|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Soil type | Rainfall(mm) | Temperature(°C) | Humidity(%) | Crop Yield(tons/hectare) |
| Clay | 1200 | 25 | 80 | 4.5 |
| Sandy | 800 | 30 | 60 | 3.0 |
| Loamy | 1000 | 28 | 70 | 4.0 |
| Clay | 1500 | 22 | 85 | 5.0 |
| Sandy | 600 | 35 | 50 | 2.5 |
| Loamy | 1100 | 27 | 75 | 4.2 |

**Terminologies**

1. **Feature**:
   * Individual measurable properties used as inputs to the model.
   * **Example**: Soil Type, Rainfall (mm), Temperature (°C), Humidity (%).
2. **Label**:
   * The output variable that the model aims to predict.
   * **Example**: Crop Yield (tons/hectare).
3. **Prediction**:
   * The output of the model when given a set of features.
   * **Example**: The model predicts that a field with clay soil, 1200 mm rainfall, 25°C temperature, and 80% humidity will yield 4.5 tons/hectare.
4. **Outlier**:
   * A data point that deviates significantly from the rest of the data.
   * **Example**: If there was a field with a crop yield of 8.0 tons/hectare in this dataset.
5. **Test Data**:
   * Data used to evaluate the performance of the model.
   * **Example**: The last row of the dataset could be set aside as test data.
6. **Training Data**:
   * Data used to train the model.
   * **Example**: The first five rows of the dataset.
7. **Model**:
   * A mathematical representation learned from data used to make predictions.
   * **Example**: A regression model predicting crop yield based on features.
8. **Validation Data**:
   * Data used to tune the model parameters.
   * **Example**: A subset of the training data set aside for validation purposes.
9. **Hyperparameter**:
   * Parameters set before training the model, influencing the training process.
   * **Example**: Learning rate, number of epochs.
10. **Epoch**:
    * One complete pass through the training dataset.
    * **Example**: If the dataset is passed through the model 100 times during training, that constitutes 100 epochs.
11. **Loss Function**:
    * A function that measures how well the model's predictions match the actual data.
    * **Example**: Mean Squared Error (MSE) used for regression models.
12. **Learning Rate**:
    * A hyperparameter that controls how much the model is adjusted with respect to the loss gradient.
    * **Example**: A learning rate of 0.01.
13. **Overfitting**:
    * When a model learns the training data too well, including noise, and performs poorly on new data.
    * **Example**: A model that has a high accuracy on training data but low accuracy on test data.
14. **Underfitting**:
    * When a model is too simple to capture the underlying patterns in the data.
    * **Example**: A model that has low accuracy on both training and test data.
15. **Regularization**:
    * Techniques used to prevent overfitting by adding constraints to the model.
    * **Example**: L2 regularization.
16. **Cross-Validation**:
    * A technique for assessing how the results of a model will generalize to an independent data set.
    * **Example**: K-fold cross-validation.
17. **Feature Engineering**:
    * The process of creating new features or modifying existing ones to improve model performance.
    * **Example**: Creating a new feature "Rainfall per Month" from "Rainfall (mm)".
18. **Dimensionality Reduction**:
    * Techniques used to reduce the number of features in a dataset.
    * **Example**: Principal Component Analysis (PCA).
19. **Bias**:
    * Error introduced by approximating a real-world problem, which may be complex, by a simpler model.
    * **Example**: A model that predicts crop yield using only the soil type might have high bias.
20. **Variance**:
    * The model's sensitivity to fluctuations in the training data.
    * **Example**: A complex model that changes significantly with small changes in the training data has high variance.