1. Mean Absolute Error (MAE)

- **Definition**: MAE is the average of the absolute differences between the predicted values and the actual values.
- **Formula**:

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

  where:

  - y_i = actual value
  - $\hat{y}_i$ = predicted value
  - n = number of data points

- **Explanation**:
  MAE gives an equal weight to all errors, irrespective of whether they are small or large. It measures how far predictions are, on average, from the actual values.

- **Example**:
  Suppose you have 3 actual values $y = [10, 15, 20]$ and predicted values $\hat{y} = [12, 14, 18]$.

$$\text{Errors} = [|10 - 12|, |15 - 14|, |20 - 18|] = [2, 1, 2]$$

$$\text{MAE} = \frac{(2 + 1 + 2)}{3} = \frac{5}{3} \approx 1.67$$

  So the average error is **1.67 units**.

- **Pros**: Simple to interpret.
- **Cons**: It does not emphasize large errors, as it treats all errors equally.

2. Mean Squared Error (MSE)

- **Definition**: MSE is the average of the squared differences between the predicted values and the actual values.
- **Formula**:

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

- **Explanation**:
MSE penalizes larger errors more because the errors are squared. This makes it sensitive to outliers.
- **Example**:
Using the same actual values $y = [10, 15, 20]$ and predicted values $\hat{y} = [12, 14, 18]$:

$$\text{Squared Errors} = [(10-12)^2, (15-14)^2, (20-18)^2] = [4, 1, 4]$$

$$\text{MSE} = \frac{(4+1+4)}{3} = \frac{9}{3} = 3$$

The MSE is **3**.

- **Pros**: Large errors have a higher impact, making MSE useful when large deviations need to be heavily penalized.
- **Cons**: Because of squaring, the units of MSE are not the same as the original data.

3. Root Mean Squared Error (RMSE)

- **Definition:** RMSE is the square root of the MSE. It measures the standard deviation of errors.
- **Formula**:

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

- **Explanation**:
  RMSE brings the error back to the same unit as the target variable, making it easier to interpret. Like MSE, it is also sensitive to large errors.
- **Example**:
  From the previous MSE calculation:

$$\text{RMSE} = \sqrt{3} \approx 1.73$$

  So the RMSE is **1.73**.

- **Pros**: The same unit as the target variable makes interpretation straightforward.
- **Cons**: Sensitive to large errors because of squaring.

4. Mean Absolute Percentage Error (MAPE)

- **Definition**: MAPE is the average of the absolute percentage differences between the predicted values and the actual values.
- **Formula**:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

- **Explanation**:
  MAPE expresses the error as a percentage of the actual value. This makes it easier to interpret the scale of errors.
- **Example**:
  Using the same values $y = [10, 15, 20]$ and $\hat{y} = [12, 14, 18]$:

$$\text{Errors (\%)} = \left[\left|\frac{10-12}{10}\right| \times 100, \left|\frac{15-14}{15}\right| \times 100, \left|\frac{20-18}{20}\right| \times 100\right]$$

$$= [20\%, 6.67\%, 10\%]$$

$$\text{MAPE} = \frac{(20 + 6.67 + 10)}{3} = \frac{36.67}{3} \approx 12.22\%$$

The average percentage error is **12.22%**.

- **Pros**: Results are expressed as percentages, which are easy to interpret.
- **Cons**: It can produce issues when actual values are close to zero, as it causes very large percentage errors.

Key Insights

1. **MAE** is more **robust to outliers** than MSE/RMSE because it does not square the errors.
2. **MSE** and **RMSE** emphasize larger errors, so they are more sensitive to outliers.
3. **RMSE** is more interpretable than MSE because it uses the same units as the target variable.
4. **MAPE** is particularly useful for understanding relative errors as percentages.

The choice of metric depends on the specific needs of the problem. For example:

- Use **RMSE/MSE** when large errors need to be penalized.
- Use **MAE** when you want to treat all errors equally.
- Use **MAPE** for interpretability as percentages (especially in business contexts).

3. Example

Suppose you train two models and compare the errors:

Metric	Model A	Model B
MAE	2.1	1.2
MSE	5.3	2.8
RMSE	2.3	1.7
MAPE (%)	12.5	6.8

Here:

- **Model B** has smaller values for all metrics.
- This indicates that **Model B** performs better and has more accurate predictions compared to **Model A**.

4. Important Notes

1. **Smaller Metrics Are Better**: Always aim for smaller error values when comparing models.
2. **Context Matters**: Compare metrics against a **baseline** (e.g., a naive model or a threshold) to judge how good "small" is.
3. **MAPE Limitations**: Be cautious with **MAPE** if actual values are close to zero because it can produce very high percentages.