*For each of these scenarios decide if you could use supervised or unsupervised techniques, or both!*

1. Define the likelihood that an individual will contract a specific disease

This would use supervised learning, since we have a target outcome (whether or not the person contracted a disease) that we use in our training of a model which will output probabilities.

1. Translate a set of images into variables for modeling

Probably unsupervised learning since there is no outcome variable to try to predict. An unsupervised model might be able to “say” something about our images in the form of output variables (maybe pixel counts, shades?).

1. An ecommerce company wants to identify power users

Perhaps this situation could lend itself to both supervised and unsupervised learning. You can build a supervised ML model if you know in advance who the power users are (outcome variable) and have other data that you can use as features to predict this outcome. I think you could also use unsupervised learning – via clustering – without explicitly “telling” the unsupervised model which group of users you were after. Perhaps the model would be able to correctly identify this group through similarities in multiple dimensions (variables) that the company tracks.

1. That same company wants to see shopping patterns in users

This seems to be an area where you could again use both supervised and unsupervised learning. In order to use supervised, the company would need to define discrete shopping patterns that they have observed (perhaps a group for customers who buy once and never again, customers who buy on a recurring basis – perhaps every two weeks, customers who buy on a recurring basis but without a discernable pattern, etc.) and use that discrete shopping pattern as an outcome variable in a supervised model. The company could also use unsupervised learning and let the model try to identify its own groups via clustering or perhaps some type of pattern learned through neural networks.

1. You want to reduce the number of variables inputting into your random forest model

In this case, we can use PCA, an unsupervised technique, to reduce the number of input variables while explaining the maximum possible variance in the output.