* The model is online (continuously receives data and the model adapts) or offline (the model is trained on a dataset and the model does not intake more data after being deployed).
* The type of learning, supervised, semi-supervised, or unsupervised, should be used to fit the data
* The model is either classification or regression, where it is based on the expected outputs.

DO definition of one e.g

Vector: akd jahsdj djak

Related: AAttribute [link]]

Fold cross validation: <https://www.openml.org/a/estimation-procedures/1>

Found a good compilation of Machine learning definitions at <http://robotics.stanford.edu/~ronnyk/glossary.html>

Unseen data:

Online model:

Related: Offline model

Offline model:

Related: Online model.

Supervised learning:

Related: semi-supervised learning, unsupervised learning.

Semi-supervised learning:

Related: Supervised learning, Unsupervised learning.

Unsupervised learning:

Related: Supervised learning, Semi-supervised learning.

Classification model:

Related: Regression model.

Regression model:

Related: Classification model.

Feature vector: Is an n-dimensional vector of numerical features that represent some object.

Target attribute:

Attribute:

Encoding:

Instances: A row in the dataset.

Training dataset:

Related: Training process

Training process:

Related: Training dataset.

Validation dataset:

Related: Testing process

Testing process:

Related: Validation dataset

Good fit:

Related: Underfit, Overfit.

Underfit:

Related: Good fit, Overfit.

Overfit:

Related: Good fit, Underfit.

Gaussian Distribution:

Also known as Normal Distribution

Hyper-parameter tuning:

Ground truth: Used in supervised learning where each instance is provided an attribute that is the target result of the non-target attributes.

Also known as target attribute.

k-Fold cross validation: “In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data.”

Source: <https://www.openml.org/a/estimation-procedures/1>