**Network Rail II R&D Workstream**

**POE Prediction model -**

**Physical attribute analysis**

**Solution document**

# **Context and overview**

The objective of this exercise was to test the performance of separate model on assets with similar physical attributes and compare the performance with the original model where all the physical attributes were considered together.

## **Problem statement**

The purpose of this exercise was to group Clamp Lock Mark 2 assets based on the physical attributes. The hypothesis was that if assets are grouped according to their physical attributes, then the traces for the assets in a group would look very similar and model performance will improve. Finally, when assets have been grouped then separate models can be trained for these groups to check the hypothesis.

## **Scope of work**

The assets in scope for this analysis were as follows:

* The analysis has been carried out on 1438 POE Clamp Lock Mark 2 assets across four routes: ANGLIA, EAST MIDS, KENT and LNW (considered from National Database).
* Scope for data coverage are as follows:

- Ellipse Data : January 2017 - January 2020

- FMS Data : January 2017 - January 2020

- Wonderware Data : January 2017 - January 2020 (No Live Data)

The actions in scope for this exercise were as follows:

* Finding out the right combination of physical attributes to group the assets.
* Training different model for the assets in different groups and check the model performance.

# **The approach**

Two approaches have been finalised to carry out this exercise.

* ***Approach 1: Using clustering algorithm to group the assets***
  + Aims to use clustering algorithm using selected features to group similar looking assets.
  + Based on the assets within a group dominant physical attributes will have to be identified to understand the relation between shape of the trace and physical attributes
* ***Approach 2: Using physical attribute data to group the assets***
  + The objective was to group the assets based on physical attributes
  + After forming the groups, sample traces were plotted for each asset to check the similarity of trace pattern within the groups
  + After forming the groups and checking the traces of the respective groups, separate models were trained and evaluated for these groups.

## **Approach 1: Using clustering algorithm to group the assets**

### **Background**

The objective was to find out the similar looking assets in terms of trace pattern and keep them in a separate group and then to correlate it with the physical attributes. The hypothesis here was that if the traces within a group look similar, then the physical attributes will also be very similar.

To carry out this work, first features were created at trace level for all the assets. Next, features which were extracted from traces were selected and aggregated at asset level. Then various clustering algorithms were tried and finally K-Means clustering algorithm was selected to group the assets.

Below numeric features were extracted and K-Means Clustering algorithm was used to cluster assets based on extracted following features Ph1\_max\_current, Ph1\_min\_current, Ph2\_max\_current, Ph3\_max\_current and Swingtime.

However, after applying K-Means clustering algorithm on 1435 Clamp Lock Mark 2 assets from four different routes, total six clusters were obtained in which the assets were divided and assets with similar trace pattern belongs to same cluster.

Table 1 shows the route-wise number of assets and total assets in the clusters after applied K-Means clustering. In the below table, clusters have been arranged based on the number of total assets in that cluster (descending order).

**Table 1 Group of assets by routes after K-Means Clustering**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number of assets in various clusters (from Multiple Routes)** | | | | | |
| **Clusters** | **Route** | | | | **Total Assets** |
| **ANGLIA\_1** | **LNW** | **EAST MIDS** | **KENT** |
| 1 | 109 | 337 | 62 | 11 | 519 |
| 4 | 130 | 202 | 55 | 12 | 399 |
| 5 | 131 | 70 | 41 | 40 | 282 |
| 2 | 16 | 81 | 24 | 14 | 135 |
| 3 | 56 | 19 | - | 9 | 84 |
| 0 | 12 | 3 | 1 | - | 16 |
| **Total Assets** | **454** | **712** | **183** | **86** | **1435** |

**Note**: There is no relationship between number of features and cluster number in Clustering Algorithm. However, number of clusters is an input for the clustering algorithm based on which it divides the data into those number of groups/clusters. We have set these parameters as six based on the multiple iterations and visualisation. This is the reason we are getting 6 clusters (0,1,2,3,4 & 5).

For visualisation, few sample assets were plotted from individual clusters. For example, figure 1(a) to 1(f) plots show few traces from different Clamp Lock Mark 2 assets on trace direction (NR & RN) where green colour traces represent RN trace and blue traces represents NR trace. The Y- axis of the plot represents the current in amperes and X- axis represents the swing time.

Figure 1(a) to 1(f) shows the visualisation of few sample assets from individual clusters and their trace pattern within the clusters.

