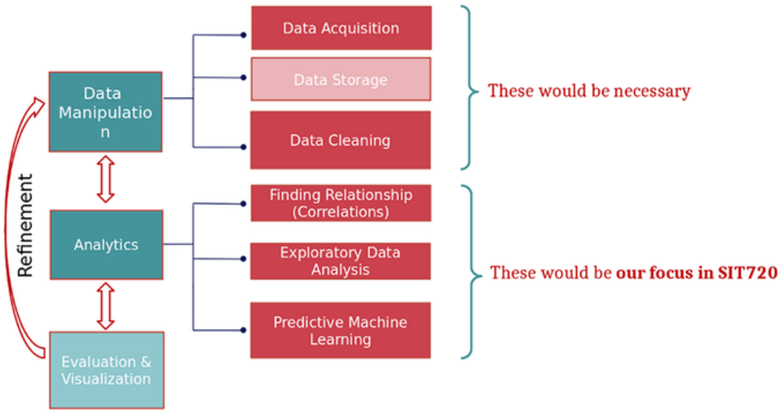

© Getty Images

Machine Learning steps

Have you ever thought about how you learn to make decisions?

Let's assume that we have been asked to research a problem and make a decision. You must have access to information (data) from which you can learn. Once you have data, what do you do with it? Most people analyse the information to find patterns and relationships. You make a decision based on your work. Finally, you evaluate the model. Did you make a good decision or could it have been better?

Let's look at each of these stages in more detail and relate them to ML.

Step 1: Data Manipulation

This is a process of data preparation. ML usually uses the largest sets of data available.

The first step in data manipulation is *Data Acquisition*. Data acquisition is the process of sampling information that illustrates real world physical conditions with a predefined measurement. Using our car example, you might measure engine size, number of doors, size of tyres etc.

The data acquired should be reliable for converting into digital numeric values that can be manipulated by a computer. Number of car doors is easy; style is less easy to define numerically.

Once the data is properly stored, any redundant, noisy, unusable parts of it should be trimmed. We call this procedure *Data Cleaning*. Data cleaning is a major step as real-world datasets are highly affected by noise, redundancy and missing values. We might delete any three-wheeled cars because they're so unusual.

Step 2: Analytics

The second main step in machine learning is *Analytics*. Analytics mainly involves *finding relationships and correlations* in the prepared data in order to design an accurate model based on that input data.

In addition, *Exploratory Data Analysis* is an approach for analysing datasets in order to summarise their main characteristics or features. Many exploratory data analysis methods use visual illustration of data, based on different features. Things like graphs, charts and tables make data easier to understand.

Finally *Predictive Machine Learning* is the last stage of Analytics. It uses a variety of statistical techniques such as predictive modelling in order to build a classifier or intelligent system for decision making. We will come back to these ideas over the next few weeks.

Step 3: Evaluation and Visualisation

The final result of Analytics is an intelligent system or model. As the last step of ML design, we have to evaluate the performance of the system. "Did I choose the right car?"

If the quality and performance of the intelligent system does not achieve a satisfactory outcome, the *Refinement* procedure is required and another round of data manipulation and analytics becomes necessary. Again, we will come back to these ideas as you move through the Unit.

In this Unit we mainly focus on the *Analytics* step. However it will be necessary to use parts of the *Data Manipulation* step to build a model.

Activity

If you would like another explanation of the scope and complexity possible and the methods used to make the machine learn, watch this 10 minute video from Android Authority.

It goes over some of the concepts we've covered so far, and introduces others.

Tip: if you watch it at twice the speed, it takes half as long!

This is an additional video, hosted on YouTube.

Does this change your own ideas about ML? Student Discussion your thoughts.
