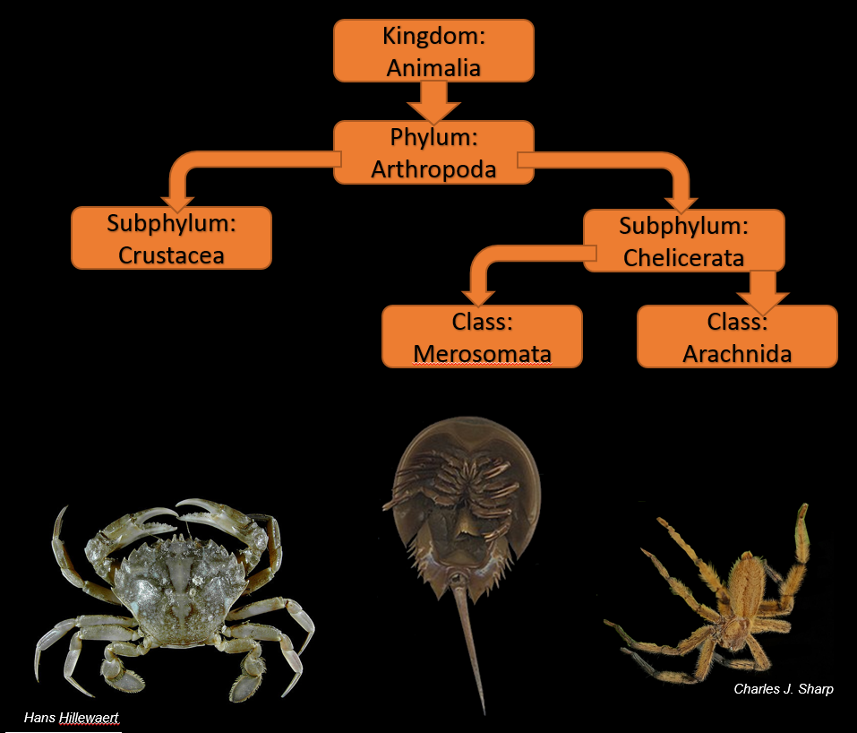
Classification is one of the key components of the modern AI toolkit in which a machine learning (ML) algorithm attempts to mimic the human ability to distinguish and categorize. The idea is that the algorithm, when confronted with a new instance of an object, is to statistically determine the class or category into which this object best fits.

Classification is one of those human activities that is deceptively simple. For example, for decades people thought that the [horseshoe crab](https://en.wikipedia.org/wiki/Horseshoe_crab) was related to the [crustaceans](https://en.wikipedia.org/wiki/Crustacean) because it could be found in the ocean. As biology progressed, it became clearer that the horseshoe crab wasn’t in the same class as the crustaceans and that it has more in common with [arachnids](https://en.wikipedia.org/wiki/Arachnid), making it the ‘spider of the seas’.



The ability of an expert to determine whether an object belongs in one category is also a subtle affair that is often as much art as it is science, as the following excerpt from Miss Marple’s speech in *A Christmas Tragedy* by Agatha Christie nicely describes:

“It’s really a matter of practice and experience. An Egyptologist, so I’ve heard, if you show him one of those curious little beetles, can tell you by the look and feel of the thing what date B.C. it is, or if it’s a Birmingham imitation. And he can’t always give a definite rule for doing so. He just *knows*. His life has been spent handling such things.”

Christie makes several important points in that brief passage. First, there is the matter of ‘practice and experience’. This translates in the domain of machine learning to training. Second, she speaks of the Egyptologist ‘handling’ the beetle and judging by the ‘look and the feel of the thing’. This requirement corresponds to having a set of percepts about the object, a point that is, arguably, the trickiest. The third point she raises is that the expert can classify the age of the object (‘what date B.C’) or can spot a counterfeit. Of course, this is the point of the ML algorithm in the first place, to be able to judge expertly a new object. The fourth and final point is that the expert can’t always give a definite rule explaining how he judged. There is no direct translation of this rule into the domain of machine learning, but more on that point below.

In thinking about an expert (Egyptologist or otherwise), we need to recognize that what makes him an expert is that he is more often right than wrong. The context for the previous excerpt is the argument that Miss Marple, a spinster sleuth of uncertain age, makes about how ‘superfluous women’ (such as herself) who engage in ‘tittle tattle’ are ‘nine times out of ten’ correct, and ‘[t]hat’s really just what makes people so annoyed about it’. So, we can’t expect our machine learning algorithm to be able to be 100% accurate, since no expert ever is; we can only hope that it is ‘accurate enough’.

It is also very likely that the algorithm will never be as accurate as a human expert for the following reason. In philosophical terms, machine classification overlaps the first and second [Acts of the Mind](http://aristotle2digital.blogwyrm.com/?p=998) (the Act of Understanding and the Act of Judgement) without, necessarily, being fully developed in either.

For humans, the first act involves apprehending the percepts provided (‘handling’ the object to get an idea of its ‘look and feel’). A baby is born with the ability to process his perceptions; to make some sort of comprehension of the sensory input from the five senses. In the second act, the person abstracts universals (or what, at least functionally, passes as such) from those sensory experiences to be able to understand ‘redness’ or ‘roundness’ or the being-qua-being of any other form. These universals allow the human to then classify and sub-classify the objects in the surrounding world.

In contrast, the machine is taught about only a small subset of possible percepts (typically digital data representing an image or a time series). Currently, no machine can expand or contract its attention when it realizes it needs to know more or is being blasted with too much information. In addition, it only knows the categories that are used to train it.

The human has a decided advantage in that he can expand or contract the number of attributes used in the classification on the fly (e.g. concentrating only the weight and the texture first and then adding in color and style as needed later) and the human can invent new attributes at need (e.g., suddenly noticing that the size matters). The machine has only two advantages: raw speed and the ability to handle an arbitrarily large number of attributes (although the number must be fixed for each situation). As a result, the machine’s ability to classify is entirely based on some statistical or probabilistic measure. The human’s ability to classify is surely rooted in probability as well, but what if anything else is going on is, at this time, anybody’s guess.

To be more concrete, consider the problem of spam emails. Determining whether a given email is spam or not is a good example of a classification problem that illustrates some of the advantages and disadvantages on both sides. The human can actually read the content of an email, comprehend the meanings, and judge the context (which may require consideration of different attributes compared to the previous email) before deciding whether the message is good or bad. However, the human can only read a limited number of emails each day and is prone to getting bored or tired and making mistakes. The machine can make sense of large amount of the associated network data, be it IP addresses, message size, number of hops and so on – data that would make little or no sense to an overwhelming number of most humans. In addition, the machine can analyze a vast number of messages in the time it takes the human to read one.

Over the coming months, this column will look at some of the more [popular ML techniques for classifying data](https://analyticsindiamag.com/7-types-classification-algorithms/) and compare the pros and cons of each technique. Some of the metrics for the comparison will be the difficulty of assembling a training set (the data that gives the required ‘practice and experience’), whether the data need to be pre-labeled into classes (e.g., a real scarab or a Birmingham imitation), or whether we can allow the algorithm to find the possible classes based on how the data cluster, the accuracy of the method compared to truth, and the application domains that experts use. In the end, we will have essentially a classification of classification algorithms.