Rapid Miner Results Holding Document

# Clustering

As part of the rule based approach taken in previous work, it is necessary to determine what state of operation a compressor is in, i.e. loaded, unloaded, idle and what % the VSD is operating at.

In order to more accurately determine this information than by simple observations, clustering was used. A k-means clustering method was used to attempt to group the variable speed compressor’s power consumption into distinct groups. It was known beforehand from observing the compressor’s behaviour in tandem with the power consumption that the different groupings would be:

1. Compressor idle (not spinning): approx.. 1 kW
2. Compressor running unloaded: approx.. 8 kW
3. Compressor running loaded minimal flow: approx.. 20 kW

In addition, the CAGI data sheet for the compressor gave the following information:

1. Compressor running loaded VSD 0-20%: 27.2 kW
2. Compressor running loaded VSD 20-40%: 31.2 kW
3. Compressor running loaded VSD 40-60%: 38.3 kW
4. Compressor running loaded VSD 60-80%: 42.6 kW
5. Compressor running loaded VSD 80-100%: 52 kW

## 2-Means Clustering

K was initially set to 2, to differentiate between loaded and unloaded operation. Initially the only input to the clustering algorithm was the power consumption of the compressor. **Figure 1** shows a simplified version of the process in RapidMiner.

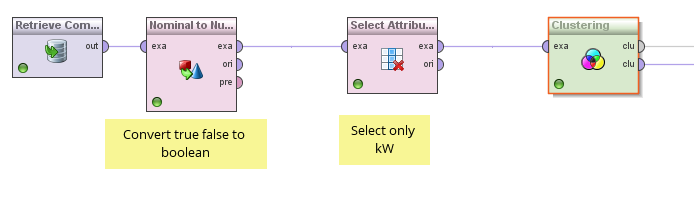


Figure 1: Simple Clustering Process

**Table 1** describes the two clusters determined using this method:

|  |  |  |
| --- | --- | --- |
| Cluster | Number of Items | kW Centroid |
| 0 | 563 | 7.20 |
| 1 | 194 | 21.27 |

Table 1: Simple Clustering Results

**Figure 2** shows that the clusters determined using this method have a definite centre, with reasonably uniform dispersal around this centroid.