Chart

Description automatically generated

**Fig 1(a): Traces of few sample assets from Cluster 1**

Chart

Description automatically generated

**Fig 1(b): Traces of few sample assets from Cluster 4**

**Graphical user interface

Description automatically generated with low confidence**

**Fig 1(c): Traces of few sample assets from Cluster 5**

Chart

Description automatically generated

**Fig 1(d): Traces of few sample assets from Cluster 2**

**Chart

Description automatically generated**

**Fig 1(e): Traces of few sample assets from Cluster 3**

**Graphical user interface

Description automatically generated**

**Fig 1(f): Traces of few sample assets from Cluster 0**

### **Findings**

Total six clusters have been identified after applying the K-Means clustering algorithm and this has been observed that -

1. The trace pattern looks similar for assets within the clusters and there is difference in trace pattern for assets between the clusters.
2. Within each cluster there is no difference based on physical attributes or there is no dominant physical attribute that decides the shape of the trace (Check Appendix 5.1)

## **Approach 2: Using physical attribute data to group the assets**

### **Background**

Another approach has been considered where grouping of the assets was based on their physical attributes and the objective was to first group the assets based on physical attributes and then find out if similar physical attributes exhibits similar trace pattern.

For this task there was a lot of input received from NR SMEs to find out the appropriate physical attributes to group the assets together.

For this analysis, assets of specific physical attributes were selected. Following physical attributes were considered to select the assets: - Full Depth, C Switch and S&C-Turnout. However, for simplicity, let's call these combinations of physical attributes as Group 1. A detail table of physical attribute has been kept in appendix (5.2) for reference.

**Table 2 Assets with physical attributes equal to Full Depth, C Switch and S&C-Turnout**

|  |  |  |  |
| --- | --- | --- | --- |
| **SWITCH\_DEPTH** | **SWITCH\_SIZE** | **SNC\_TYPE** | **TOTAL\_ASSETS** |
| Full Depth | C Switch | S&C - Turnout | 454 |

Further, as per suggestion, more physical attributes were considered on Group 1 assets to find evidence whether more granular level combination of physical attributes exhibits same pattern of traces.

Table 3 shows the group with different physical attributes (using more granular level of physical attributes) and total number of assets in the respective groups.

**Table 3 Total number of assets in each group of assets (SET\_NAME) with more granular level combination of physical attributes**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **SWITCH\_DEPTH** | **SWITCH\_SIZE** | **SNC\_TYPE** | **ROLLERS\_LEFT** | **ROLLERS\_RIGHT** | **SUPP\_DRIVE** | **TOTAL\_ASSETS** | **SET\_NAME** |
| Full Depth | C Switch | S&C - Turnout | No Rollers | No Rollers | 1 | 212 | Set 1 |
| Full Depth | C Switch | S&C - Turnout | Schwihag | Schwihag | 1 | 108 | Set 2 |
| Full Depth | C Switch | S&C - Turnout | No Rollers | No Rollers | blank | 84 | Set 3 |
| Full Depth | C Switch | S&C - Turnout | Schwihag | Schwihag | blank | 24 | Set 4 |
| Full Depth | C Switch | S&C - Turnout | Schwihag | No Rollers | 1 | 5 | Set 5 |
| Full Depth | C Switch | S&C - Turnout | No Rollers | Schwihag | 1 | 4 | Set 6 |
| Full Depth | C Switch | S&C - Turnout | Schwihag | Schwihag | 2 | 4 | Set 7 |
| Full Depth | C Switch | S&C - Turnout | Vortok | Vortok | 1 | 4 | Set 8 |
| Full Depth | C Switch | S&C - Turnout | blank | blank | 1 | 2 | Set 9 |
| Full Depth | C Switch | S&C - Turnout | blank | No Rollers | 1 | 1 | Set 10 |
| Full Depth | C Switch | S&C - Turnout | No Rollers | No Rollers | 2 | 1 | Set 11 |
| Full Depth | C Switch | S&C - Turnout | No Rollers | Other | 1 | 1 | Set 12 |
| Full Depth | C Switch | S&C - Turnout | No Rollers | Schwihag | blank | 1 | Set 13 |
| Full Depth | C Switch | S&C - Turnout | Other | Other | 1 | 1 | Set 14 |
| Full Depth | C Switch | S&C - Turnout | Schwihag | No Rollers | blank | 1 | Set 15 |
| Full Depth | C Switch | S&C - Turnout | Vortok | No Rollers | 1 | 1 | Set 16 |
| **Grand Total** | | | | | | **454** |  |

For visualisation, few assets have been selected from different physical attributes groups and from different routes. For example, figure 2(a) to 2(e) plots show few traces from different Clamp Lock Mark 2 assets on trace direction (NR & RN) where green colour traces represent RN trace and blue traces represents NR trace. The Y- axis of the plot represents the current in amperes and X- axis represents the swing time.

