📘Tutorial: Evaluating Regression Models with MAE, RMSE, and R^2

When we train a regression model, we need to check **how good it is**. The three most common metrics are:

- **MAE (Mean Absolute Error)** → Average size of errors
- **RMSE (Root Mean Squared Error)** → Like MAE, but punishes big mistakes
- **R^2 (Coefficient of Determination)** → How well the model explains the data

---

🩺1. MAE – Mean Absolute Error

Formula

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

👉Steps:

1. Find absolute error for each prediction.
2. Take the average.

🕐2. RMSE – Root Mean Squared Error

Formula

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

👉Steps:

1. Square the errors.
2. Take average.
3. Take square root.

🩻3. R^2 – Coefficient of Determination

Formula

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

👉Steps:

1. Compute total variance of data ((SS_{tot})).
2. Compute error of model ((SS_{res})).
3. Compare:
4. If model is perfect → (R^2 = 1)
5. If model is average guess → (R^2 = 0)
6. If model is worse → (R^2 < 0)

📊Example 1: Small Dataset

(y) (Actual)	(\hat{y}) (Predicted)
10	12
8	9
12	11
15	14
6	10

Step 1: Errors

$$Errors = [-2, -1, 1, 1, -4]$$

- Absolute errors = [2, 1, 1, 1, 4]
- Squared errors = [4, 1, 1, 1, 16]

Step 2: MAE

$$MAE = \frac{2 + 1 + 1 + 1 + 4}{5} = 1.8$$

Step 3: RMSE

$$RMSE = \sqrt{\frac{4+1+1+1+16}{5}} = \sqrt{4.6} \approx 2.14$$

Step 4: R^2

$$\bar{y} = 10.2, \quad SS_{tot} = 48.8, \quad SS_{res} = 23$$

$$R^2 = 1 - \frac{23}{48.8} = 0.53$$

👈**Results**: MAE = 1.8, RMSE = 2.14, R^2 = 0.53

📊Example 2: Better Model

(y) (Actual)	(\hat{y}) (Predicted)
20	21
15	14
25	26
30	29
18	18

Step 1: Errors

$$Errors = [-1, 1, -1, 1, 0]$$

- Absolute errors = [1, 1, 1, 1, 0]
- Squared errors = [1, 1, 1, 1, 0]

Step 2: MAE

$$MAE = \frac{1+1+1+1+0}{5} = 0.8$$

Step 3: RMSE

$$RMSE = \sqrt{\frac{1+1+1+1+0}{5}} = \sqrt{0.8} \approx 0.89$$

Step 4: R^2

$$\bar{y} = 21.6, \quad SS_{tot} = 125.2, \quad SS_{res} = 4$$

$$R^2 = 1 - \frac{4}{125.2} = 0.968$$

👉**Results**: MAE = 0.8, RMSE = 0.89, R^2 = 0.97 (much better fit).

🔗Conclusion

- **MAE** → average size of mistakes.
- **RMSE** → like MAE, but bigger mistakes count more.
- **R^2** → how much variation is explained by the model.

In Example 1 → weak model (errors larger, R^2 = 0.53). In Example 2 → strong model (errors small, R^2 = 0.97).

Now let's create a **Jupyter Notebook** with these two examples:

👉Done! I created the Jupyter Notebook:

Download regression_metrics_tutorial.ipynb

It contains:

- Explanations of **MAE, RMSE, and R^2**
- Two full examples (manual + scikit-learn calculations)
- Output printing for easy comparison

Would you like me to also **add plots** (error distributions, residual plots, etc.) so students can *see* the differences, not just numbers?