Abstract: Animal scientists have become more enthusiastic about developing machine learning (ML) models to improve the predictability of variables of interest. However, adequate data is limited either because ways to collect the data are scarce or because the process is expensive, time- or labor-consuming, or it simply takes too long to be collected. In addition to the usual hurdles in developing ML models, e.g., appropriate technique, the number of layers, and their activation, enough high-quality data is an essential requirement for ML that is often neglected. Should the data come from longitudinal or cross-sectional experiments of how many subjects? When enough high-quality, reliable data is limiting, one alternative is to create synthetic datasets that reflect the correlation among input and output variables of interest. The probability distribution for each variable needs to be well defined, and the correlation among variables must be taken into account. The normal distribution may not always be a reasonable assumption for all variables; thus, variable-specific distributions must be used with their appropriate parameters. An adequate range (min and max) must be provided for each variable to represent it. Shortcomings might occur when nonlinear relationships occur between variables. The synthetic dataset will also fail to provide good predictability when the new inputs do not have similar correlations, as did the variables used to build the synthetic dataset. It is common to standardize or normalize each variable during ML development. Each variable is normalized based on its minimum and maximum values. A common mistake during the ML development process is that training and evaluation subsets are normalized independently when both should be normalized using the range of the complete dataset; otherwise, normalization becomes dependent on the subset, and the ML weights will differ from one epoch to another. This might weaken the ML predictability because the correct range for back-transforming the output to its original value is unknown.

Keywords: artificial intelligence, big data, modeling

Agent-Based Modeling: A Historical Perspective and Comparison to Other Modeling Techniques.
Karun Kaniyamattam1, Texas A&M University

Abstract: The earliest known history of agent-based modeling (ABM) is traced back to Von Neumann’s self-reproducing cellular automata, designed in the 1940s. The usage of ABM in different scientific fields accelerated in the 1990s, supported by the phenomenal increase in computing capabilities. ABM is gaining acceptance in animal systems, as daily decision-making is becoming complex due to multiple competing outcomes of interest to food systems stakeholders. ABM typically replicates complex real-world systems, for example, a herd of animals that dynamically interact based on simple rules. ABM simulate the heterogenous, stochastic characteristics of agents observed in the real-world. The dynamic interaction of agents replicates the observable real-world complex system patterns. Application of ABM in animal systems ranges from modeling cell behavior, precision nutrition and herd management to predicting the spread of epidemics, and food animal supply chain optimization. Animal science’s most widely used alternative modeling techniques include system dynamic models, differential equations-based models, and statistical modeling. The list of unrealistic assumptions that limit the utility of these modeling techniques includes the assumptions of linearity, homogeneity, normality, and stationarity. The advantageous characteristics of an agent in ABM that set it apart from other techniques include being identifiable, capable of existing in an environment where it interacts with other agents while being autonomous and self-directed, goal-oriented behavior, flexible learning, and capability of adaptations in its behavior over time, based on experience. Identifying the purpose of the model, the questions the model will answer, and the potential users are the key decision variables modelers should ponder upon before embarking on building ABM. The commonly used ABM software includes Repast, Swarm, Netlogo, and MASON. Developing an ABM has several highly interleaved stages: concept development, requirements definition, design, implementation, and operationalization, each of which will be illustrated during the hands-on training session on ABM.

Keywords: agent-based models, decision-making, animal systems