**Route: ANGLIA\_1**

Graphical user interface

Description automatically generated with low confidence

**Fig 2(a): Traces of few sample assets from ANGLIA\_1 route**

Chart

Description automatically generated

**Fig 2(b): Traces of few sample assets from ANGLIA\_1 route**

**Route: EAST\_MIDS**

Chart

Description automatically generated with low confidence

**Fig 2(c) Traces of few sample assets from EAST\_MIDS route**

**Route: KENT**

Chart

Description automatically generated with medium confidence

**Fig 2(d) Traces of few sample assets from Kent route**

**Route: LNW**

**Chart

Description automatically generated with medium confidence**

**Fig 2(e) Traces of few sample assets from LNW**

### **Findings**

After exploratory data analysis it was found that the trace pattern varies from one asset to another asset. Also, the variability in trace patters varies in degree.

A thorough inspection of traces within the groups was done then it was found that there is less correlation between physical attributes and the shape of the trace. The shape of the traces could not be explained using the given physical attributes. There are many other assets show different behaviour, however, showing few assets for visualisation purpose.

Finally, to confirm the above findings, it was decided to check the model performance for each of these sets.

### **The results**

To continue with the experiment, the finalised set of features, the trace segmentation logic and the finalised machine learning model (Isolation Forest) was used to identify the traces into anomalous (-1) and normal (1). The output of Isolation Forest model has been used as input to train the finalised classification model (Random Forest). However, the output of classification algorithms produce two classes and confidence score which is nothing but probability score belonging to each output class. Based on the selected probability cut-off criteria a trace predicted with high probability score will be mark as anomalous and low probability score will be mark as normal.

The model was trained on trace level data and then the trace level predictions were aggregated at day level for evaluation.

Here, True Positives (TP) represents the number of successful predictions of unique failures within a 7 day window from a fault date and False Positives (FP) depicts the number of days predicted as anomalous by the model out of all the days outside the 7 day window.

Table 4 shows the model evaluation performance on different set of assets and their performance metrics like True Positive Percentage (TP%) within specified FP% cut-off and Approx. False Triggers/Month/Assets (AFTMA). Model was built on Set 1 to Set 3 assets group and not considered other sets due to a smaller number of assets.

**Table 4 Random Forest result for Set 1 to Set 3**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Random Forest Result** | | | | | | | | | | | |
| **Set** | **Assets Considered** | | **Unique Failures (On Test)** | **Within 10% FP** | | **10%-15% FP** | | **15%- 20% FP** | | **20%-25% FP** | |
| **Train** | **Test** | **TP%** | **AFTMA** | **TP%** | **AFTMA** | **TP%** | **AFTMA** | **TP%** | **AFTMA** |
| Set 1 | 212 | 204 | 97 | 29.90% | 1.73 | 40.21% | 2.51 | 49.48% | 3.26 | 54.64% | 3.96 |
| Set 2 | 105 | 105 | 28 | 17.86% | 1.36 | 21.43% | 2.17 | 25.00% | 2.58 | - | - |
| Set 3 | 80 | 80 | 47 | - | - | 38.30% | 2.64 | 51.06% | 3.31 | 55.32% | 4.18 |

Overall, the results from different group of physical attributes were inconsistent.

# **Conclusion**

The two exercise that were carried out led to the conclusion that there is less correlation between shape of the trace and the physical attributes hence building statistical model on different group of physical attributes will not be worthy and initial hypothesis can be rejected that assets with similar physical attributes exhibits similar trace pattern.

# **Next steps**

Different models based on different clusters can be built.

