Predictive Master

# **Prepare** **Data**

• Impute missing values

• Handle unbalanced datasets

• Create training and validation datasets

# **Investigate** **Data**

• Perform feature engineering

• Reduce feature dimensionality using PCA

• Select predictor variables using correlation metrics

• Appropriately standardize predictor variables

• Interpret data investigation reports

# **Train** **Models**

• Train predictive models for classification analysis (Gradient Boosting, Logistic Regression, Naïve Bayes, Decision Tree, SVM, Random Forest, Neural Network)

• Train predictive models for regression analysis (Linear Regression, Gamma Regression, Linear regression with stepwise reduction of variables)

• Train models for time series forecasting (ETS and ARIMA)

• Train models for cluster analysis

• Train prescriptive models for optimization

# **Compare Models**

• Compare and evaluate predictive models using Model-Comparison and Cross-Validation tools

• Compare the performance of a full model vs. a model with reduced dimensionality

• Determine if two models are statistically equivalent

• Compare time series forecasting models

# **Interpret** **Fit** **Statistics**

• Interpret error measures (RMSE, MAPE, MASE)

• Compare precision measures of a class

• Compare F1 scores and average accuracy of models

• Interpret cluster information statistics

# **Interpret** **and** **Apply** **Results**

• Identify misclassified records

• Interpret coefficients of a linear regression model and determine the impact of specific variables

• Interpret a plot of residuals vs. fitted values

• Interpret confusion matrices

• Determine if an interaction effect exists

• Interpret PCA measures of proportion of variance

• Identify the best predictor variable

• Forecast future periods

• Score new data with a trained model

• Assign instances to clusters