**Machine Learning Summaries**

**Linear Regression**

Linear Regression is a form of supervised learning. It is used to predict a dependant variable (Y) based on a given independent variable (X), finding out the linear relationship between input (X) and output (Y). A line of best fit is used on a graph to highlight this relationship. A practical use of linear regression would be as a tool to predict the effect of depreciation on a car. In this example, X would be the price of the car after years 1 and 2, with Y being the price in the future.

**Logistic Regression**

Logistic Regressions is a form of supervised learning. It is similar to Linear Regression in that it uses observations to make predictions, however where linear regression outputs continuous number values, Logistic Regression gives us a probability value which can then be mapped to two or more discrete classes. Logistic Regression could be used to show the probability of a shopper buying coffee or tea depending on what time of the day it is.

**Decision Tree**

A Decision Tree is a form of supervised learning. It makes decisions based on the conditions present in the data. They are most effective when producing categorical outcomes (True/False) but can also be used as a regression tree to give outputs that are continuous. Decision Trees may be used by banks to decide whether or not to grant someone a loan. It would use independent variables such as salary, age and outgoings in its decision making.

**Support Vector Machine (SVM)**

A Support Vector Machine is a form of supervised learning. It uses classification algorithms to classify data into two distinct categories based on the training data it receives by using a hyperplane to differentiate the data by the largest margin. It can categorize data into more than two categories, if the differences are non-linear, by using Kernels. A classic example of an SVM would be as a tool to classify email as spam.

**Naive Bayes**

Naive Bayes is a form of supervised learning. It assumes that all predictors are independent, which is to say that the presence of one feature in a class doesn’t affect the presence of another one. Naive Bayes is named after Bayes’ Theorem which it applies to find the probability of outcomes occurring and make predictions based on that probability. It is extremely fast and scalable, making it ideal for large data sets. Naive Bayes is often used as a text classifier, a spam email filter being a good example of this.

**K Nearest Neighbours (KNN)**

KNN is a form of supervised learning that can be used for both regression and classification. It tries to predict the correct class for the test data by calculating the distance between the test data and the training data. It calculates the probability of the test data belonging to each different class of ‘K’ data, the class holding the largest probability being the one that will be selected. When calculating regression the algorithm uses the mean of the ‘K’ selected training points. KNN may be used by streaming platforms to recommend content for people based on the shows they’ve already watched.

**K-Means**

K-Means is a form of unsupervised learning. It divides objects into clusters that share similarities and are dissimilar to objects from other clusters. The amount of clusters can be decided by the user, with the algorithm then taking over to divide the objects into the amount of clusters specified. It is used in search engines.

**Random Forest**

Random Forest is a form of supervised learning. It uses multiple decision trees to output the class which is the mode (classification) or mean (regression) prediction of the individual trees. While decision trees consider all the possible feature splits, random forests select only a subset of those features. These characteristics make random forests more precise predictors than regular decision trees as they account for the potential variability of the data. Random Forests are used in finance as a precise way of detecting fraud and evaluating individuals with high credit risks.