Anatomy of Data Analytics, Machine Learning and Deep Learning -- Demystified - Part IV

We have been discussing the need, methodologies and algorithms of Shallow and Deep Learning thus far.

This edition presents a case study where all the theory is put to practice. The described process flow is applied to a large Delaware basin data set comprising of 5716 horizontal wells in the Wolfcamp formation. The original database contains a total of 131 predictor variables containing 26 reservoir, 21 completion, 22 well architecture. 53 production and 9 reservoir fluid related. The enormity of making sense out of this data stack is either too cumbersome for direct



human interpretability or insurmountably time consuming (and often impractical) for physics based models. The Machine Learning techniques unravel the underlying trends and interrelationships between



the driver and response variables. The deep and shallow learning algorithms should be applied in a systematic manner, traversing step-by-step methodology of data preparation, exploratory data analysis, model selection, model validation, model parameter tuning, selection of variable of importance and model application. In particular, data sets are prepared for both Supervised Regression (continuous) and Classification (categorical) methods. Post exploratory data analysis, multivariate regression

along with Multicollinearity/Variation Inflation Factor and outlier tests are applied to reduce the predictor variable list. Thereafter, models of Shallow Learning (e.g.Gradient Boosting, Support Vector Machine, k-Nearest neighbors, Decision Trees, Random Forest etc.) and Deep Learning (e.g. kerasRegressor/Classifier) are progressively disciplined on training data sets to be tested on the hold-out

data sets. Models are also optimized based of GridSearch techniques. Accuracy of predictability is compared against standard goodness-of-fit metrics.

The figure above shows the relative comparisons of the performance and accuracy of the various methods used. The plots show the crossplot of model predicted and actual EURB (target) variables. The adjoining figure shows the



variables with relative importance. Some conclusions are intuitive e.g. if the well has a high 3 month BOE it is likely it will have good EUR. However, there are other conclusions that are not as intuitive!