Labelling the Clusters

The labelling of clusters is a procedure implemented explicitly in the attached code of the Assignment and utilized in both K-Means and KNN algorithms. However, the methodology that led to the labelling of the clusters will be explained thoroughly in this file. Before describing our way of thinking, it should be noted that the stated topology was taken into consideration for the determination of the labels. Furthermore, the proposed formulation defines the clusters efficiently and the results are always the same for the implemented algorithm

Generator Outage

To define in which of these clusters there is a generator outage, we assumed that when a generator is out of the system, the power flow through the line that connects each generator bus with the adjacent one will be negligible and therefore the voltage drop between the two buses will be very small, if not zero. That means that the generator and the bus that is connected through the line will have almost similar voltages for all the states that belong in that cluster. Therefore, the proposed procedure involved the calculation of the voltage drop between the generator buses and their adjacent lines for all the given states. We ended up in the mean value of each one of these voltage drops that correspond to each defined cluster. The cluster that has the minimum value, in this case **Cluster 3** will be assigned as Generator Outage as stated below where the voltage drop is calculated close to zero.

Cluster ID	Voltage Drop (Generator Bus - Bus)
0	0.05150862909181784
1	0.06861039250294303
2	0.05001075859564901
3	3.850874395663636E-4

Line Outage

The idea behind the Line Outage determination cluster is similar with the Generator Outage. Basically, we thought that when a line between two buses will be disconnected, then these two previously connected buses will have maximum voltage drop compared to any other case. That can be explained from the fact that the current, as the line is disconnected, must find an alternative path to reach the specified bus. As a consequence, the increased resistance that the current will meet until it reaches the specified bus will result in an increased voltage drop compared to the case when the line was healthy. So, in this case, we calculated the mean voltage drop in each line, for all the states and we specified the maximum occurred voltage drop. In this case, **Cluster 2** is labelled as Line Outage.

Cluster ID	Voltage Drop (Bus - Bus)
0	0.19457756278988916
1	0.2714591392401274
2	0.42082164239896364
3	0.10233977087926809

Load Rate

The load rate determination is based on the fact that when the load rate is high in a power system, then more current will flow through the lines in order to meet the increased demand. That can be translated to greater power losses in the lines which means increased voltage drops and lower voltages in the load buses of the system. On the contrary, when the load demand is low, then the amount of power losses is sufficiently lower which means that the voltage drop in the lines is lower and the voltage in the load buses higher. Therefore, for the determination of the load rate clusters, we took into account only the voltage in each bus. For each cluster, we calculated the mean value of voltage in each bus and subsequently the mean value of voltage in every bus of the system regardless of whether they are load or generator buses. The maximum mean value of voltage in each bus is translated with the Low Load Rate label, whereas the minimum mean voltage represents the High Load Rate label. Below, our procedure gives that **Cluster 0** is labelled as Low Load Rate and **Cluster 1** as High Load Rate.

Cluster ID	Mean Bus Voltage
0	1.000334882620507
1	0.9280084469620307
2	0.9809656544663281
3	0.9896411671084692