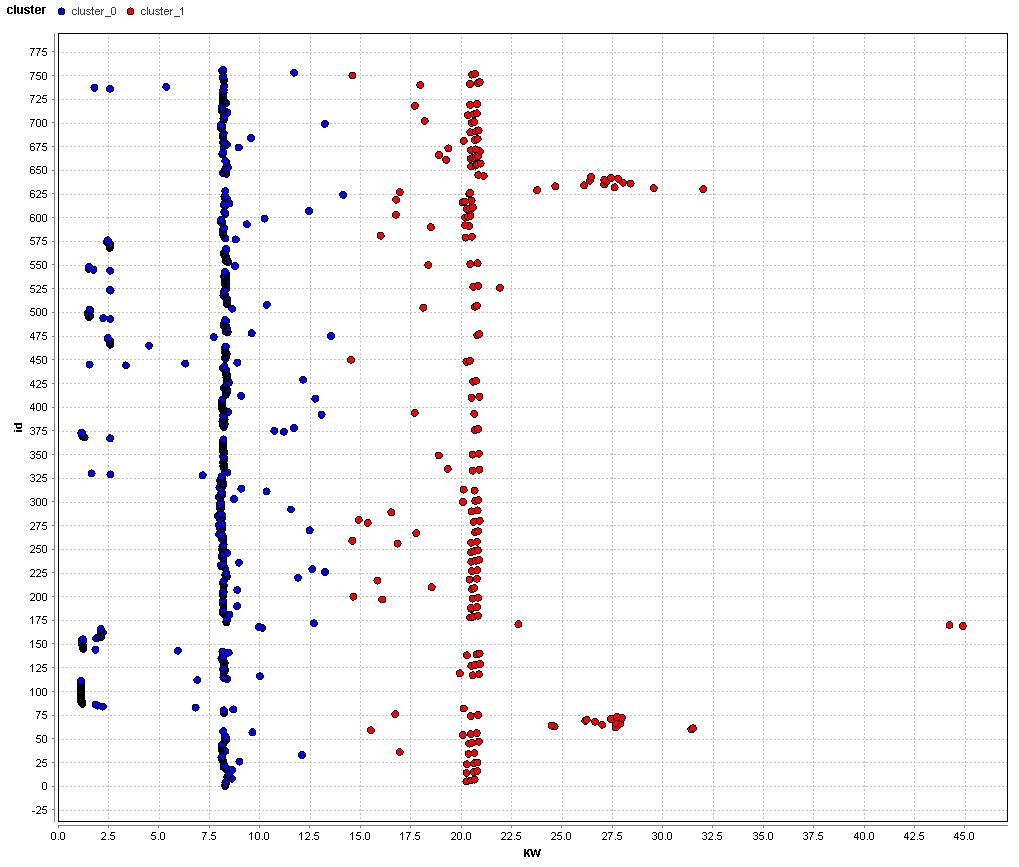


Figure 2: Simple Clustering Scatter Chart

Clearly this method effectively finds suitable centroids for loaded and unloaded operation, based on the initial observations already made from visual analysis of the data.

## 3-Means Clustering

K was increased to 3 to attempt to find more discrete centroids for the compressor power data.

|  |  |  |
| --- | --- | --- |
| Cluster | Number of Items | kW Centroid |
| 0 | 165 | 19.89 |
| 1 | 562 | 7.19 |
| 2 | 30 | 28.60 |

Table 2: 3-Means Clustering Parameters

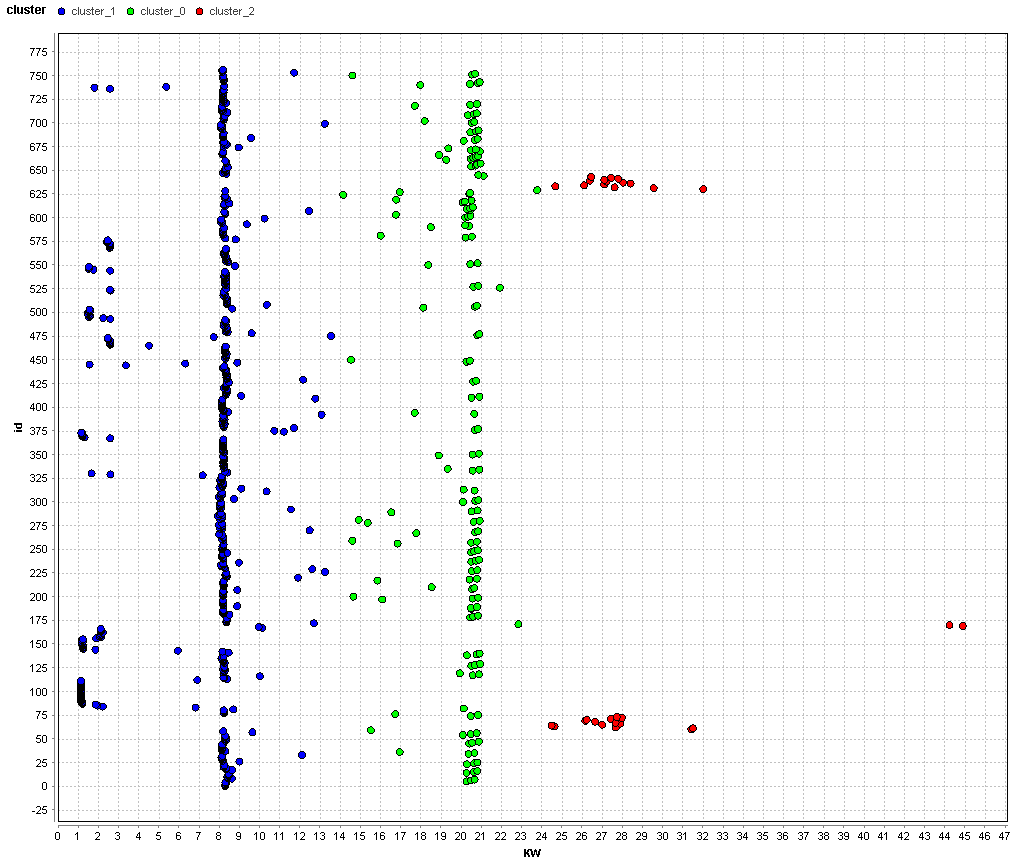


Figure 3: 3-Means Clustering Scatter Chart

Increasing K to 3 has the effect of effectively determining clusters for the following conditions, as comparted to visually observing the data:

