. There are many initiatives in the pork industry to implement sensors to capture additional data to predict outcomes. We view the problem as not a lack of data but a failure to apply advanced analytics to the vast amounts of existing data collected over many years. The use of data collected routinely on sow farms, combined with advanced data analysis techniques, will increase farm efficiency, and create value for all parts of the swine production system.

Key Findings:

- Machine learning models can predict the number of pigs weaned from an individual sow at the service time with a mean squared error of 0.01.
- Machine learning models can predict the gestation length of individual sows with 91% accuracy.
- Machine learning models can predict farrowing predictions with greater than 98% accuracy.
- Using a Monte Carlo Simulation, it is possible to calculate the number of available weaned at breeding.

Keywords: Machine learning, predictive, pig flow, management, analytics

Abstract:

The intersection of animal health and production is crucial as animal production systems strive to maximize overall farm efficiency. Production is a function of genetics, management, nutrition, and health. Human management decisions determine genetics, management, and nutrition—each which can be readily observed and measured. Health remains the primary source of unknown variation in production systems. Therefore, variation from the expected level of production serves as a proxy for the population's health. The prediction of production has been unreliable with the accuracy of models low and the application of traditional statistical methods difficult with a high number of predictor variables. Machine learning allows complex relationships of many variables to be identified and used to make such predictions. In total data was collected from a total of ~250,000 individual sows, over a timeframe of 5 years. This represents >1.1million recorded service events. The data was composed of 62 unique variables, routinely collected on the sow farms represented. After data processing, 885824 instances, service events, remained and 655 variables representing the original 62 and an additional 593 generated through feature engineering. Machine learning predictive models utilizing both gradient boosting and neural nets were generated for gestation length, farrowing, and total born for an individual sow at the time of service. Accuracy of the classification model for if a sow would farrow was the most accurate model with an accuracy of >98%. >91% of sows gestion length could be predicted within a day of farrowing. The number of piglets born by a sow also could be predicted with a mean squared error of <0.1 piglets. Monte Carlo simulations of individual sows within a breed group could be added together a serve as a method to generate a 95% confidence interval of wean piglets per week. These results demonstrate that machine learning may be a valid method to predict production and in turn monitor variations in the health of large groups of animals.

Introduction:

Machine learning techniques are used in many different industries such as crop agriculture (Díaz et al., 2017; Shakoor et al., 2017), forestry (Diamantopoulou and Özçelik, 2018), energy (Fischetti and Fraccaro, 2018), and manufacturing (Li, 2016; Meredig, 2017). These industries and many others utilize the data produced in their system daily to predict and drive productivity. In animal agriculture research using machine learning with the measurement from a variety of sensors has been used to predict parturition in cattle (Borchers et al., 2017) as well as its use in estrous detection (Higaki et al., 2019). The poultry industry has produced several algorithms to help predict the health and production of birds within their systems. One such algorithm was created to forecast the egg production curve over a 5-day interval as a warning signal for problems within flocks and was able to achieve an accuracy of 0.9854 (Ramírez et al., 2016). Researchers have also used machine learning to identify and predict disease prevalence of broilers using production data with algorithms performing with accuracies ranging from 0.78-0.99469 (Hepworth et al., 2012; Zhuang et al., 2018). The use of machine learning to form predictions to aid in productivity is a field quickly gaining popularity. However, a thorough literature search revealed no such techniques currently being developed in the swine industry. Researchers have employed different statistical techniques, such as statistical process control charts, to allow swine production data to serve as an early indicator disease and production problems (Silva et al., 2017; Vries and Reneau, 2009), but such techniques are limited by complexed data. Because of the widespread adoption of machine learning in other industries, the adaptation of these technologies for use in swine production needs to occur if the industry is to remain competitive.

The objective was to develop novel approaches, using of both traditional statistical and advanced machine learning techniques, to predict swine production outcomes, which over the long term will serve as a foundation for detecting changes in swine health. Specifically, to predict the availability of weaned pigs from sow farms. With the ability to accurately predict both gestational length and number of weaned pigs produced for an individual sow at the time of breeding, the possibility of accurately predicting pig flow from sow farm through wean/finish farms exists. Application of these predictions creates value across the entire production chain from farrowing management to the marketing of lean hogs and serves as a foundation to understand changes in animal health when actual production varies from predicted.

Objectives:

1. Deploy various data analysis modalities to design an algorithm capable of predicting an individual sow's gestation length with greater than 90% accuracy.
2. Develop a similar algorithm to predict the number of weaned piglets produced by an individual sow for a given service with accuracy of greater than 90%
3. Use the predictions of gestation length and number of weaned piglets in combination with the distributions of production parameters such as farrowing rate or sow death rate to produce 95% confidence intervals of the predicted number of piglets available to be weaned on a given day.

