**Generative AI Consortium (Ltd)**

**AI/ML Internship: Assignment 1 (Simple Machine Learning Problem)**

**Name: DEEPIGA P**  
**Email:** deepikagowtham24@gmail.com

**Airline Management: Machine Learning Concepts**

**Dataset Example:**

| **ID** | **Age** | **Flight Cost** | **Purchased** | **Flight Class** | **Is Outlier** |
| --- | --- | --- | --- | --- | --- |
| 1 | 30 | 200 | Yes | Economy | No |
| 2 | 45 | 500 | Yes | Business | No |
| 3 | 35 | 300 | No | Economy | No |
| 4 | 55 | 700 | Yes | First Class | No |
| 5 | 25 | 150 | No | Economy | No |
| 6 | 40 | 400 | Yes | Business | No |
| 7 | 120 | 10000 | No | Outlier | Yes |

**Definitions:**

**Feature:** Individual independent variables that serve as inputs in the system.  
*Example:* Age, Flight Cost, Flight Class.

**Label:** The identification or target value in the dataset.  
*Example:* Purchased.

**Prediction:** Projecting a probable outcome based on the input data.  
*Example:* For a new record with Age=32 and Flight Cost=350, the model might predict No.

**Outlier:** Data that significantly deviates from other data.  
*Example:* id=7 where the outlier=yes.

**Test Data:** Data used to ensure the model works effectively.  
*Example:* Records of id=6 and id=7.

**Training Data:** Data used to train the model.  
*Example:* Records from id=1 to id=5.

**Model:** A program capable of making decisions from previously unseen datasets.  
*Example:* Random Forest, Neural Network.

**Validation Data:** A subset of data withheld from training for model validation.  
*Example:* Records of id=3 and id=4.

**Hyperparameter:** Parameters set before training a model to control the learning process.  
*Example:* The number of trees in a random forest.

**Epoch:** One complete pass of the training dataset through the algorithm.  
*Example:* One pass through records of id=1 to id=5.

**Loss Function:** Quantifies the difference between predicted outputs and actual values.  
*Example:* Cross-Entropy Loss, Mean Absolute Error.

**Learning Rate:** A parameter that determines the step size at each iteration while moving towards a loss function's minimum.  
*Example:* Starting with a learning rate of 0.01 and reducing it by half every 20 epochs.

**Overfitting:** When the model performs well on training data but poorly on new data.  
*Example:* A model that memorizes training data instead of learning general patterns.

**Underfitting:** When a model is too simple and fails to capture the data patterns.  
*Example:* A linear model for a complex dataset.

**Regularization:** Techniques to reduce overfitting.  
*Example:* L1 Regularization, Dropout in Neural Networks.

**Cross-Validation:** Resampling technique to evaluate models on different subsets of data.  
*Example:* k-Fold Cross-Validation.

**Feature Engineering:** Creating new variables from existing data to improve model performance.  
*Example:* Creating a 'Loyalty' feature by binning frequent flyer miles.

**Dimensionality Reduction:** Reducing the number of variables in a dataset.  
*Example:* Principal Component Analysis (PCA).

**Bias:** Systematic error due to incorrect assumptions in the model.  
*Example:* Assuming all flights have the same cost regardless of class.

**Variance:** Changes in the model's performance with different data portions.  
*Example:* A highly complex model showing significant performance changes with small dataset variations.