1. Unloaded
2. Minimally Loaded
3. Greater than minimally loaded

## 6-Means Clustering

K was increased to 6 in order to try to determine centroids for the following conditions of operation:

1. Idle
2. Unloaded
3. Minimally Loaded
4. VSD 0-20%
5. VSD 20-40%
6. VSD 40-60%
7. VSD 60-80%

There were no observed instances of the compressor operating at VSD 80-100%, and so this category was omitted.

|  |  |  |
| --- | --- | --- |
| Cluster | Number of Items | kW Centroid |
| 0 | 441 | 8.20 |
| 1 | 26 | 16.59 |
| 2 | 100 | 1.72 |
| 3 | 141 | 20.64 |
| 4 | 27 | 29.054 |
| 5 | 22 | 11.97 |

Table 3: 6-Means Cluster Parameters

It is clear from **Table 3** and **Figure 4** that 6-means clustering does not provide satisfactory results for categorisation of compressor operation. Of the centroids obtained, only the following clusters are meaningful:

1. Cluster 0: Unloaded
2. Cluster 2: Idle
3. Cluster 3: Minimally Loaded
4. Cluster 4: VSD 0-20%

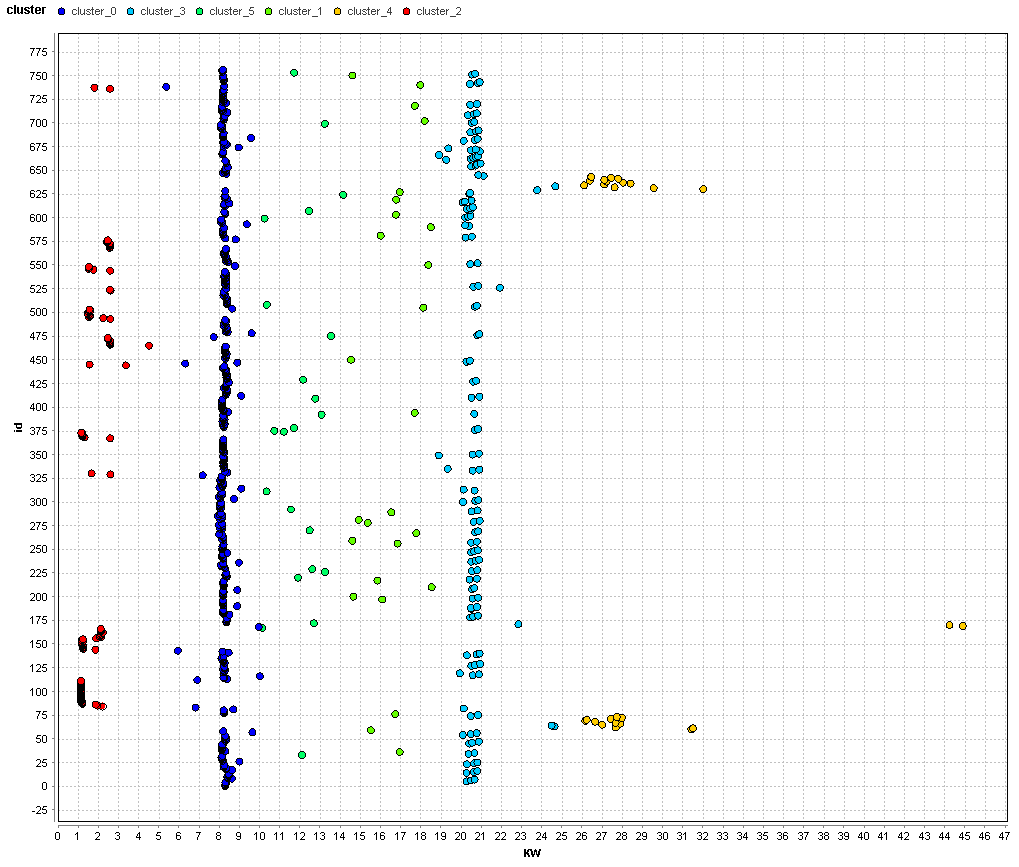


Figure 4: 6-Means Cluster Scatter Chart

## 4-Means Clustering

Having seen that four useful clusters were obtained when trialling 6-means clustering, K was reduced to four to attempt to improve on the four centroids determined using 6-means clustering.

|  |  |  |
| --- | --- | --- |
| Cluster | Number of Items | kW Centroid |
| 0 | 463 | 8.382 |
| 1 | 164 | 19.93 |
| 2 | 100 | 1.72 |
| 3 | 30 | 28.6 |

Table 4: 4-Means Cluster Parameters

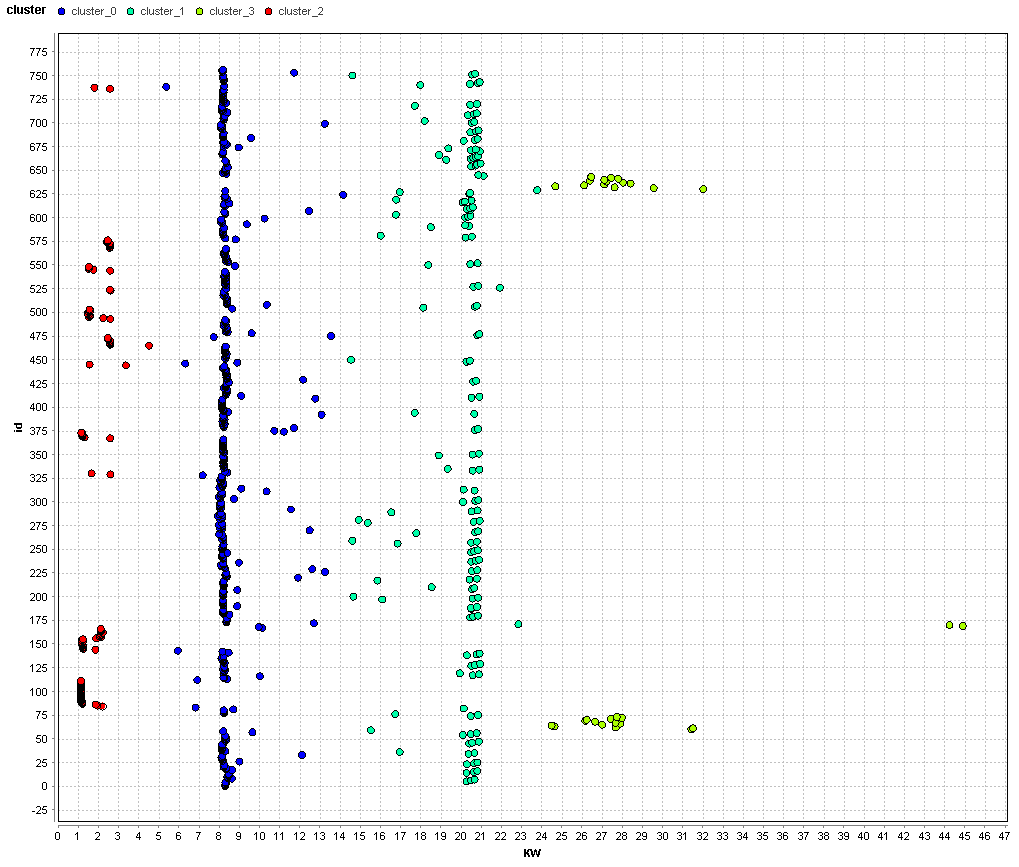


Figure 5: 4-Means Clustering Scatter Chart

Changing K to 4 created well defined centroids for the following categories of compressor operation:

1. Cluster 0: Unloaded
2. Cluster 1: Minimally Loaded
3. Cluster 2: Idle
4. Cluster 3: VSD 0-20%