Materials and Methods:

Data collection and processing:

Retrospective production data from six individual sow farms, each between 2500 and 6000 active females, with greater than five years production history from one swine production system is available for analysis. Data underwent a series of processes to ensure data quality, while maximizing the usefulness of the raw data in training a prediction model. These processes included data quality analysis, data cleaning, feature engineering, and feature selection. All numerical features underwent standardization or normalization depending on normality so that all features lie between -1 and 1. Dummy variables created for all categorical variable so that each category is transformed into a Boolean feature. These procedures were completed to maximize data quality and quantity, increasing model accuracy without overfitting the model, hindering the model's real-world application.

Training a machine-learning model:

Machine learning is a rapidly evolving field. Many algorithms and techniques exist to produce accurate prediction models, and additional algorithms and techniques are being developed every day. In the case of predicting gestation length and number of weaned pigs of an individual sow, a target variable exists and is continuous, therefore a supervised regression model was most appropriate. Although many algorithms exist, due to the number of expected features and instances and the tabular structure of the data, an artificial neural network and a gradient boosting algorithm was chosen.

An artificial neural network (ANN) was selected because of its ability to learn from large complex datasets and make some of the most accurate predictions of any algorithm today. ANN are composed of many layers of nodes and connections like the human brain's neurons and synapses. Each connection is initially weighted with a value of importance and the node gives an output based on a non-linear function which are combined to predict an outcome. As inputs enter the ANN the predicted output and the actual output from the dataset are compared and through a process of backpropagation, the weights of the connections between nodes are adjusted until the difference of predicted outcome and actual outcome is minimized. Due the complexity of the ANN, distributive computing, utilizing the processing power of several computers must be used to complete such a task in a timely manner. Even though ANN requires significant computing power in comparison to other algorithms, it excels when working with large complex data sets making it an appropriate option for this project.

The power of ANN and its ability of learn from data is undisputable and until recently it was thought to be the only answer in formulating extremely accurate predictions from very large complex sets of data. Gradient boosting algorithms however have allowed for accurate predictions from large complex tabular data sets like ANN without excessive computer processing requirements. Gradient boosting is a method where the final prediction model is composed of an ensemble of weak prediction models, usually decision trees. Combining multiple weak predictors creates a single strong predictor. Because of these attributes, gradient boosting is a strong choice when trying to formulate a predication model from the expected data.

Training of both the gestation length model and the pigs weaned model will occurred in a similar fashion. All training was performed using the SciKitLearn and TensorFlow libraries in Python. 70% of the data available was randomly selected to form the training dataset and the remaining 30% of the data composed the test dataset. The model was fitted to the training data using the default parameters. Once trained, an exhaustive grid search technique was employed for model optimization and minimization of the mean square error of test predictions. Optimal model parameters were chosen to both maximize the model's accuracy and minimize overfitting of the dataset allowing for optimization of the model's application in other datasets. Models required an accuracy of greater than 90% prior to further use.

Building a Stochastic Model:

The final prediction of the number of piglets available to wean on a specific day employed a Monte Carlo simulation approach. A Monte Carlo simulation was generated within Excel. This simulation was used to generate a 95% confidence interval for the number of piglets a sow should wean given the predictions of the models. This was accomplished by generating a probability distribution for the prediction of farrowing and total born. For farrowing, a Bernoulli distribution was generated from the probability of farrowing reported for an individual sow. Within this distribution 1 equals farrowing and 0 equals failure to farrow. Then the prediction model for total born was trained 100 times with unique, randomly selected data to generate distribution of the prediction of piglets born from an individual sow. This distribution is a normal distribution around the mean predicted value for a sow. Finally, the mean and standard deviation of weekly pre-wean mortality was calculated from the dataset. Then using one minus the mean weekly pre-wean mortality and the standard deviation a final normal distribution for pre-wean survival was created, to the describe the percentage of piglets weaned from a litter. Once all three distributions were generated numbers were randomly sampled from each distribution. These samples where then multiplied together to represent the weaning outcome of an individual sow. This was repeated 1000 times to build the final distribution of weaning outcome for an individual sow (figure 2). Then using the gestation length prediction, the weaning outcomes of sows farrowing on Friday, Saturday, Sunday, Monday where added together and Tuesday, Wednesday, Thursday. This was done to represent twice weekly weaning within a farm.

Results

In total data was collected from a total of ~250,000 individual sows, over a timeframe of 5 years. This represents >1.1million recorded service events. The data was composed of 62 unique variables, routinely collected on the sow farms represented. After data processing, 885824 instances, service events, remained and 655 variables representing the original 62 and an additional 593 generated through feature engineering were used in training prediction models.

