|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SEPAL LENGTH | SEPAL WIDTH | PETAL LENGTH | PETAL WIDTH | FLOWER SPECIES |
| 5.1 | 3.5 | 1.4 | 0.2 | Setosa |
| 6.7 | 3.1 | 4.7 | 1.5 | Versicolor |
| 7.2 | 3.2 | 6.0 | 1.8 | Virginica |
| 5.9 | 3.0 | 4.2 | 1.5 | Versicolor |
| 4.9 | 2.5 | 4.5 | 1.7 | Virginica |
| 5.8 | 2.7 | 3.9 | 1.2 | Setosa |

TERMINOLOGIES:

* FEATURE:

Specific quantifiable attributes that are fed into the model.  
For instance: Petal Length (cm), Petal Width (cm), Sepal Length (cm), and so on.

* LABEL:

The variable that the model attempts to forecast as an output.  
For instance, flower species.

* PREDICTION:

The model's output after being supplied with a set of features.  
Example: According to the model, a flower of the genus Setosa has petals that are 1.4 cm in length, 3.2 cm in width, 5.1 cm in sepal length, and 3.5 cm in width.

* OUTLIER:

A single data point that substantially differs from the rest of the data.

For instance, let's say that this dataset had a flower with a sepal length of 10.0 cm.

* TEST DATA:

Information utilized to assess the model's performance.

As an illustration, the dataset's final row may be designated as test data.

* TRAINING DATA:

Data that the model was trained on.

Example: The dataset's first five rows.

* MODEL:

A prediction-making mathematical model that is learned from data.

Example: A classification model that uses features to estimate the species of flowers.

* VALIDATION DATA:

Information used to adjust the model's parameters.

An example would be a portion of the training data reserved for validation.

* HYPERPARAMETER:

Setting parameters prior to training the model has an impact on the training procedure.

Example: Learning rate and the number of trees in a random forest.

* EPOCH:

One thorough iteration of the training set.

For instance, 100 iterations of the model running over the dataset during training equals 100 epochs.

* LOSS FUNCTION:

A metric that assesses how well the actual data and the model's predictions match.

For instance, classification models employ cross-entropy loss.

* LEARNING RATE:

A hyperparameter governing the degree of model adjustment in relation to the loss gradient.

For instance, a 0.01 learning rate.

* OVERFITTING:

When a model performs poorly on fresh data after learning the training set of data which includes noise.

An illustration would be a model that performs well on training data but poorly on test data.

* UNDERFITTING:

When an overly simplistic model fails to identify the underlying patterns in the data.  
A model with poor accuracy on test and training data, for instance.

* REGULARIZATION:

Methods for limiting the model's fit in order to avoid overfitting.

L2 regularization, for instance.

* CROSS-VALIDATION:

A method for evaluating a model's ability to generalize to a different set of data.

K-fold cross-validation, for instance.

* FEATURE ENGINEERING:

The procedure for adding new features or changing current ones in order to enhance model performance.

An illustration would be to take "Petal Length" and "Petal Width" and create a new feature called "Petal Area".

* DIMENTIONALITY REDUCTION:

Methods for lowering the quantity of characteristics in a dataset.

PCA stands for Principal Component Analysis.

* BIAS:

Error caused when a simpler model is used to approximate a real-world problem, which may be complex.

For instance, a model that solely uses sepal length to predict the species of flower may be highly biased.

* VARIANCE:

The training data variations' impact on the model.

Example: A high variance model is one that alters dramatically even with slight changes in the training set.