|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| AGE | BLOOD PRESSURE | CHOLESTEROL LEVEL | SMOKING | DISEASE |
| 45 | 140 | 230 | 1 | 1 |
| 50 | 130 | 220 | 0 | 0 |
| 55 | 150 | 250 | 1 | 1 |
| 60 | 135 | 240 | 0 | 0 |
| 65 | 160 | 260 | 1 | 1 |
| 70 | 125 | 210 | 0 | 0 |

TERMINOLOGIES:

* FEATURE:

Specific quantifiable attributes that are fed into the model.  
Example: Smoking (1 = Yes, 0 = No), Age, Blood Pressure (mmHg), and Cholesterol Level (mg/dL).

* LABEL:

The variable that the model attempts to forecast as an output.  
Illness, for instance, (1 = Yes, 0 = No).

* PREDICTION:

The model's output after being supplied with a set of features.  
Example: According to the model, a person who is 45 years old, has a blood pressure of 140 mmHg, a cholesterol level of 230 mg/dL, and smokes 1 is likely to have the condition (1 = Yes).

* OUTLIER:

A single data point that substantially differs from the rest of the data.  
For instance, suppose someone in this dataset had a blood pressure of 200 mmHg.

* 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: Based on features, a logistic regression model forecasts the likelihood of a disease.

* 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.

For instance, learning rate and epoch count.

* 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, binary classification models employ the Binary 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 the fundamental patterns in the data are not sufficiently captured by a model.

An illustration would be a model with poor accuracy on test and training sets.

* REGULARIZATION:

Methods for keeping the model from overfitting by introducing limitations.

L2 regularization is one example.

* CROSS-VALIDATION:

A method for determining how well a model's output will transfer to a different collection of data.

K-fold cross-validation is one example.

* FEARTURE ENGINEERING:

The process of enhancing model performance by adding new features or changing current ones.

Example: Taking "Weight" and "Height" and creating a new feature called "BMI".

* DIMENTIONALITY REDUCTION:

Methods for cutting down on a dataset's feature count.

Principal Component Analysis (PCA) is one example.

* BIAS:

Error brought about by using a simpler model to approximate a real-world problem, which may be complex.

Example: A model with high bias may be one that just uses age to predict the likelihood of a disease.

* VARIANCE:

The sensitivity of the model to variations in the training set.

For instance, a sophisticated model with high variance would alter dramatically even with slight modifications to the training set.