After a descriptive analysis of the data no simple relationships between the 655 variables or the outcome variables could be discovered. Both an artificial neural network and a gradient boosting algorithm were trained with the data. Ranking of the most accurate prediction models within this study are as follows: Farrowing, Total Born, Live Born, Gestation Length, Wean Pigs.

Current accuracy of the farrowing classification model is >98%. The gestation length predictive model, a regression model, has a mean squared error of 0.519, accurately predicting the gestation length of sow within 1 day 91.01% of the time. Total born is currently the best model to predict piglets when compared to wean pigs or live born per litter with a mean squared error of 0.01 (fig.1).

Since the predictions of the machine learning models are probabilistic in nature a Monte Carlo simulation was used to determine the probability distribution of wean pigs for an individual sow (Fig 2). The summation of the individual simulations provided an opportunity to generate a probability distribution of number of wean pigs available on a given day (Fig. 3).

Discussion:

Swine production is by its nature is highly variable resulting in imprecise predictions of future performance. This variation forces producers to make production decisions where a high degree of ambiguity surrounds the facts, such as the number of pigs available for weaning on a specific day in the future, which support those decisions. Other industries have improved their decision-making by increasing the precision of the predictions about future outcomes thought the use of machine learning techniques. The objective of this project was to adapt analytical methods, proven in other industries, to make predictions about outcomes in swine production though the use of existing data. Specifically, we aimed to improve animal flow management within sow farms to optimize the operational efficiency of both the sow farm and the resulting downstream flow.

A highly predictable production chain creates value for all involved. The ability to predict the number of pigs weaned on a specific day from a sow farm provides the opportunity to improve sow breeding management to ensure that the required number pigs are available without creating excess supply. With existing production forecasting methods, production systems buffer capacity buy producing weaned pigs more than demand (space available) to insure achievement of system throughput goals. Practically this means attempting to maximize production on the sow farm with the assumption that the growing pig system will absorb the pigs over its "capacity." This leads to various "work around" solutions that solve short-term needs for space but create long-term inefficiencies in production operations including variable market weights, reduced growth and increased mortality and feed conversion. While systems have become very skilled in managing this unpredictable variation, controlling it represents a significant opportunity for improving production efficiency and financial returns to the US pork industry.

This project is a first step in controlling what is now unpredictable variation in pork production. Machine learning allows us to make what is complex simple and what appears to be unpredictable, predictable. The use of advanced machine learning techniques to train predictive algorithms can aid in pig flow management. Machine learning models could generate accurate predictions of gestation length, the probability that an individual sow farrows and the number of total-born for an individual sow at the time of breeding. These models then used in conjunction with a Monte Carlo simulation can generate the number of pigs available to wean on a specific day. In total, this project makes it possible to predict the number of pigs available to wean prior to or at the time of service, facilitating the alternation of breeding groups, in both number and composition, to meet weaned pig targets with a higher degree of precision.

References

Borchers, M.R., Chang, Y.M., Proudfoot, K.L., Wadsworth, B.A., Stone, A.E., Bewley, J.M., 2017. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle 5664–5674.

Diamantopoulou, M.J., Özçelik, R., 2018. Tree-bark volume prediction via machine learning: A case study based on black alder's tree-bark production 151, 431–440. https://doi.org/10.1016/j.compag.2018.06.039

Díaz, I., Mazza, S.M., Combarro, E.F., Giménez, L.I., Gaiad, J.E., 2017. Machine learning applied to the prediction of citrus production 15.

Fischetti, M., Fraccaro, M., 2018. Machine learning meets mathematical optimization to predict the optimal production of offshore wind parks. Comput. Oper. Res. 0, 1–9. https://doi.org/10.1016/j.cor.2018.04.006

Hepworth, P.J., Nefedov, A. V, Muchnik, I.B., Morgan, K.L., 2012. Broiler chickens can benefit from machine learning: support vector machine analysis of observational epidemiological data 1934–1942.

Higaki, S., Miura, R., Suda, T., Andersson, L.M., Okada, H., Zhang, Y., Itoh, T., Miwakeichi, F., Yoshioka, K., 2019. Theriogenology Estrous detection by continuous measurements of vaginal temperature and conductivity with supervised machine learning in cattle. Theriogenology 123, 90–99. https://doi.org/10.1016/j.theriogenology.2018.09.038

Li, H., 2016. An Approach to Improve Flexible Manufacturing Systems with Machine Learning Algorithms 54–59.

