|  |
| --- |
| FFfffff |
| **Generative AI Consortium (Ltd)**  **AI/ML Internship**: Assignment 1 (Simple Machine Learning Problem)  **Name**: JAYA NIRENJAN A C |
| **Email:** [**jayanirenjan@gmail.com**](mailto:jayanirenjan@gmail.com)  The list of terminologies using an example dataset of Laptops.  Here's an example dataset:   | **Laptop ID** | **Brand** | **Model** | **Processor Type** | **RAM Size** | **Storage Capacity** | **Price (USD)** | **Customer Rating** | | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | Dell | Inspiron 15 | Intel i5 | 8GB | 256GB SSD | 650 | 4.3 | | 2 | HP | Pavilion x360 | Intel i7 | 16GB | 512GB SSD | 850 | 4.5 | | 3 | Lenovo | ThinkPad X1 | Intel i7 | 16GB | 1TB SSD | 1200 | 4.7 | | 4 | Apple | MacBook Air | Apple M1 | 8GB | 256GB SSD | 999 | 4.8 | | 5 | Acer | Aspire 5 | AMD Ryzen 5 | 8GB | 512GB SSD | 600 | 4.2 |   **TERMINOLOGIES EXPLAINED:**  **1. Feature:** Individual measurable properties or characteristics used as inputs to the model.   * **Example:** Laptop ID, Brand, Model, Processor Type, RAM Size, Storage Capacity, Price, Customer Rating.   **2. Label:** The output variable that the model aims to predict.   * **Example:** Price (if we are trying to predict the price of a laptop based on the other features).   **3. Prediction:** The output of the model after it has been trained, given new input data.   * **Example:** Predicting the price for a new laptop based on its processor type, RAM size, storage capacity, etc.   **4. Outlier:** A data point that deviates significantly from the rest of the data.   * **Example:** If there was a laptop with a price of $10,000 in the dataset.   **5. Test Data:** A subset of the dataset used to assess the performance of the model.   * **Example:** The last two records in the table (Laptop ID 4 and 5) can be used as test data.   **6. Training Data:** The subset of the dataset used to train the model.   * **Example:** The first three records in the table (Laptop ID 1, 2, and 3) can be used as training data.   **7. Model:** The mathematical representation of the relationship between features and labels.   * **Example:** A linear regression model predicting the price of a laptop based on its RAM size and processor type.   **8. Validation Data:** A subset of the dataset used to tune the hyperparameters of the model.   * **Example:** A separate set of laptop records not in the training or test set.   **9. Hyperparameter:** Parameters whose values are set before the learning process begins.   * **Example:** The number of layers in a neural network or the learning rate.   **10. Epoch:** One complete pass through the entire training dataset.   * **Example:** In a neural network, an epoch would mean the model has seen each laptop record in the training data once.   **11. Loss Function**A function that measures the discrepancy between the predicted and actual labels.   * **Example:** Mean Squared Error (MSE) used in regression tasks to predict laptop prices.   **12. Learning Rate:** A hyperparameter that controls how much the model's parameters are adjusted with respect to the loss gradient.   * **Example:** A learning rate of 0.01.   **13. Overfitting:** When a model performs well on training data but poorly on unseen data.   * **Example:** If our model predicts the laptop prices perfectly on training data but fails on test data.   **14. Underfitting:** When a model is too simple to capture the underlying pattern in the data.   * **Example:** If our model has a low prediction accuracy on both training and test data for laptop prices.   **15. Regularization:** Techniques to prevent overfitting by penalizing large coefficients.   * **Example:** L1 or L2 regularization applied to a regression model predicting laptop prices.   **16. Cross-Validation:** A technique for assessing how the results of a statistical analysis will generalize to an independent dataset.   * **Example:** k-fold cross-validation used to evaluate the performance of a model predicting laptop prices.   **17. Feature Engineering:** The process of using domain knowledge to create features that make machine learning algorithms work.   * **Example:** Creating a new feature 'Price per GB of RAM' from the price and RAM size of the laptops.   **18. Dimensionality Reduction:** Techniques to reduce the number of input variables.   * **Example:** Principal Component Analysis (PCA) used to reduce the number of features in a dataset of laptop specifications.   **19. Bias:** Error due to overly simplistic assumptions in the learning algorithm.   * **Example:** High bias could mean the model is too simple and fails to accurately predict laptop prices (underfitting).   **20. Variance:** Error due to too much complexity in the learning algorithm.   * **Example:** High variance could mean the model is too complex and predicts laptop prices accurately on training data but poorly on test data (overfitting). |