IDC | Accelerating Time Series Analysis with Automated Machine Learning | DOWNLOAD NOW

Target Variable

What is a Target Variable in Machine Learning?

The target variable of a dataset is the feature of a dataset about which you want to gain a deeper understanding. A supervised machine learning algorithm uses historical data to learn patterns and uncover relationships between other features of your dataset and the target.

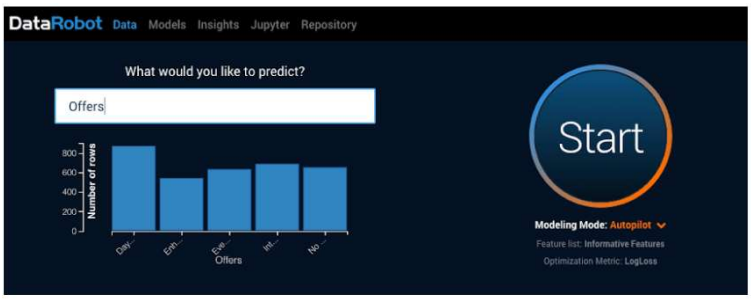
The target variable will vary depending on the business goal and available data. For example, let's say you want to use sentiment analysis to classify tweets whether tweets about your company's brand are positive or negative. Some aspects of a tweet that might be useful that you can use as features are word tokens, parts of speech, and emoticons. A model can't lear how those features relate to sentiment without first being given examples of which tweets ar happy or sad (the target). Targets are often manually labeled in a dataset, but there are ways automate this process (see semi-supervised machine learning).

Why are Target Variables important?

Without a labeled target, supervised machine learning algorithms would be unable to map available data to outcomes, just as a child would be incapable of figuring out that cats are called "cats" without having been told so at least a few times. It is important to have a well-defined target since the only thing an algorithm does is learn a function that maps relationsh between input data and the target. The model's outcomes will be meaningless if your target doesn't make sense.

Target Variables + DataRobot

DataRobot makes it easy to select a target and start building supervised models. Once a use uploads a dataset and indicates which feature they want to understand, and DataRobot will d the rest of the data science heavy lifting.



Once a user chooses a target and hit "Start," DataRobot automatically uncovers insights that show you how features relate to the target and how much each trained model has learned about the target. This helps you easily pick the best model to deploy in your production application and spot issues that are notoriously difficult to discern, such as target (data) leakage.

Model Fitting

Model Interpretability

Model Tuning

Multiclass Classification

Neural Network

Open Source Model Infrastructure

Overfitting

Prediction

Prediction Explanations

Regression

Scoring Data

Semi-Supervised Machine Learning

Stacked Predictions

Supervised Machine Learning

Target Leakage

Target Variable

Text Mining

Training Sets, Validation Sets, and Holdout Sets

Underfitting

Unsupervised Machine Learning

Related materials

White Paper: Data Preparation for Automated Machine Learning

White Paper: Model Deployment with DataRobot

Blog: Act Now on Artificial Intelligence

One International Place, 5th Floor, Boston, MA 02110

Contact: info@datarobot.com

DataRobot, Inc. © 2019

Privacy and Terms of S

We use cookies and other tracking technologies to improve your browsing experience on our site. To learn more, please read our updated privacy

policy. By choosing "OK", you consent to our use of cookies and other tracking technologies. Find out more.

OK