Supervised learning	
Linear regression	This algorithm is used for determining a relationship between an independent variable (predictor) and a dependent variable (response) (Swaminathan, 2018a) when the dependent variable is continuous.
LASSO and ridge regression	LASSO and ridge regression are closely related extensions of linear regression. They work in a similar way to linear regression, but avoid the problem of "overfitting", by discounting outliers and extreme values in the analysis. The main difference between these two extensions is that the regularisation in the LASSO method is an absolute value, which has a crucial impact on the trade-off between underfitting and overfitting a model (Chakon, 2017).
Logistic regression	This algorithm is used for determining a relationship between an independent variable (predictor) and a dependent variable (response) when the dependent variable is categorical (Swaminathan, 2018b). The output of the logistics regression is a value between 0 and 1. As we will see later, this classification property makes it very useful for creating artificial neural networks.
Decision trees	Used predominantly for classification, although not exclusively, this algorithm is based on a successive decision process (Bonaccorso, 2017:155). Starting at the "root", a variable is assessed, and one of two "branches" is selected. This process is repeated until the target variable, in the final "leaf", is identified (Bonaccorso, 2017:155).
Random decision forests	The random decision forest algorithm builds multiple decision trees, which are then combined to create a more accurate model (Donges, 2018). This is referred to as an "ensemble method", as it combines the outputs of many individual models to reach a decision.
Support Vector Machine (SVM)	In this algorithm, each variable is plotted in n-dimensional space, which separates the variables into different classes (Patel, 2017b). Its main application is to classify new data in relation to existing data, for example in fraud detection in financial transaction data, or email spam filters.
Unsupervised learning	
k-means	Using unstructured data, the k-means algorithm clusters data according to features of similarity, with k denoting the number of clusters (Trevino, 2016). The value k is a "hyperparameter", that is, chosen by the data scientist before the model runs.
Naïve Bayesian Classifier	Bayes' Theorem can be used in machine learning too, and it works well for both clustering and classification. When used in clustering, any new datapoint updates the "prior", and thus builds the clusters from the ground up. Its main advantage is that it works well where there is only little data available to make predictions.
Reinforcement learning	
Q-learning	A reinforcement learning method where the values of actions in states are stored in a Q-table to calculate the expected future rewards for each action in each of the possible states (ADL, 2018). The algorithm performs, in effect, a true "trial and error" strategy, and changes its behaviour in relation to past experience.
Policy gradient	Policy gradient algorithms work just like Q learning, but here a hyperparameter is chosen by the data scientist that sets the balance between "exploring" new solutions, versus "exploiting" solutions that were found to work in past rounds.
Deep learning	
Convolutional neural networks (CNN)	Convolutional neural networks are the classic method used for image recognition. The algorithm uses different types of layers that extract and pool features from input images to differentiate different objects in the image (Saha, 2018).
Recurrent neural networks (RNN)	A form of neural network that includes loops or connections between the different layers in the network, and neurons in the same layer. "A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor" (Banerjee, 2018). The most efficient RNN is called LSTM, which stands for "Long Short-Term Memory", and is an RNN that in effect has a memory function that can store the last value. This is very helpful when working on sequential data like language or speech.

Encoders and transformers	The most recent developments of neural networks in natural language processing are so-called encoder-decoder models, and transformer models. They combine several neural networks within an architecture that allows for the inputs and outputs to have different formats. This is very helpful when analysing or generating language but can also be used for image processing. **ChatGPT**, **Claude**, and **Gemini** are very powerful examples of modern transformer models for language generation and analysis, which combined are often referred as “LLMs – large language models”.