## **Understanding the Problem:**

### ****Regression:****

* Predicts a **continuous numerical value.**

### ****Classification:****

* Predicts a **discrete class label.**
* Example Predicting if an employee gets promoted (Yes = 1, No = 0).

### ****Regression as Classification:****

We can train a **linear regression model** to output a continuous score, then **map that score to a class label** using decision boundaries (thresholds).

**For example:**

* y < 0.5y ​< 0.5 → Class A (Not Promoted)
* y ≥ 0.5y ≥ 0.5 → Class B (Promoted)

## **Dataset Description:-**

|  |  |  |
| --- | --- | --- |
| **Feature** | **Type** | **Description** |
| date\_of\_birth | string | Age in years |
| date\_of\_joining | string | Years of Service |
| gender | integer | 0 = Female, 1 = Male,2=Other |
| promoted | integer | 0 (not promoted) / 1 (promoted) |

## **Process to Implement:-**

**Step 1 – Train a Multi-Variable Linear Regression Model**

* Features: gender, age, years of service.
* Target: promoted (0 = No, 1 = Yes).

**Step 2 – Convert Regression Output to Class**

* Use threshold = 0.5:
  + If prediction < 0.5 → Class A (Not promoted)
  + Else → Class B (Promoted)

## **Changes Needed in Multi-Variable Linear Regression Code for Classification:-**

1. **Target Variable Encoding**
   * Regression: y is continuous.
   * Classification: Encode categories (0, 1, 2, …).
2. **Post-processing Predictions**
   * Regression: Use the continuous output.
   * Classification: Apply threshold or mapping function to assign class labels.
3. **Evaluation Metrics**
   * Regression: MSE, RMSE, R2.
   * Classification: Accuracy, Precision, Recall, F1-score, Confusion Matrix.
4. **Optional Upgrade**
   * Instead of LinearRegression + threshold, use Logistic Regression for better performance in binary classification.

## **Conclusion:-**

This experiment demonstrates that a regression model can be adapted for classification tasks by converting continuous predictions into discrete categories using thresholds. Although this approach works for simple binary classification problems, it is often less accurate than using specialized classification algorithms such as logistic regression, decision trees, or ensemble methods. For datasets like employee promotions, where class boundaries are clear, direct classification models generally offer better interpretability and performance.