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 main difference between training data and testing data is that training data is the subset of original data that is used to train the machine learning model, whereas testing data is used to check the accuracy of the model. The training dataset is generally larger in size compared to the testing dataset.

Ground truth in machine learning refers to the reality you want to model with your supervised machine learning algorithm. Ground truth is also known as the target for training or validating the model with a labelled dataset.

Ground truth labels categorize data points into groups based on assignment by a human or an existing algorithm. These types of metrics do their best to suggest the correct number of clusters but can be deceiving when used without context.

A model starts the training process with random parameter values and adjusts them throughout. Whereas hyperparameters are the components set by you before the training of the model. The values of hyperparameters might improve or worsen your model's accuracy.