# **Appendix**

## **Different Physical attributes in same clusters**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PHYSICAL ATTIBUTES** | | | **CLUSTER NAME** | | | | | | **TOTAL ASSETS** |
| **SWITCH\_DEPTH** | **SWITCH\_SIZE** | **SNC\_TYPE** | **1** | **4** | **5** | **2** | **3** | **0** |
| Full Depth | C Switch | S&C - Turnout | 165 | 110 | 111 | 32 | 29 | 6 | 453 |
| Full Depth | C Switch | S&C - Catch / Trap Point | 3 | 2 | 5 |  |  | 1 | 11 |
| Full Depth | C Switch | S&C - Scissors | 2 | 1 | 3 |  | 2 |  | 8 |
| Full Depth | C Switch | S&C - Switch Diamond | 1 | 1 |  |  |  |  | 2 |
| Full Depth | C Switch | S&C - Fixed Diamond Double Slip |  |  |  | 1 |  |  | 1 |
| Full Depth | B Switch | S&C - Turnout | 42 | 26 | 13 | 2 | 8 | 1 | 92 |
| Full Depth | B Switch | S&C - Catch / Trap Point | 15 | 17 | 8 | 2 | 3 |  | 45 |
| Full Depth | B Switch | S&C - Scissors | 8 | 2 |  |  | 4 |  | 14 |
| Full Depth | B Switch | S&C - Switch Diamond Single Slip | 2 |  | 2 |  | 3 |  | 7 |
| Full Depth | B Switch | S&C - Fixed Diamond Single Slip | 2 | 1 |  |  |  |  | 3 |
| Full Depth | B Switch | S&C - Switch Diamond | 1 |  |  |  |  |  | 1 |
| Full Depth | B Switch | S&C - Switch Diamond Double Slip | 1 | 1 | 4 |  | 3 |  | 9 |
| Full Depth | B Switch | blank |  | 4 |  |  |  |  | 4 |
| Full Depth | B Switch | S&C- Tandem Or Three Throw |  | 1 |  |  |  |  | 1 |
| Full Depth | B Switch | S&C - Fixed Diamond Double Slip |  | 2 | 3 | 4 |  |  | 9 |
| Full Depth | D Switch | S&C - Turnout | 65 | 47 | 34 | 8 | 5 | 2 | 161 |
| Full Depth | D Switch | S&C - Switch Diamond | 3 | 1 | 2 |  |  |  | 6 |
| Full Depth | D Switch | S&C - Fixed Diamond Single Slip | 2 |  |  |  |  |  | 2 |
| Full Depth | D Switch | S&C - Fixed Diamond Double Slip |  |  |  | 1 |  |  | 1 |
| Full Depth | E Switch | S&C - Turnout | 31 | 43 | 17 | 7 | 4 | 5 | 107 |
| Full Depth | E Switch | S&C - Switch Diamond |  |  |  |  | 1 |  | 1 |
| Full Depth | A Switch | S&C - Turnout | 5 | 1 | 2 | 3 |  |  | 11 |
| Full Depth | A Switch | S&C - Catch / Trap Point | 2 | 4 |  |  |  |  | 6 |
| Full Depth | A Switch | S&C - Switch Diamond Double Slip | 2 |  |  |  |  |  | 2 |
| Full Depth | A Switch | S&C - Fixed Diamond Double Slip | 2 | 3 |  | 1 |  |  | 6 |
| Full Depth | A Switch | S&C - Fixed Diamond Single Slip | 1 |  |  |  |  |  | 1 |
| Full Depth | blank | S&C - Switch Diamond | 4 | 6 | 7 | 5 | 3 |  | 25 |
| Full Depth | blank | S&C - Switch Diamond Double Slip | 3 |  |  |  | 1 |  | 4 |
| Full Depth | blank | S&C - Switch Diamond Single Slip |  |  |  |  | 1 |  | 1 |
| Full Depth | F Switch | S&C - Turnout | 6 | 6 | 8 | 5 |  |  | 25 |
| Full Depth | H Switch | S&C - Turnout |  |  | 1 |  |  |  | 1 |
| Full Depth | H Switch | S&C - Scissors |  | 1 |  |  |  |  | 1 |
| Full Depth | Tandem A Switch | S&C - Fixed Diamond Single Slip |  | 1 |  |  |  |  | 1 |
| Full Depth | Tandem A Switch | S&C - Fixed Diamond Double Slip |  |  |  | 1 |  |  | 1 |
| Full Depth | G Switch | S&C - Turnout | 1 | 1 | 2 |  |  |  | 4 |
| Full Depth | Tandem B Switch | S&C - Fixed Diamond Single Slip | 1 |  |  |  |  |  | 1 |
| Shallow Depth | C Switch | S&C - Turnout | 34 | 20 | 24 | 13 | 2 |  | 93 |
| Shallow Depth | C Switch | S&C - Catch / Trap Point | 1 | 1 |  | 1 |  |  | 3 |
| Shallow Depth | E Switch | S&C - Turnout | 15 | 10 | 7 | 9 |  |  | 41 |
| Shallow Depth | D Switch | S&C - Turnout | 16 | 5 | 1 | 8 | 1 |  | 31 |
| Shallow Depth | B Switch | S&C - Turnout | 5 | 2 | 2 |  | 1 |  | 10 |
| Shallow Depth | B Switch | S&C - Catch / Trap Point | 1 | 2 |  |  |  |  | 3 |
| Shallow Depth | blank | S&C - Switch Diamond | 4 | 4 |  | 5 |  |  | 13 |
| Shallow Depth | blank | S&C - Switch Diamond Double Slip |  |  | 2 |  |  |  | 2 |
| Shallow Depth | F Switch | S&C - Turnout | 3 | 2 | 2 | 3 |  |  | 10 |
| Shallow Depth | SG Switch | S&C - Turnout |  |  | 4 |  |  |  | 4 |
| blank | blank | blank | 69 | 71 | 16 | 23 | 13 | 1 | 193 |
| blank | blank | S&C - Scissors |  |  | 1 |  |  |  | 1 |
| blank | blank | S&C - Catch / Trap Point |  |  |  | 1 |  |  | 1 |
| blank | C Switch | S&C - Turnout | 1 |  | 1 |  |  |  | 2 |
| **Grand Total** | | | **519** | **399** | **282** | **135** | **84** | **16** | **1435** |