Meredig, B., 2017. Industrial materials informatics: Analyzing large-scale data to solve applied problems in R & D, manufacturing, and supply chain. Curr. Opin. Solid State Mater. Sci. 21, 159–166. https://doi.org/10.1016/j.cossms.2017.01.003

Ramírez, I., Rivero, D., Fernández, E., Pazos, A., 2016. Early warning in egg production curves from commercial hens: A SVM approach 121, 169–179. https://doi.org/10.1016/j.compag.2015.12.009

Shakoor, T., Rahman, K., Rayta, S.N., Chakrabarty, A., 2017. Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.

Silva, G.S., Schwartz, M., Morrison, R.B., Linhares, D.C.L., 2017. Monitoring breeding herd production data to detect PRRSV outbreaks. Prev. Vet. Med. 148, 89–93. https://doi.org/10.1016/j.prevetmed.2017.10.012

Vries, A. De, Reneau, J.K., 2009. Application of statistical process control charts to monitor changes in animal production systems 1 11–24. https://doi.org/10.2527/jas.2009-2622

Zhuang, X., Bi, M., Guo, J., Wu, S., Zhang, T., 2018. Development of an early warning algorithm to detect sick broilers. Comput. Electron. Agric. 144, 102–113. https://doi.org/10.1016/j.compag.2017.11.032

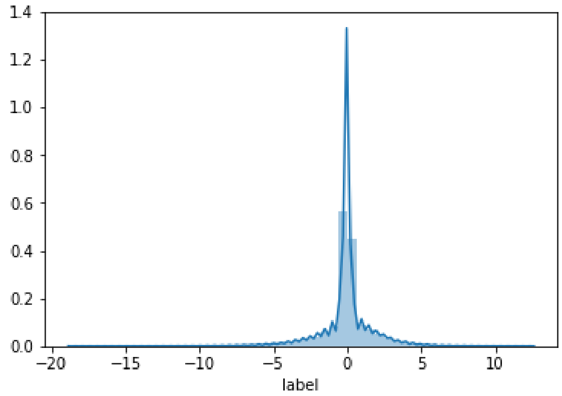

Figure 1 Distribution of prediction errors for Total Born

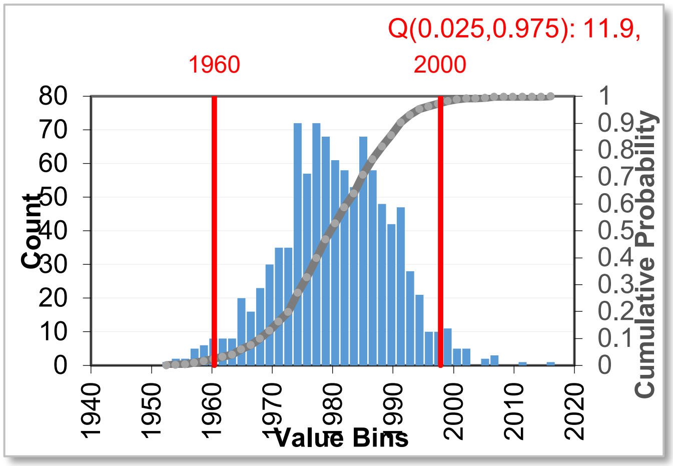

Figure 2. Sample Probability Distribution of the number of weans pigs of an individual sow

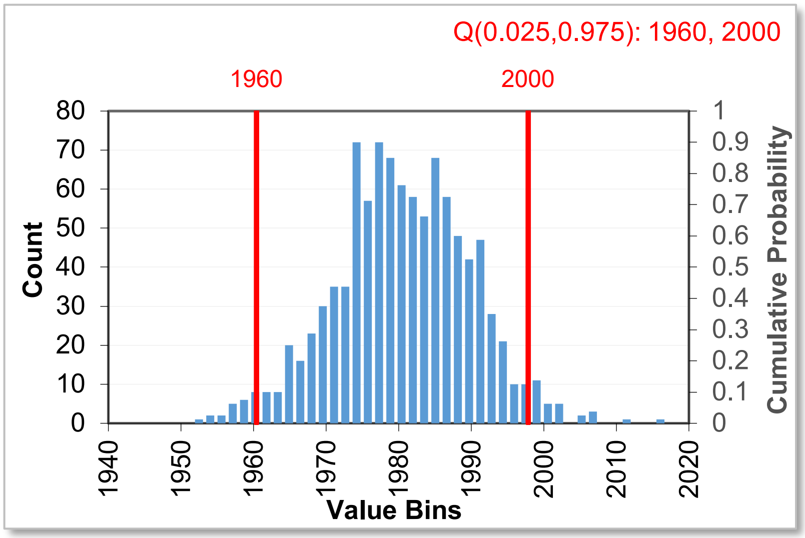

Figure 3. Sample Probability Distribution number of wean pigs available from a breed group