

Intel® AI for Manufacturing Certificate Course

Week-5 – Assignment: Evaluation Metrics Report

Name: Ayan Memon

Submission Date: 16 – 03 – 2025

1. Introduction

In this assignment, we explore different evaluation metrics used for assessing machine learning models. We focus on:

- **Regression Metrics** – Used for models that predict continuous values.
- **Classification Metrics (Confusion Matrix)** – Used for models that classify inputs into categories.

The goal is to understand these metrics and apply them using an example AI model.

2. Evaluation Metrics for Regression

Regression models predict continuous values (e.g., sales revenue, temperature). We evaluate them using the following metrics:

2.1 Mean Absolute Error (MAE)

Definition: Measures the average absolute difference between actual and predicted values.

Formula:

$$\text{MAE} = (1/n) * \sum | \text{Actual} - \text{Predicted} |$$

Example Use Case: Used in real estate price prediction, where small errors matter.

2.2 Mean Squared Error (MSE)

Definition: Similar to MAE but squares the differences, giving more weight to large errors.

Formula:

$$\text{MSE} = (1/n) * \sum (\text{Actual} - \text{Predicted})^2$$

Example Use Case: Used in weather forecasting where large deviations are critical.

2.3 Root Mean Squared Error (RMSE)

Definition: The square root of MSE, bringing the error metric back to the original unit.

Formula:

$$\text{RMSE} = \sqrt{\text{MSE}}$$

Example Use Case: Used in stock price prediction to measure deviations.

2.4 R-squared (R² Score)

Definition: Measures how well the model explains the variability of the target variable.

Formula:

$$R^2 = 1 - (\text{SS}_{\text{residual}} / \text{SS}_{\text{total}})$$

where:

$$\text{SS}_{\text{residual}} = \sum (\text{Actual} - \text{Predicted})^2$$

$$\text{SS}_{\text{total}} = \sum (\text{Actual} - \text{Mean}(\text{Actual}))^2$$

Example Use Case: Used in finance to determine how well a model predicts market trends.

3. Confusion Matrix & Classification Metrics

Classification models predict categories (e.g., Spam vs. Not Spam).

3.1 What is a Confusion Matrix?

A confusion matrix is a table that summarizes a model's performance in classification. It helps in calculating **Precision, Recall, and F1-score**.

3.2 Example AI Model: Robot Identifying Humans in Pictures

Our AI model predicts whether an image contains a human (1) or not (0).

Actual vs. Predicted Results

Input #	Actual	Predicted
1	0	0
2	0	1
3	1	1
4	0	1
5	1	0
6	1	1
7	0	0
8	1	1
9	0	1
10	1	1

From this dataset, we build the **Confusion Matrix**:

Actual / Predicted	Human (1)	Not Human (0)
Human (1) (TP + FN)	3 (TP)	1 (FN)
Not Human (0) (FP + TN)	3 (FP)	3 (TN)

3.3 Performance Metric Calculations

Precision (Positive Predictive Value)

Definition: Measures how many predicted "Humans" are actually correct.

Formula:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Calculation:

$$\text{Precision} = 3 / (3 + 3) = 0.5$$

Meaning: When the AI predicts "Human," it is correct 50% of the time.

Recall (Sensitivity or True Positive Rate)

Definition: Measures how many actual humans were correctly identified.

Formula:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Calculation:

$$\text{Recall} = 3 / (3 + 1) = 0.75$$

Meaning: The AI successfully identifies 75% of actual humans.

F1-Score (Balance Between Precision & Recall)

Definition: The harmonic mean of Precision and Recall.

Formula:

$$F1 = 2 * (Precision * Recall) / (Precision + Recall)$$

Calculation:

$$F1 = 2 * (0.5 * 0.75) / (0.5 + 0.75) = 0.6$$

Meaning: The AI achieves a 60% balance between precision and recall.

4. Conclusion

- We explored regression and classification evaluation metrics.
- **MAE, MSE, RMSE, and R²** were used for regression models.
- **Confusion Matrix, Precision, Recall, and F1-score** were applied to a classification problem.
- The AI model identified humans with **50% precision and 75% recall**, meaning it missed fewer humans but also made some false detections.

This structured approach ensures accurate model evaluation before deployment.

Submission Files

- **Jupyter Notebook Link:**

https://github.com/AyanMemon296/Intel-AI-Certification/blob/main/Weekly_Assignments/Week-05/Ayan_Week-05_Evaluation_Metrics.ipynb