**Generative AI Consortium (Ltd)**

**AI/ML Internship: Assignment 1**

**NAME**:GOWRI SHANKAR V

**EMAIL**:shankar4112004@gmail.com

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Player ID** | **Age** | **Hours Played per Week** | **Total Spend(USD)** | **Level Ahieved** | **VIP Status** | **Is Outlier** |
| 1 | 22 | 10 | 20 | 15 | NO | NO |
| 2 | 30 | 25 | 150 | 30 | YES | NO |
| 3 | 28 | 5 | 5 | 10 | NO | NO |
| 4 | 35 | 40 | 250 | 50 | YES | NO |
| 5 | 18 | 3 | 0 | 5 | NO | NO |
| 6 | 26 | 70 | 5000 | 60 | YES | YES |

**Explanation of Machine Learning Terminologies**

**Feature**: These are the individual measurable properties or characteristics of the data.

* **Example**: In the dataset, Age, Hours Played per Week, Total Spend, and Level Achieved are features.

**Label**: This is the output variable that the model aims to predict.

* **Example**: VIP Status is the label in this dataset.

**Prediction**: This is the model’s output for a given input.

* **Example**: For a new player with Age=24 and Hours Played per Week=15, the model might predict VIP Status as No.

**Outlier**: A data point that differs significantly from other observations.

* **Example**: Player ID=6 where Is Outlier=Yes

**Test Data**: Data used to evaluate the performance of the model.

* **Example**: Records of Player ID=5 and Player ID=6

**Training Data**: Data used to train the model

* **Example**: Records from Player ID=1 to Player ID=4

**Model**: A mathematical representation used to predict outcomes.

* **Example**: Logistic Regression, Decision Tree.

**Validation Data**: Data used to tune the model’s hyperparameters.

* **Example**: Records of Player ID=3 and Player ID=4.

**Hyperparameter**: Parameters set before the learning process begins and control the training process.

* **Example**: The maximum depth of a decision tree, the number of neighbors in a KNN model.

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

* **Example**: One pass through records of Player ID=1 to Player ID=4.

**Loss Function**: A function that measures how well the model’s predictions match the actual data.

* **Example**: Binary Cross-Entropy, Mean Squared Error.

**Learning Rate**: A hyperparameter that controls how much to change the model in response to the estimated error.

* **Example**: Starting with a learning rate of 0.01 and reducing it by a factor of 0.1 every 10 epochs.

**Overfitting**: When a model learns the training data too well, including noise, leading to poor performance on new data.

* **Example**: If the model perfectly predicts VIP Status on the training data but performs poorly on test data.

**Underfitting**: When a model is too simple to capture the underlying structure of the data.

* **Example**: If a linear model fails to capture the non-linear relationship between Hours Played per Week and VIP Status.

**Regularization**: Techniques to prevent overfitting by adding a penalty for larger coefficients in the model.

* **Example**: L1 Regularization, Dropout in neural networks.

**Cross-Validation**: A technique for assessing how the model will generalize to an independent dataset.

* **Example**: Using k-fold cross-validation to evaluate model performance.

**Feature Engineering**: The process of using domain knowledge to create features that make machine learning algorithms work.

* **Example**: Creating a new feature Engagement Level by combining Hours Played per Week and Total Spend.

**Dimensionality Reduction**: Techniques to reduce the number of input variables.

* **Example**: Using Principal Component Analysis (PCA) to reduce the number of features.

**Bias**: Systematic error due to overly simplistic assumptions in the learning algorithm.

* **Example**: If the model assumes a linear relationship between Age and VIP Status, causing it to systematically miss the actual pattern.

**Variance**: Error due to excessive sensitivity to small fluctuations in the training set.

* **Example**: A complex model that changes significantly with small changes in the training data has high variance.