samsi
NSF·Duke·NCSU·UNC

Deep Learning Program
Triangle Machine Learning Day
September 20, 2019

SPEAKER TITLES/ABSTRACT

Wayne Thompson
SAS

"What You Didn't Learn About Machine Learning in School"

Machine learning is all the rage these days. As a data scientist at SAS I thought I would be talking more with customers about machine learning algorithms. You know the stuff we learned in school or by taking online classes. What algorithms to use for what use case? How to tune the model and evaluate for generalization. How to interpret these highly complex nonlinear models? Instead most customers want to know more about putting models into production and managing models once they're deployed. Indeed, model deployment and model management are the two must-have steps to get machine learning right.

A machine learning pipeline is comprised of data wrangling + feature engineering and extraction + model formulae. It may also be layered with rules. Each model includes a lot of data preparation logic. You must aggregate many data sources, include the model formulae, and layer it with rules or policies. Most organizations don't have enough rigor and metadata to re-create the data wrangling phase for scoring. As a result, many of the backward data source dependencies for deriving the new scoring tables get lost. This is the biggest reason why most organizations take too long to put a model to work.

Models begin to degrade as soon as they are deployed. It is important to monitor model drift, retrain champion models and evaluate new challengers. Model fairness and bias must be addressed. Advanced organizations are running model training services at the edge. I believe one of the holy grails of machine learning is being able to orchestrate a continuous learning platform in real-time. One that can adapt as the population is changing. We will discuss these strategies and more to help you get more value out of machine learning.