|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Generative AI Consortium (Ltd)**  **AI/ML Internship: Assignment 1 (Simple Machine Learning Problem) Assignment)**  **Name: DHAMAYANDHI R** | | **Email:** [**mailto:kavindhamayandhi@gmail.com**](mailto:kavindhamayandhi@gmail.com) | | | | | | |  |
| **ID** | **Height** | **Weight** | **Gender** | **Activity Level** | **Health Status** |
| 1 | 165 cm | 60 kg | Female | Active | Good |
| 2 | 175 cm | 80 kg | Male | Sedentary | Average |
| 3 | 160 cm | 55 kg | Female | Active | Excellent |
| 4 | 180 cm | 90 kg | Male | Moderate | Poor |
| 5 | 170 cm | 70 kg | Female | Active | Good |
| 6 | 155 cm | 50 kg | Female | Sedentary | Average |

FFfffff

**Terminology Explanations:**

**Feature:** The attributes or characteristics used as input to the model (e.g., Height, Weight, Gender).

**Label:** The outcome or target variable that the model aims to predict (e.g., Health Status).

**Prediction:** The model’s forecast based on input features (e.g., predicting the Health Status).

**Outlier:** A data point that significantly deviates from the others (e.g., a record with unusually high Weight).

**Test Data:** The subset of data used to evaluate the model’s performance (e.g., record with ID 6).

**Training Data:** The subset of data used to train the model (e.g., records with ID 1 to 5).

**Model:** The algorithm or method used to make predictions (e.g., Linear Regression).

**Validation Data:** Data used to fine-tune and validate the model’s parameters (e.g., a separate portion of the dataset).

**Hyperparameter:** Pre-set parameters that influence the learning process (e.g., the number of iterations).

**Epoch:** One complete pass through the entire training dataset during the training process.

**Loss Function:** A measure of how well the model’s predictions match the actual values (e.g., Mean Absolute Error).

**Learning Rate:** The step size used to adjust the model’s parameters during training.

**Overfitting:** When the model performs well on training data but poorly on new data.

**Underfitting:** When the model is too simplistic and fails to capture underlying patterns in the data.

**Regularization:** Techniques used to reduce overfitting by adding constraints to the model (e.g., L1 Regularization).

**Cross-Validation:** A technique to evaluate model performance by partitioning the data into training and validation sets multiple times.

**Feature Engineering:** The process of creating new features from existing data to improve model performance.

**Dimensionality Reduction:** Techniques to reduce the number of input variables while retaining essential information (e.g., t-SNE).

**Bias:** Systematic error introduced by the model’s assumptions (e.g., model bias).

**Variance:** The variability of the model’s predictions due to changes in the training da**ta.**