DS 4.1.3

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

Supervised learning will work in this scenario. We are defining the probability that an individual contracts a disease so the target variable will be binary/categorical. We can use data about each individual to predict this

1. Translate a set of images into variables for modeling

Unsupervised learning will work here. WE don’t have an outcome variable we are trying to predict. The goal is simply to translate images into variables.

1. An ecommerce company wants to identify power users

Either supervised or unsupervised learning will work. Unsupervised learning can take data about users and put them in clusters. Supervised learning will have a target variable of a power user (1) or not a power user (1). This can be predicted based on data for each user.

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

Unsupervised learning will be best here. Clusters can be made for users who share similar shopping patterns.

1. Unsupervised PCA will reduce the number of variables and random forest will be a supervised learning model.