Summary of Stakeholder Engagement for Data-to-Decisions in Agricultural & Environmental Modeling

Background & Meeting Goals

Researchers at N.C. State University received a grant from the U.S. Department of Agriculture Food and Agriculture Cyberinformatics and Tools Initiative to develop computational tools that facilitate use of machine learning models in agricultural and environmental management. To gauge the practicality of these tools, stakeholders who conduct data analysis and environmental/agricultural modeling were invited to attend a one-day forum to:

(1) Ascertain if advances in machine learning modeling could benefit their day-to-day work; and,

(2) Identify ways these modeling advances could be made more accessible for those who could benefit from them.

This document summarizes the broad learning from the stakeholder engagement meeting on February 27, 2020 at the N.C. State University James B. Hunt Jr. Library. Stakeholders who attended represented academia, extension, industry, consulting, and local government and had a range of machine learning experience.

Meeting Summary

After a machine learning presentation by **Natalie Nelson,** Assistant Professor, N.C. State University, Department of Biological & Agricultural Engineering & Principal Investigator, Biosystems Analytical Lab and **Sheila Saia**, Postdoctoral Research Scholar, Biosystems Analytics Lab, participants were asked two broad questions:

(1) Given opportunities/case studies that could benefit from machine learning modeling, what *barriers* might you encounter when using machine learning models; and,

(2) What *solutions* could mitigate or overcome those barriers?

Current Barriers

Limited Definition and Scope: Stakeholders noted that there is *no accepted definition of machine learning.* Without a common definition, it is difficult to come to an agreement about what machine learning will/will not accomplish. With limited common baseline understanding, stakeholders reported resistance to using machine learning especially when companies have established workflows with process-based models. And, a related concern is ethical: process-based modelers are trained to know which parameters may be adjusted based on uncertainty and which parameters should not be modified. That understanding ensures that process-based models are not forced into producing a desired outcome, but rather, represent real-world patterns. Machine learning models, with their "black box" approach, do not provide a standard method or commonly accepted guideline for objectively developing research protocols.

Limited Guidance & Acknowledgement of Costs: Some of the concerns noted above are due to its newness, but there is also a limited amount of literature or broader community guidelines regarding machine learning model selection for case studies or situations at hand. Machine learning models also require large amounts of high-quality data before they begin to provide useful results. If stakeholders lack guidance outside of their own understanding, they may not have the knowledge or skills to apply machine learning models effectively. Additionally, implementation inevitably incurs a cost - everything from staff training to data integration. There are currently very limited resources on how the life cycle cost and resource requirements of machine learning models compare to standard (process-based) models.

Limited Effective Communication: Finally, the language surrounding machine learning is especially important to its success. Stakeholders described decisionmaker discomfort with describing machine learning models as "black boxes". Tension between scientists and decision makers occur with any technical approach, but the problem is enhanced with machine learning because, unlike process-based models, it is more difficult (but not impossible) to link a result to a particular underlying physical process or human activity. By trusting data and algorithms, researchers can automate analytical processes. However, if that automation means replacing someone's job with an algorithm, a decision maker may be reluctant to embrace machine learning. These qualitative considerations must be addressed to move machine learning modeling forward.

Potential Solutions to Improve Adoption

The barriers for machine learning deployment then broadly include the language and understanding surrounding its use, the lack of available resources for data researchers, including the data set requirements for developing a machine learning model. Stakeholders offered three thematic areas where solutions to these barriers could improve adoption of these machine learning tools.

Interpretation & Communication: First, stakeholders suggested training on how to implement and use machine learning models as well as how to effectively communicate results. Just like any other new technology, training should focus on the change to mind-set, skill-set, and associated work infrastructure needed for its deployment. Rather than using the phrase "black box" to describe machine learning models, stakeholders suggested reframing advantages and disadvantages of machine learning models in more precise ways. This educational component, they believed, could improve integration of machine learning modeling approaches for decision making in environmental and agricultural sciences/policy.

Cost-benefits of Integrating Machine Learning Approaches: Guidelines that document the cost of running a machine learning model as compared with a process-based model would illustrate where machine learning approaches are most successful and how they may improve current workflows. Stakeholders also suggested guidelines cover how to select and limit inputs and outputs; and, application specific situations such as the minimum dataset required for a certain type of application. In other words, guidelines on the minimum number of variables or observations for developing a model for specific applications (i.e., predicting nutrient flow or plant disease forecasting). A pilot project that pairs a real-world situation with machine learning guidance would go a long way in developing practical on the ground training.

Building a Dynamic Research Community: Finally, stakeholders noted the benefit from collaboration with others as they apply machine learning models to different situations. Attendees represented a range of interests from forestry, to soil science, to water quality to aquaculture and some had considerable experience with machine learning and others little to no experience. Therefore, solutions might include dividing stakeholders by interest and machine learning knowledge so that sharing occurs among similar disciplines and experience.