ASSIGNMENT-1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Dish Name | Category | Description | Price (INR) | Vegetarian | Spicy Level | Availability |
| 201 | Masala Dosa | Main Course | Crispy rice crepe filled with spicy potato | 50 | Yes | Medium | Available |
| 202 | Idli | Breakfast | Steamed rice cakes served with chutney | 30 | Yes | None | Available |
| 203 | Vada | Snack | Deep-fried lentil doughnut | 20 | Yes | Medium | Available |
| 204 | Pongal | Breakfast | Savory rice and lentil porridge | 40 | Yes | Low | Available |
| 205 | Bajji | Snack | Deep-fried fritters made from various vegetables | 25 | Yes | High | Available |
| 206 | Paniyaram | Snack | Savory or sweet dumplings made from fermented batter | 35 | Yes | Medium | Available |
| 207 | Parotta | Main Course | Flaky, layered flatbread | 30 | No | Low | Available |
| 208 | Kothu Parotta | Main Course | Chopped parotta stir-fried with vegetables and meat | 70 | No | High | Available |
| 209 | Murukku | Snack | Crunchy, spiral-shaped snack made from rice flour | 15 | Yes | Low | Available |

From the above table,

* **Feature:** The input variable for the model (e.g., Dish Name, Category, Price (INR), Vegetarian, Spicy Level, Availability).
* **Label:** The output variable that the model predicts (e.g., Vegetarian).
* **Prediction:** The result produced by a ML model based on input features (e.g., predicting whether a dish is vegetarian or not based on its features).
* **Outlier:** A data deviates from the other data. (e.g., Price (INR) of one data is 200 or 1000)
* **Test Data:** An unseen data used to evaluate the performance of a trained model. (e.g., a portion of the dish data not used during training to assess model accuracy).
* **Training Data:** Data used to train the ML model (e.g., the dish data used to teach the model about various features and their relationships).
* **Model:** A mathematical representation of a real-world process that is trained on training data to make predictions (e.g., a model trained to predict whether a dish is vegetarian based on its features).
* **Validation Data:** Set of data from the training data to validate the model after training (e.g., a portion of the dish data used to adjust the learning rate or other hyperparameters).
* **Hyperparameter:** A parameter whose value is set before the learning process starts (e.g., the learning rate or the number of epochs).
* **Epoch:** A complete pass through the entire training dataset (e.g., the model training process making one complete pass through all the dish data).
* **Loss Function:** Measures how well the model's predictions match the actual data and guides the training process (e.g., a function that calculates the difference between predicted and actual vegetarian status).
* **Learning Rate:** A hyperparameter that controls how much the model's weights are adjusted with respect to the loss (e.g., a small learning rate ensures small updates, preventing the model from overshooting the optimal solution).
* **Overfitting:** The trained model works well in the training data but does not works well in testing data (e.g., if the model predicts "Vegetarian" perfectly on the training set but poorly on the test set).
* **Underfitting:** The trained model does not work well in training model (e.g., a model that cannot accurately predict whether a dish is vegetarian because it doesn't consider enough features).
* **Regularization:** The Technique used to prevent overfitting (e.g., adding a regularization term to the loss function).
* **Cross-Validation:** It is the dividing of data into subsets and train the model with different combination of these subsets and also testing with different subset (e.g., using k-fold cross-validation to evaluate the vegetarian prediction model).
* **Feature Engineering:** To improve the model’s performance by creating new features. It also includes modifying of features also (e.g., creating a new feature that combines "Price (INR)" and "Spicy Level" to better predict vegetarian status).
* **Dimensionality Reduction:** This is the techniques used to reduce the number of features by retaining only the important features in the dataset only (e.g., using PCA to reduce the features of dish data).
* **Bias:** The error which predicts only the maximum appeared result not the minimum appeared one. (e.g., for the above dataset the model only predicts the vegetarian but not the non veg.)