# Clustering logs for increased log analysis efficiency

Theresa Spiel, Marius Führer, Stephan Grünfelder
2026

## Introduction

Automated tests for Riedel Communications Austria GmbH MediorNet broadcasting equipment generate roughly 1000 logs. Logs from failed test runs are currently manually triaged by testers per test system. This carries the risk of duplicated work, which costs time and money. This research aims to support manual analysis by clustering failed test runs. The objective is to increase analysis efficiency by grouping them based on similarity.

## Primary Research Question:

To what extent can clustering of failed test run log data identify related faults?

## Approach

The following prototype was evaluated within this work:

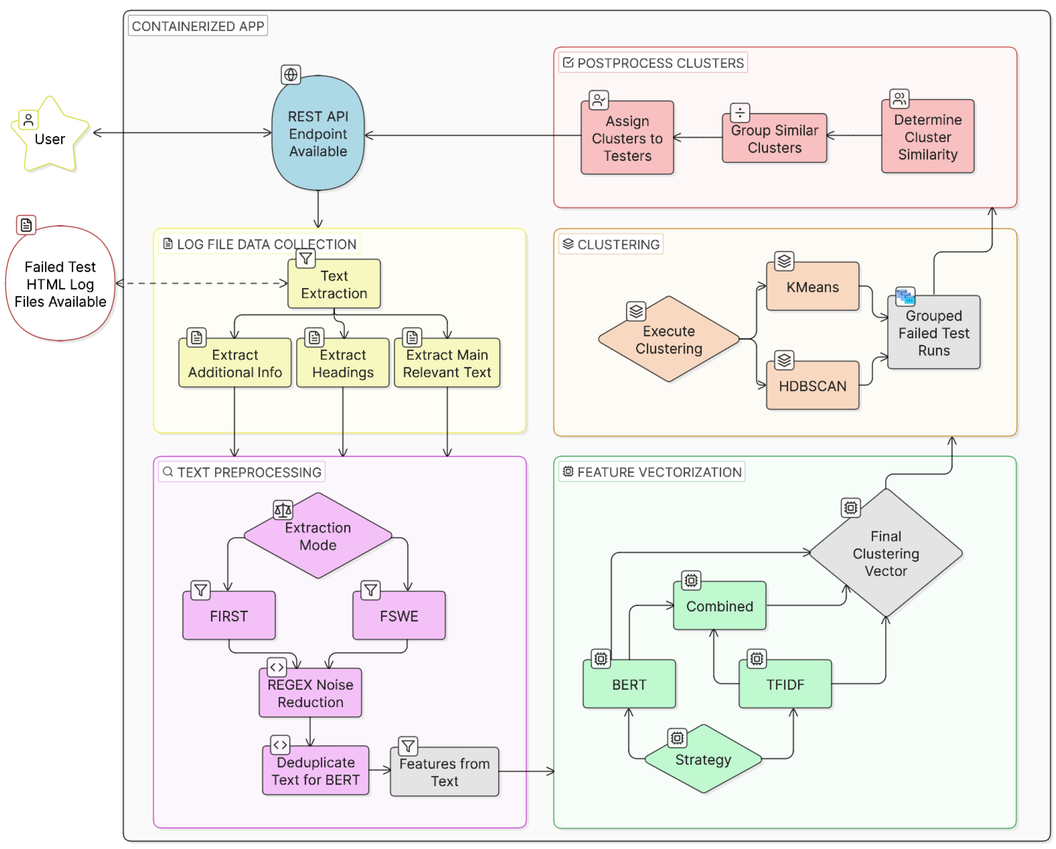


Figure 1: Application Prototype

The main steps are:
1) log data collection (yellow)
2) domain-specific preprocessing (purple)
3) feature vectorization (green)
4) clustering (brown)
5) cluster to test engineer assignment (red)

**Data**: 113 expert-labelled test campaigns with failed test runs only.

**Clustering algorithms**: HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) and KMeans [1].

## Evaluation

Expert-to-algorithm cluster alignment is evaluated via ARI and V-Measure, internal cluster consistency is evaluated via Silhouette and Calinski-Habarasz for both HDBSCAN and KMeans.

Additionally, internal metric Density-Based Clustering Validation for HDBSCAN and sum-of-squared errors in combination with the elbow metric for KMeans are applied for internal-external metric alignment.

