kNN stands for k-Nearest-Neighbors in Machine Learning. This is a type of algorithm used in Machine Learning that can be used for regression and classification. The kNN algorithm will predict a new input's class. This is determined by utilizing the classes of the k-Nearest-Neighbors in the data. Similarly, when used in regression, the kNN algorithm finds the value by predicting using the k-Nearest-Neighbors and finding their mean.

Another algorithm utilized in Machine Learning is Decision Trees. This algorithm is also used for classification and regression. Decision Trees will split data into subsets in a recursive fashion. This is based on specific features of the data. The recursive manner of the algorithm begins at the root node of the tree and will result in leaf nodes. The leaf nodes correspond to the prediction, given as a value or class.

There are three clustering methods used in Step 3 of this assignment: k-means clustering, Hierarchical clustering, and Density-based clustering. K-means clustering is a type of clustering that will separate or partition the given data into different clusters (k number of them.) This method of clustering will also use randomization to give initial values to the different clusters' centroids. Additionally, it will utilize iteration and change the centroids to equal the average of all its points. If there are too many iterations or there is no change to the centroids, the algorithm will terminate. Hierarchical clustering is based on a hierarchy of different clusters. The structure of this Hierarchy is called a Dendrogram. If the points are all in the same cluster, the algorithm will terminate. Density-based clustering is a different type of clustering that relies heavily on the density of the data. This algorithm will assume that the clusters are located where the most-dense data is. In finding the density, the algorithm will utilize a metric for distance to determine the areas of dense data.

PCA is Principal Component Analysis. This is a technique for analyzing data in machine learning. It helps, in a machine learning context, to diminish the dimensionality of given data. On the other hand, LDA stands for Linear Discriminant Analysis. This method also reduces dimensionality, but LDA is considered a “supervised” type of analysis while PCA is considered “unsupervised.” Both methods can help to diminish the complexity of given data. This, in turn, might result in an increase in speed for computing.