**how to handle a large dataset**

In real life, all the data we collect are in large amounts. To understand this data, we need a process. Manually, it is not possible to process them. Here’s when the concept of feature extraction comes in...

**What is Feature Extraction?**

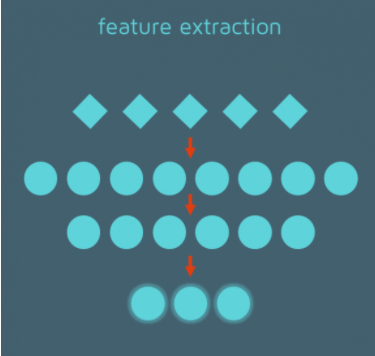
Feature extraction is a part of the dimensionality reduction process, in which, an initial set of the raw data is divided and reduced to more manageable groups. The most important characteristic of these large data sets is that they have a large number of variables. So Feature extraction helps to get the best feature from those big data sets by select and combine variables into features, thus, effectively reducing the amount of data

In short,, the reduction of the data helps to build the model with less machine’s efforts and also increase the speed of learning and generalization steps in the machine learning process.

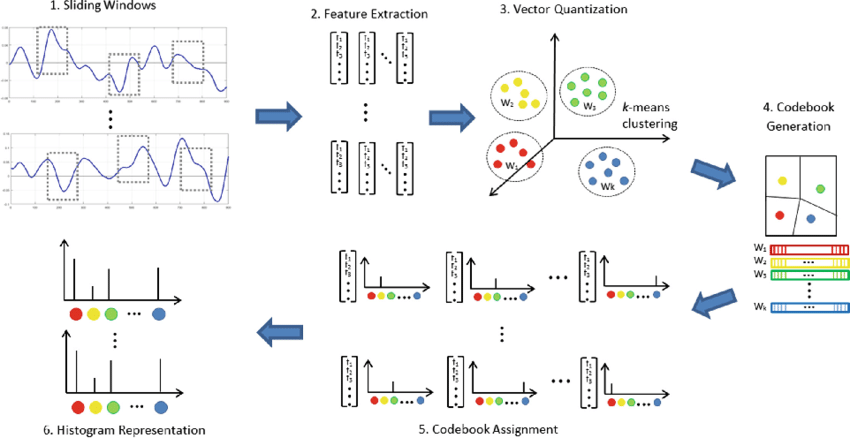
nguồn : <https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/>



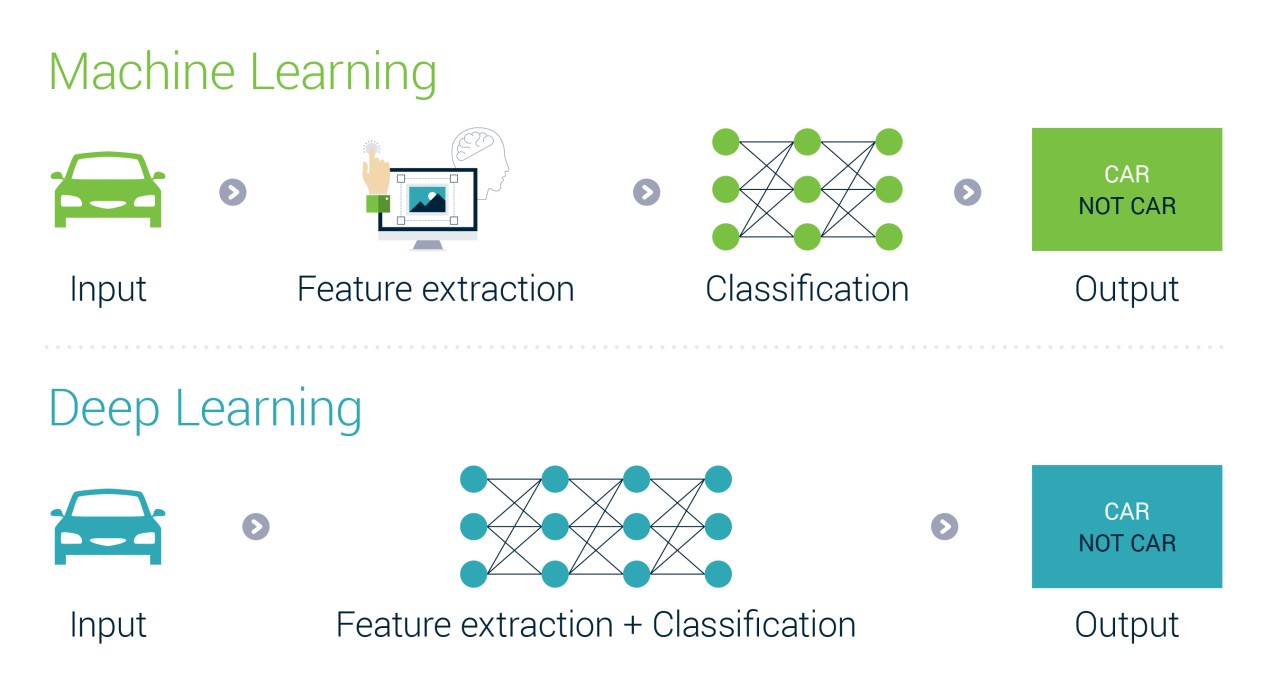
How Feature Extraction works



**Bag of Words-**Bag of Words is themost used technique for natural language processing. In this process they extract the words or the features from a sentence, document, website, etc. and then they classify them into the frequency of use. So in this whole process feature extraction is one of the most important parts



**Image Processing** –Image processing is one of the best and most interesting domain. In this domain basically you will start playing with your images in order to understand them. So here we use many many techniques which includes feature extraction as well and algorithms to detect features such as shaped, edges, or motion in a digital image or video to process them.



**Auto-encoders:**The main  purpose of the [auto-encoders](https://www.mygreatlearning.com/blog/autoencoder/) is efficient data coding which is unsupervised in nature. this process comes under unsupervised learning . So Feature extraction procedure is applicable here to identify the key features from the data to code by learning from the coding of the original data set to derive new ones.