## Quantitative clustering results

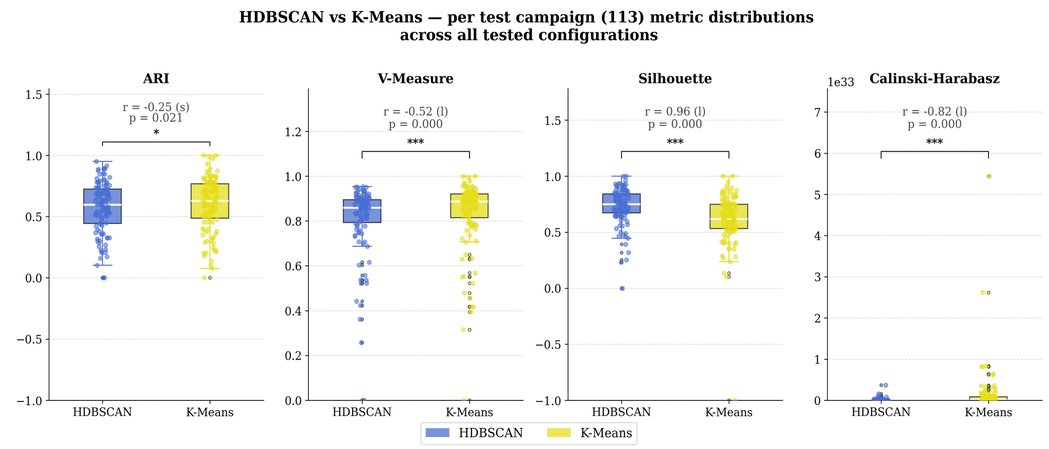


Figure 2: Quantitative metrics results

Algorithms perform almost equally well in external metrics, Silhouette favours HDBSCAN and Calinski-Habarasz is ignored due to numerical instabilities.

**HDBSCAN** is **selected** due to:
1) alignment with internal metrics
2) better inherent suitability for the data structure
3) noise treated as single-entry clusters

## Golden Sample qualitative results - HDBSCAN

ARI μ=0.665±0.088 / V-Measure μ=0.90±0.05

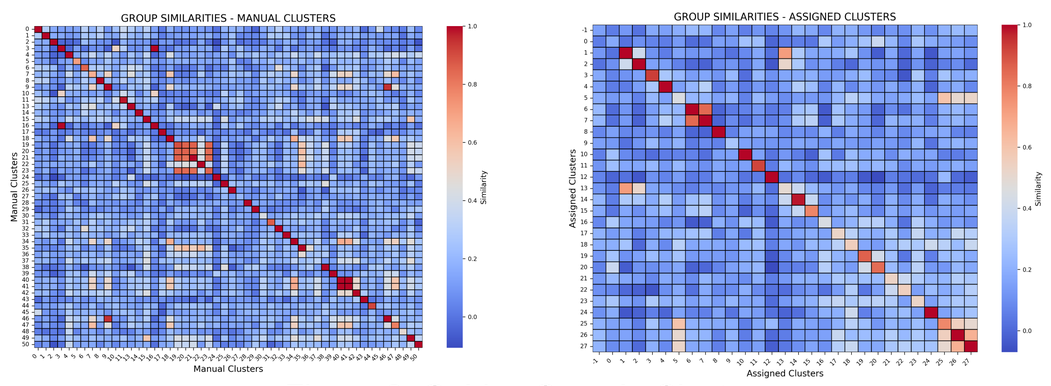


Figure 3: Golden Sample Clusters
Cluster cosine similarity heatmaps, blue (low) to red (high)

Similarity of expert-based (left) and algorithm-based clusters (right). While the former are sometimes split or aggregated, the latter remain at least thematically consistent. Even when diverging from expert-based clusters, the algorithm still groups thematically related failed test runs together.

Workload distribution among six engineers:

**Before** (by system): 7, 13, 26, 45, 22 → one receives no failed test runs for analysis.
**After** (by similarity): 18 failed runs for one, 19 for the remaining five test engineers.

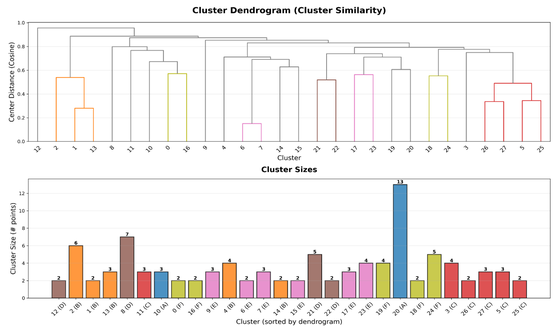


Figure 4: Cluster-to-Tester Assignment

## Conclusion

Unsupervised machine learning can identify related faults when testing RIEDEL MediorNet broadcasting equipment.

Homogeneous related faults are better identified than heterogeneous ones.

Initial feedback from testers **positive**: Tool helps in analysis and workload is distributed more evenly.

Reference:

[1] M. Abbas, A. Hamayouni, M. H. Moghadam, M. Saadatmand, and P. E. Strandberg, “Making sense of failure logs in an industrial devops environment”, in ITNG 2023 20th International Conference on Information Technology-New Generations, S. Latifi, Ed., Cham: Springer International Publishing, 2023, pp. 217–226, ISBN: 978-3-031-28332-1.