As data scientists, we are always pursuing “The Best” machine learning model that solves the problem at hand. To evaluate the performance of our model, we choose the appropriate statistical metric. We work on building the model that results in the best accuracy score, the R\_squared score that is closest to 1, or the best Precision and Recall.

If you're working in research, statistical metrics are the most important, and you should focus solely on these. But, if you're like the majority of data scientists, working in a business that relies on data mining to provide better solutions to its customers, then you should go further.

In the end, a "Good" machine learning model is one that performs well on unseen data and helps the business move forward to provide customers with more innovative solutions that are backed by data.

# Two Types of Model Evaluation

There are two types of model evaluation metrics. Each metric addresses a different side of the business problem we are dealing with. Evaluating your model on both metrics ensures that you're doing the right job for the business.

## 

## Intrinsic Evaluation

Intrinsic evaluation metrics care about the model's success in performing the required task. How accurate is the model in classifying spam or ham emails? Or How accurately does it predict the demand for items in our stores? This is what we were trained to focus on, and I think we are comfortable to a great extent in choosing the right intrinsic metric for evaluating our model. But in business, this isn't just enough.

## 

## Extrinsic Evaluation

On the other hand, Extrinsic evaluation focuses on the model's success in performing the final business objective. How does it support the company in delivering better services to the customer? or How much time does it save our users?

Let’s consider a spam detection model as an example. The intrinsic evaluation metric will be Precision and Recall. While the extrinsic metric that the company cares about will be the amount of time users spend on spam emails. Or the time a user wasted because a genuine email went to their spam folder.

## 

## If the extrinsic model evaluation is what matters, why do intrinsic evaluation at all?

In industrial projects, Machine Learning models are built to provide solutions for business problems or to improve the services that companies provide. This means that the most important metric in the model evaluation process should be the extrinsic metric because it has a direct impact on the business objectives.

**If so, why bother doing an intrinsic evaluation of our model?**

For two reasons. First, extrinsic evaluation is a much more expensive process, as it involves decision-makers and executives outside the AI team. So, it's easier to perform an intrinsic evaluation.

Second, intrinsic evaluation can serve as a proxy for extrinsic evaluation. Bad results in intrinsic evaluation often imply bad extrinsic evaluation results. The opposite is rarely true. Intrinsic evaluation metrics can help us predict the model's performance when evaluated on extrinsic metrics, and enable us to make quick iterations to reach a better model with minimal cost. And only after reaching good results in the intrinsic evaluation should we pursue the extrinsic evaluation.