data iku

AN INTRODUCTION TO KEY DATA SCIENCE CONCEPTS

Definitions of 10 fundamental terms for data science and machine learning.

model (n)

['mɑdəl] / noun

1. a mathematical representation of a real world process; a predictive model forecasts a future outcome based on past behaviors.

algorithm (n)

['ælgəˌrɪðəm] / noun

1. a set of rules used to make a calculation or solve a problem.

training (v)

['treɪnɪŋ] / verb

1. the process of creating a model from the training data. The data is fed into the training algorithm, which learns a representation for the problem, and produces a model. Also called "learning".

regression (n)

[rə'grɛʃən] / noun

1. a prediction method whose output is a real number, that is, a value that represents a quantity along a line. Example: predicting the temperature of an engine or the revenue of a company.

classification (n)

[ˌklæsəfə'keɪʃən] / noun

1. a prediction method that assigns each data point to a predefined category, e.g., a type of operating system.

target (n)

['tɑrgət] / noun

1. in statistics, it is called the dependent variable; it is the output of the model or the variable you wish to predict.

training set (n)

['treɪnɪŋ sɛt] / noun

1. a dataset used to find potentially predictive relationships that will be used to create a model.

test set (n)

[tɛst sɛt] / noun

1. a dataset, separate from the training set but with the same structure, used to measure and benchmark the performance of various models.

feature (n)

['fitfər] / noun

1. also known as an independent variable or a predictor variable, a feature is an observable quantity, recorded and used by a prediction model. You can also engineer features by combining them or adding new information to them.

overfitting (v)

[ˌoʊvər'fɪtɪŋ] / verb

1. a situation in which a model that is too complex for the data has been trained to predict the target. This leads to an overly specialized model, which makes predictions that do not reflect the reality of the underlying relationship between the features and target.