

What is a “good” accuracy for machine learning models?

Authored by
stats writer

June 28, 2024

RECOMMENDED CITATION

stats writer (2024). *What is a “good” accuracy for machine learning models?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=156267>

A good accuracy for machine learning models refers to the percentage of correct predictions made by the model. It is a measure of how well the model is able to accurately classify or predict data. A higher accuracy indicates a better performing model, while a lower accuracy may indicate the need for improvement. The determination of a "good" accuracy for a specific model depends on the specific task and data being analyzed. Generally, a model with an accuracy of 70% or higher is considered to be good, but this may vary depending on the complexity of the problem and the industry standards. Overall, a good accuracy is essential for producing reliable and effective results in machine learning.

What is a "Good" Accuracy for Machine Learning Models?

When using in machine learning, one metric we often use to assess the quality of a model is accuracy.

Accuracy is simply the percentage of all observations that are correctly classified by the model.

It is calculated as:

Accuracy = (# True Positives + # True Negatives) / (Total Sample Size)

One question that students often have about accuracy is:

What is considered a "good" value for the accuracy of a machine learning model?

While the accuracy of a model can range between 0% and 100%, there is no universal threshold that we use to determine if a model has "good" accuracy or not.

Instead, we typically compare the accuracy of our model to the accuracy of some baseline model.

A baseline model is one that simply predicts every observation in a dataset to belong to the most common class.

In practice, any classification model that has a higher accuracy than a baseline model can be considered "useful" but obviously the greater the difference in accuracy between our model and a baseline model, the better.

The following example shows how to roughly determine if a classification model has "good" accuracy or not.

Example: Determining if a Model Has "Good" Accuracy

Suppose we use a model to predict whether or not 400 different college basketball players get drafted into the NBA.

The following summarizes the predictions made by the model:

		Predicted	
		Drafted = Yes	Drafted = No
Actual	Drafted = Yes	120 (True Positive)	40 (False Negative)
	Drafted = No	70 (False positive)	170 (True Negative)

Here is how to calculate the accuracy of this model:

$$\text{Accuracy} = (\# \text{ True Positives} + \# \text{ True Negatives}) / (\text{Total Sample Size})$$
$$\text{Accuracy} = (120 + 170) / (400)$$
$$\text{Accuracy} = 0.725$$

To get an idea of whether or not that is accuracy is "good", we can calculate the accuracy of a baseline model.

In this example, the most common outcome for the players was to not get drafted. Specifically, 240 out of 400 players did not get drafted.

A baseline model would be one that simply predicts every single player to not get drafted.

The accuracy of this model would be calculated as:

$$\text{Accuracy} = (\# \text{ True Positives} + \# \text{ True Negatives}) / (\text{Total Sample Size})$$
$$\text{Accuracy} = (0 + 240) / (400)$$
$$\text{Accuracy} = 0.6$$

This baseline model would correctly predict the outcome for 60% of players.

In this scenario, our logistic regression model offers a noticeable improvement in accuracy compared to a baseline model so we would consider our model to at least be "useful."

In practice, we would likely fit several different classification models and choose the final model as the one that offers the greatest boost in accuracy compared to a baseline model.

Cautions on Using Accuracy to Assess Model Performance

Accuracy is a commonly used metric because it's easy to interpret.

For example, if we say that a model is 90% accurate, we know that it correctly classified 90% of observations.

However, accuracy does not take into account how the

data is distributed.

For example, suppose 90% of all players do not get drafted into the NBA. If we have a model that simply predicts every player to not get drafted, the model would correctly predict the outcome for 90% of the players.

This value seems high, but the model is actually unable to correctly predict any player who gets drafted.

An alternative metric that is often used is called the F1 Score, which takes into account how the data is distributed.

For example, if the data is highly imbalanced (e.g. 90% of all players do not get drafted and 10% do get drafted) then F1 score will provide a better assessment of model performance.

Read more about the differences between accuracy and F1 score .

Additional Resources

The following tutorials provide additional information

about metrics used in classification models in machine learning:

ARABPSYCHOLOGY.COM