Metrics are used to evaluate the outcome of your model. While there are recommendations on which metric is appropriate for the model you are currently using, each model object has different variations that may provide additional insight.

Regression Metrics

- Compares predicted output values with real output values
- The difference between predicted and real value determines the model performance

R2

R2 measures the proportion of variance between values. It is somewhat related to correlation and has a value that ranges from -1 to 1. The higher the R2 value, the better the model.

RMSE - Root Mean Squared Error

RMSE measures the standard deviation of prediction errors to target values. Another term for this is residuals. The lower the RMSE, the better the model.

Classification Metrics

- Compares the predicted label with the real label value
- Options to do comparison are either 2 label classification or multiple label classification

Quick Refresher on Prediction Outcomes

The following metrics are calculated by how well a model can classify a dataset.

The table below shows the 4 types of model results:

	Positive	Negative
True	TP - Correctly labeled as Yes	TN - Correctly labeled as No
False	FP - Incorrectly labeled as Yes	FN - Incorrectly labeled as No

For example, you work at a publishing company and you want your model to identify who out of 100 people have read a specific book and here are the results:

- 40 TP - people who have read the book and are labeled as Yes by the model
- 35 TN - people who have not read the book and are labeled as No by the model
- 15 FP - people who have not read the book and are labeled as Yes by the model
- 10 FN - people who have read the book and are labeled as No by the model

Accuracy

Accuracy is calculated by the total correct predictions divided by the total number of data points.

Calculated by:

$$accuracy = \frac{TP + TN}{Total}$$

Using the example above the accuracy would be calculated as follows:

$$accuracy = \frac{40 + 35}{100} = \frac{75}{100} = 0.75$$

Precision

Precision is another metric related to accuracy but explains how good the model is at identifying the relevant label. Calculated by the number of true positives divided by the number of true and false positives.

$$precision = \frac{TP}{TP + FP}$$

Using the example above the precision would be calculated as follows:

$$precision = \frac{40}{40 + 15} = \frac{40}{55} = 0.7272$$

Recall

Measures how many relevant labels are actually selected. Calculated by the number of true positives divided by the number of true positives and false negatives.

$$recall = \frac{TP}{TP + FN}$$

Using the example above the recall would be calculated as follows:

$$recall = \frac{40}{40 + 10} = \frac{40}{50} = 0.8$$

Trade-Offs

While default metrics for algorithms are good for initial evaluation, it is recommended to calculate several metrics to have a better understanding of overall model performance

Additional Resources

- For a list of metrics from Scikit-learn, check out their **metrics page** in the documentation.
- A blog post about **20 Popular ML Metrics** provides some additional information about metrics not mentioned here.