Print

Predictive Modeling Terminology

The goal of predictive modeling is to predict, or score, future values of a target variable based on the existing values of inputs. The process begins with partitioning a data set into separate training and validation data sets. The model is built using the training data and then assessed using the validation data. After a best model is chosen, the model is deployed to make predictions on new data using a process called scoring.

In predictive modeling, the predictor variables are often referred to as inputs, but are also known as features, explanatory variables, or independent variables. The response variables are often called targets, but are also known as outcomes or dependent variables. The observations, known as cases, are sometimes called instances or records.

The measurement scale and variable type of the inputs and targets can be varied across different applications. They can be continuous variables like income or age, and categorical variables such as occupation or city. They are often binary variables, such as a positive or negative response to a survey, or the presence or absence of some characteristic.

A predictive model consists of either a formula or rules, depending on the type of analysis that you use. The predictive models in this lesson are based on regression models, which are parametric and have formulas. Predictive models can also be based on nonparametric models such as decision trees and random forests, which predict new cases based on a sequence of decisions, or rules, based on the values of the inputs.

Close