## **Physical Attribute Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **SWITCH\_DEPTH** | **SWITCH\_SIZE** | **SNC\_TYPE** | **TOTAL\_ASSETS** |
| Full Depth | C Switch | S&C - Turnout | 454 |
| blank | blank | blank | 193 |
| Full Depth | D Switch | S&C - Turnout | 161 |
| Full Depth | E Switch | S&C - Turnout | 108 |
| Shallow Depth | C Switch | S&C - Turnout | 93 |
| Full Depth | B Switch | S&C - Turnout | 92 |
| Full Depth | B Switch | S&C - Catch / Trap Point | 45 |
| Shallow Depth | E Switch | S&C - Turnout | 41 |
| Shallow Depth | D Switch | S&C - Turnout | 31 |
| Full Depth | blank | S&C - Switch Diamond | 25 |
| Full Depth | F Switch | S&C - Turnout | 25 |
| Full Depth | B Switch | S&C - Scissors | 14 |
| Shallow Depth | blank | S&C - Switch Diamond | 13 |
| Full Depth | C Switch | S&C - Catch / Trap Point | 12 |
| Full Depth | A Switch | S&C - Turnout | 11 |
| Shallow Depth | B Switch | S&C - Turnout | 10 |
| Shallow Depth | F Switch | S&C - Turnout | 10 |
| Full Depth | B Switch | S&C - Fixed Diamond Double Slip | 9 |
| Full Depth | B Switch | S&C - Switch Diamond Double Slip | 9 |
| Full Depth | C Switch | S&C - Scissors | 8 |
| Full Depth | B Switch | S&C - Switch Diamond Single Slip | 7 |
| Full Depth | A Switch | S&C - Catch / Trap Point | 6 |
| Full Depth | A Switch | S&C - Fixed Diamond Double Slip | 6 |
| Full Depth | D Switch | S&C - Switch Diamond | 6 |
| Full Depth | B Switch | blank | 4 |
| Full Depth | B Switch | S&C - Fixed Diamond Single Slip | 4 |
| Full Depth | blank | S&C - Switch Diamond Double Slip | 4 |
| Full Depth | G Switch | S&C - Turnout | 4 |
| Shallow Depth | SG Switch | S&C - Turnout | 4 |
| Shallow Depth | B Switch | S&C - Catch / Trap Point | 3 |
| Shallow Depth | C Switch | S&C - Catch / Trap Point | 3 |
| Full Depth | A Switch | S&C - Switch Diamond Double Slip | 2 |
| Full Depth | C Switch | S&C - Switch Diamond | 2 |
| Full Depth | D Switch | S&C - Fixed Diamond Single Slip | 2 |
| Shallow Depth | blank | S&C - Switch Diamond Double Slip | 2 |
| blank | C Switch | S&C - Turnout | 2 |
| Full Depth | A Switch | S&C - Fixed Diamond Single Slip | 1 |
| Full Depth | B Switch | S&C- Tandem Or Three Throw | 1 |
| Full Depth | B Switch | S&C - Switch Diamond | 1 |
| Full Depth | blank | S&C - Switch Diamond Single Slip | 1 |
| Full Depth | C Switch | S&C - Fixed Diamond Double Slip | 1 |
| Full Depth | D Switch | S&C - Fixed Diamond Double Slip | 1 |
| Full Depth | E Switch | S&C - Switch Diamond | 1 |
| Full Depth | H Switch | S&C - Turnout | 1 |
| Full Depth | H Switch | S&C - Scissors | 1 |
| Full Depth | Tandem A Switch | S&C - Fixed Diamond Single Slip | 1 |
| Full Depth | Tandem A Switch | S&C - Fixed Diamond Double Slip | 1 |
| Full Depth | Tandem B Switch | S&C - Fixed Diamond Single Slip | 1 |
| blank | blank | S&C - Scissors | 1 |
| blank | blank | S&C - Catch / Trap Point | 1 |
| **Grand Total** | | | **1439** |