|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Generative AI Consortium (Ltd)**  **AI/ML Internship: Assignment 1 (Simple Machine Learning Problem)**  **Name: KAMALESHWARAN A** | | **Email: kamaleshathappan@gmail.com**  **Employee Promotion Prediction** | | | | | | |  |
| **Emp ID** | | **Age** | **Department** | **Years of Experience** | **Last Promotion Years Ago** | **Promotion Status** |
| 1 | | 30 | Sales | 5 | 1 | Not Promoted |
| 2 | | 35 | Engineering | 8 | 2 | Not Promoted |
| 3 | | 28 | HR | 3 | 1 | Not Promoted |
| 4 | | 40 | Finance | 12 | 3 | Promoted |
| 5 | | 32 | Sales | 7 | 2 | Not Promoted |
| 6 | | 45 | Engineering | 15 | 5 | Promoted |
| 7 | | 27 | HR | 5 | 3 | Promoted |

FFfffff

**Feature:** Individual measurable properties used as inputs to the model.

* **Example:** Age, Department, Years of Experience, Last Promotion Years Ago.

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

* **Example:** Promotion Status (Promoted or Not Promoted).

**Prediction:** The process of using a trained model to make predictions or decisions based on input data.

* **Example:** Predicting whether an employee will be promoted based on their features.

**Outlier:** A data point that deviates significantly from the rest of the data.

* **Example:** An employee with exceptionally high or low years of experience compared to others in the same department.

**Training Data:** Data used to train the machine learning model.

* **Example:** The first 5 rows in the provided table.

**Test Data:** Data used to evaluate the performance of the trained model on unseen data.

* **Example:** The last row in the provided table could be used as a test data point.

**Model:** A trained machine learning algorithm capable of making predictions.

* **Example:** A model trained to predict promotion status based on employee data.

**Validation Data:** Data used to tune hyperparameters and evaluate the model during training.

* **Example:** A subset of the training data reserved for validating model performance.

**Hyperparameter:** Parameters that are set before the learning process begins.

* **Example:** Learning rate, number of epochs, choice of optimizer.

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

* **Example:** Training a model for 10 epochs means it has seen the entire dataset 10 times.

**Loss Function:** Measures how well the model's predictions match the actual results.

* **Example:** Cross-entropy loss, which quantifies the difference between predicted and actual promotion status.

**Learning Rate:** Determines how much to change the model in response to the estimated error each time the model weights are updated.

* **Example:** A low learning rate means slower convergence but potentially more accurate results.

**Overfitting:** When a model learns both the underlying patterns and noise in the training data.

* **Example:** A model that memorizes specific employee details rather than generalizing promotion criteria.

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

* **Example:** A linear model trying to predict promotion decisions with limited accuracy.

**Regularization:** Technique used to prevent overfitting by adding a penalty term to the loss function.

* **Example:** L2 regularization, which penalizes large coefficients in the model.

**Cross-Validation:** Technique used to assess how the results of a statistical analysis will generalize to an independent dataset.

* **Example:** Performing k-fold cross-validation to evaluate model performance across different subsets of the training data.

**Feature Engineering:** Process of using domain knowledge to select and transform features that are most relevant to the predictive modeling problem.

* **Example:** Creating new features such as "Years since Last Promotion" from existing data.

**Dimensionality Reduction:** Technique used to reduce the number of input variables in predictive modeling while still retaining most of the relevant information.

* **Example:** PCA (Principal Component Analysis), which transforms the original features into a smaller set of principal components.

**Bias:** Error that occurs when a model consistently predicts outcomes that are systematically off-target.

* **Example:** A model consistently underestimating the likelihood of promotion for certain age groups.

**Variance:** Error that occurs when a model is too sensitive to small fluctuations in the training data, leading to overfitting.

* **Example:** A model that performs well on the training set but poorly on new employee data.