Boosted Models

Summary:

STEP 1: Create sample

In any classification problem you will need to set an estimation sample and a validation sample of your data. This helps us compare different classification models to see which better fit the data.

Useful Alteryx tool: Create Sample

STEP 2: Model Settings

Select a target variable and predictor variables, you can include as many predictor variables as you would like because the model will only use variables that work best. For a Boosted model it is best to set the target type in the model customization tab. Your options are Continuous, Count, Binary Categorical or Multinomial Categorical.

Useful Alteryx tool: Boosted Model

STEP 3: Interpreting the Report

The Variable Importance Plot is a bar graph that's length indicates the importance of the predictor variables. The Number of Iterations Assessment Plot illustrates how the deviance (loss) changes with the number of trees included in the model. The vertical blue dashed line indicates where the minimum deviance occurs using the specified assessment criteria

Useful Alteryx tools: Boosted Model

STEP 4: Model Comparison

Use the fit and error measures, Accuracy which represents the overall accuracy, the number of correct predictions of all classes divided by total sample number. The F1 score is calculated the following way, precision * recall / (precision + recall) You can read more about precision and recall.

The Confusion Matrix is a matrix (or table) that lists out all of the possible prediction results when we validate our model against our validation set. This confusion matrix is one of the best methods to review the accuracy and precision of your model as well as to understand any model bias in classifying your data points.

Useful Alteryx tool: Model Comparison

STEP 5: Score Data

Apply the model by attaching a score tool to the data set you are trying to classify and the model object.

Useful Alteryx tool: Score