**Curneu MedTech Innovations Assessment -Task 1**

**K-nearest neighbors classifier**

**Introduction:**

The K-nearest neighbors (KNN) algorithm is a type of supervised machine learning algorithms.  is extremely easy to implement in its most basic form, and yet performs quite complex classification tasks. It is a lazy learning algorithm since it doesn't have a specialized training phase. Rather, it uses all of the data for training while classifying a new data point or instance. KNN is a non-parametric learning algorithm, which means that it doesn't assume anything about the underlying data. This is an extremely useful feature since most of the real world data doesn't really follow any theoretical assumption e.g. linear-separability, uniform distribution, etc.

**Problem Statement:**

A dataset is given with labels based on fruits height, width, mass, colour score in a excel file. The task is to develop a KNN classifier machine learning model from scratch by analysing various combinations of parameters using scatter plot to find the best suited parameter combination to build the classifier model and to find out the best value of k with highest r2 score. Also to run three test cases on the parameters and to access the fruit using classifier.

**Approach to the problem:**

Given dataset is read using pandas into a variable named dat. The description of the dataset and the first 5 rows of the dataset are printed. Then scatter plots are made for different combinations of features such as fruit height, width, mass and colour score. The correlation coefficient for every combination is also found to find the best possible combination of parameter.