|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Book ID** | **Book Name** | **Price (INR)** | **Borrowed** | **Popularity Level** | **Is Special Edition** |
| 1 | "Mystery of the Manor" | ₹3,750 | No | Medium | No |
| 2 | " Science of the Cosmos " | ₹7,500 | Yes | High | No |
| 3 | "History of the World" | ₹5,625 | No | Medium | No |
| 4 | "Modern Economics" | ₹9,000 | Yes | High | No |
| 5 | "Beginner's Guide to Cooking" | ₹3,000 | No | Low | No |
| 6 | "Advanced Physics" | ₹6,750 | Yes | High | No |
| 7 | "Rare Artifacts of Ancient Times" | ₹3,75,000 | No | Special | Yes |

ASSIGNMENT – 1

**NAME: BAIRAVI PL**

**EMAIL: bairavi2407@gmail.com**

* **Feature**: Individual attributes that serve as inputs in a model.
  + **Example**: Price, Popularity Level, Is Special Edition.
* **Label**: Target variable representing the outcome.
  + **Example**: Borrowed.
* **Prediction**: Forecast of outcomes based on new data.
  + **Example**: For a new book with the name "Science for Beginners" and Price=₹6,000, the model might predict "No" for Borrowed.
* **Outlier**: Data point significantly different from other data points.
  + **Example**: Book ID=7 where Is Special Edition=Yes.
* **Test Data**: Data used to evaluate the model's performance.
  + **Example**: Records of Book ID=6 and Book ID=7.
* **Training Data**: Data used to train the model.
  + **Example**: Records from Book ID=1 to Book ID=5.
* **Model**: Algorithmic structure that makes predictions based on data.
  + **Example**: Decision Trees, Ensemble Models.
* **Validation Data**: Data used to fine-tune the model during training.
  + **Example**: Records of Book ID=3 and Book ID=4.
* **Hyperparameter**: Settings configured before training a model, controlling the learning process.
  + **Example**: The number of trees in a random forest model.
* **Epoch**: One complete pass of the training dataset through the learning algorithm.
  + **Example**: One pass through records of Book ID=1 to Book ID=5.
* **Loss Function**: Quantifies the difference between predicted outputs and actual target values.
  + **Example**: Mean Squared Error (MSE), Mean Absolute Error (MAE).
* **Learning Rate**: Tuning parameter in an optimization algorithm that determines the step size at each iteration.
  + **Example**: Starting with a learning rate of 0.1 and reducing it by a factor of 0.5 every 10 epochs.
* **Overfitting**: Occurs when a model gives accurate predictions for training data but not for new data.
  + **Example**: A complex model that fits the training data too well.
* **Underfitting**: When a model is too simple and fails to capture the patterns in the training data.
  + **Example**: A linear model applied to highly non-linear data.
* **Regularization**: 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.
* **Feature Engineering**: Creating new variables from existing data to improve model performance.
  + **Example**: Creating a "Popularity Level" feature by categorizing borrow counts.
* **Dimensional Reduction**: Methods of reducing the number of variables in a dataset.
  + **Example**: Principal Component Analysis (PCA).
* **Bias**: Systematic error introduced by incorrect assumptions in the model.
  + **Example**: Sample Bias.
* **Variance**: The model's sensitivity to fluctuations in the training data.
  + **Example**: A complex model that changes significantly with small changes in the training data has high variance.