|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Generative AI Consortium (Ltd)**  **AI/ML Internship: Assignment 1 (Simple Machine Learning Problem) Assignment)**  **Name: GUNAL C** | | **Email:** [**mailto:g**](mailto:g)**unalcsg0087@gmail.com** | | | | | | |  |
| **Record No.** | **Years of Experience** | **Annual Income** | **Employed** | **Education Level** | **Is Anomaly** |
| 1 | 2 | 45000 | No | Bachelor | No |
| 2 | 15 | 80000 | Yes | Master | No |
| 3 | 10 | 60000 | Yes | Bachelor | No |
| 4 | 20 | 95000 | Yes | PhD | No |
| 5 | 1 | 35000 | No | High School | No |
| 6 | 12 | 70000 | Yes | Master | No |
| 7 | 50 | 2000000 | No | Outlier | Yes |

FFfffff

**Feature:** Individual independent variables that act like an input in your system.

**Example:** Years of Experience, Annual Income, Education Level.

**Label:** Identification of raw data.

**Example:** Employed.

**Prediction:** Project a probable dataset that relates back to original data.

**Example:** For a new record in the dataset with Years of Experience=8 and Annual Income=65000, the model might predict Yes.

**Outlier:** Data that is unique/different from other data.

**Example:** Record No. 7 where Is Anomaly=Yes.

**Test Data:** Ensure that the model works for the given testing data.

**Example:** Records of Record No. 6 and Record No. 7.

**Training Data:** Data that is used to train the model.

**Example:** Records from Record No. 1 to Record No. 5.

**Model:** Program that can make decisions from previously unseen datasets.

**Example:** Ensemble trees, Decision tree.

**Validation Data:** Uses a sample of data that is with-held from training.

**Example:** Records of Record No. 3 and Record No. 4.

**Hyperparameter:** Parameters that are set before training a model and controlling the learning process.

**Example:** The topology and size of a neural network.

**Epoch:** Each time a dataset passes through an algorithm, it is said to have completed one epoch. Therefore it refers to the one complete passing of training data through the algorithm. **Example:** One pass through records of Record No. 1 to Record No. 5.

**Loss Function:** Quantifies the difference between predicted outputs of a machine learning algorithm and actual target values.

**Example:** Mean Square Error, Mean Absolute Error.

**Learning Rate:** Tuning parameter in an optimization algorithm that determines the step size at each iteration while moving towards a minimum of a loss function.

**Example:** Starting with a learning rate of 0.1 and reduce it by a factor of 0.5 every 10 epochs.

**Overfitting:** A behavior that occurs when the learning model gives accurate predictions for training data but not for new data.

**Example:** The model accurately predicts employment status for training data but fails for new data.