

The Role of Data

Data is at the core of machine learning. It is central to the current evolution and further advancement of this field. Just as it is for humans, it is the same way for machines.

Learning is not possible without data.

Humans learn how to perform tasks by collecting information from the Environment. This information is the data the brain uses to construct patterns and gain an understanding of the Environment. For a human being, data is captured through the sense organs. For example, the eyes capture visual data, the ears capture auditory data, the skin receives tactile data, while the nose and tongue detect olfactory and taste data, respectively.

As with humans, this same process of learning from data is replicated with machines. Let's take, for example, the task of identifying spam emails. In this example, the computer is provided email examples as data. It then uses an algorithm to learn to distinguish spam emails from regular emails.

The Cost of Data

Data is expensive to collect, and high-quality data is even more costly to capture due to the associated costs in storing and cleaning the data. Over the years, the paucity of data had limited the performance of machine learning methods. However, in the early 1990s, the Internet was born, and by the dawn of the century, it became a super highway for data distribution. As a result, large and diverse data became readily available for the research and development of machine learning products across various domains.

In this chapter, we covered the definition and history of machine learning and the importance of data. Next, we will take it further by discussing the principles of machine learning in [Chapter 14](#).