

Generative AI Consortium (Ltd)

AI/ML Internship: Assignment 1

Name: HARSHITHA KANNAN

Email: harshithakannan3@gmail.com

Year Released	Brand	Model	Storage (GB)	RAM (GB)	Condition	Resale Price (INR)
2019	Apple	iPhone X	64	3	Good	35000
2018	Samsung	Galaxy S9	128	4	Excellent	30000
2020	One Plus	8 Pro	256	12	Good	45000
2017	Xiaomi	Mi Mix 2	64	4	Fair	15000
2021	Google	Pixel 5	128	8	Excellent	50000

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

- **Example:** Year Released, Brand, Model, Storage (GB), RAM (GB), Condition.

Label: Identification of raw data.

- **Example:** Resale Price.

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

- **Example:** For a new record in the dataset with Year Released=2019, Brand=Apple, and Storage=64, the model might predict Resale Price=35000.

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

- **Example:** If there was a smartphone with a resale price at 100000 INR in this dataset, it would be considered an outlier.

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

- **Example:** Records of id=4 and id=5..

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

- **Example:** Records from id=1 to id=3.

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 id=2 and id=3.

Hyper parameter: 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 an algorithm.

- **Example:** One passes through records of id=1 to id=3.

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 reducing it by a factor of 0.5 every 10 epochs.

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

- **Example:** If the model perfectly predicts the resale prices of the training data but fails to predict prices in the test data accurately.

Underfitting: When a model is too simple and has not learned the patterns in the training data well and is unable to generalize well on the new data.

- **Example:** If the model predicts all smartphone prices around 35000 INR regardless of their features.

Regularization: Set of methods to reduce overfitting.

- **Example:** L2 Regularization.

Cross-validation: Technique of resampling different portions of training data for validation on different iterations.

- **Example:** K-fold cross-validation, where the dataset is divided into K subsets and the model is trained and validated K times.

Feature Engineering: Technique that leverages data to create new variables that aren't in the training set.

- **Example:** Converting the "Year Released" to the "Age" of the smartphone.

Dimensional Reduction: Method of reducing variables in a training dataset used to develop machine learning models.

- **Example:** Principal component analysis(PCA).

Bias: Systematic error that occurs in the model itself due to incorrect assumptions on the machine learning process.

- **Example:** Sample Bias.

Variance: Changes in the model when using different portions of the training dataset.

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