**HLT – Common Machine Learning Algorithm**

**Linear regression** **algorithm** is a widely used form of predictive analysis. Regression aims at examining the relationship between two variables. It does this by applying a linear equation to the observed data. The variables are in 2 forms, independent and dependent. It is mostly used to forecast and determine cause and effect relationships among variables. Linear regression is supervised learning and the intended goal of the regression algorithm is to plot a best fit line between the data. The benefits of linear regression are that it is widely used, runs fast, is highly interpretable and the base for other methods. A good example of linear regression can be in the real estate sector where potential buyers look at factors such as the cost of properties against factors such as transport links, crime, number of bedrooms/size and local amenities before make a house purchase. You can fit a linear regression model using by taking 2 factors such as the number of bedrooms to determine the predict the house prices. Another example can be in the medical field where scientists are able to do a study showing a direct correlation between administering various dosages of a drug to patients and observe how it reacts to their cancer recovery. A linear regression model using dosage as the predictor variable and cancer recovery as the responsive variable.

**KNN (K nearest neighbours)** – This is an algorithm based on supervised learning. It is used mostly for classification and regression. K nearest neighbour assumes the similarity between new data and already available data and puts the new data in a group with the most similarities working on the analogy that birds of a feather stick together. An example is if there is an image of an animal such a cat or dog. To clearly identify what it is, The KNN algorithm can be used to identify the animal based on its similarities. The model will find similar features and place them in either a cat of dog category. Another example where KNN is used is in the Banks networks whereby KNN can be used to predict whether a customer is likely to have their loan application approved based on similar traits to a previous defaulter.

**K-means algorithm** is an unsupervised learning algorithm, which groups the unlabelled datasets into various clusters. It allows us to cluster the data into specific groups and provides a simple technique to determine the categories of groups in an unlabelled dataset without any training. K-means performs the division of items into clusters that share similarities and are dissimilar to the items belonging to another cluster. An example is when you go to the supermarket and similar items are arranged in particular sections. For example, carrots, cucumbers and courgettes are all arranged in the vegetable section. Carrots are placed in one place and cucumbers are placed in another place in the vegetable section. They will apply to staple foods like types of pasta and rice will also arranged separately according to the characteristics.