Clustering-Based Anomaly Detection for Connected Trucks

Abdulkadir Bilge[1], Erdem Ergen[1], Onur Ciliz[1], **Aslihan Reyhanoglu[1]**,
Feyzi Ege Kumec[1], Hira Bakirhan[2], Bugra Turan[3], and Sinem Coleri[3]

[1]Koc University Ford Otosan Automotive Technologies Laboratory (KUFOTAL), Istanbul, Turkey
[2]Ford Otosan, Sancaktepe, Istanbul, Turkey
[3]Department of Electrical and Electronics Engineering, Koc University, Istanbul, Turkey

E-mail: areyhanoglu@ku.edu.tr

2026

About the Presenter

- Aslıhan Reyhanoglu — AI R&D Engineer (Koc University), Istanbul
- Focus area: applied ML/AI, telemetry-driven analytics, predictive modeling, anomaly monitoring, communication networks, vehicular communications and connected vehicle systems.
- Education: MSc in Radio and Mobile Communication Systems; BSc in Electronic and Communication Engineering
- Research areas and publications:
  - C-V2X Mode 4 reliability analysis (IEEE VTC-Fall 2022)
  - NR-V2X QoS prediction (IEEE VNC 2023)
  - Federated learning for pedestrian detection (IEEE BlackSeaCom 2023)
  - Network-aware bitrate prediction for remote driving (IEEE BlackSeaCom 2025)

Outline

Introduction

- Proactive fault detection is critical for enhancing autonomous driving reliability and safety.
- Predictive Maintenance (PdM) aims to identify potential system failures before they result in downtime.
- Unsupervised learning offers a framework for discovering patterns in large-scale, unlabeled truck telemetry.
- This study benchmarks five clustering-based anomaly detection methods using real-world engine data.

Problem Statement

- Heterogeneous sensor data from connected trucks presents challenges like irregular sampling and missing values.
- Centralized systems face scalability issues with high-dimensional telemetry streams.
- Existing literature often lacks integrated multi-algorithm benchmarking on real truck subsystems.

System Model: Data and Preprocessing

- 14 days of telemetry from two identical trucks.
- Raw dataset: 952,815 data points, 90 features.
- After preprocessing: 43,858 samples, 6 selected features.
- Real-world telemetry collected from instrumented Ford F-Max test trucks.
- High-frequency multivariate signals (vehicle health and operational status streams).
- Removed invalid/out-of-range records; handled missing values and time misalignments.
- Selected vehicle-health related signals; derived a compact feature vector per time window.
- Standardized features (e.g., z-score) to make signals comparable across scales.
- A normalized multivariate dataset ready for clustering and anomaly scoring.

Feature Correlation Analysis & Selected Features

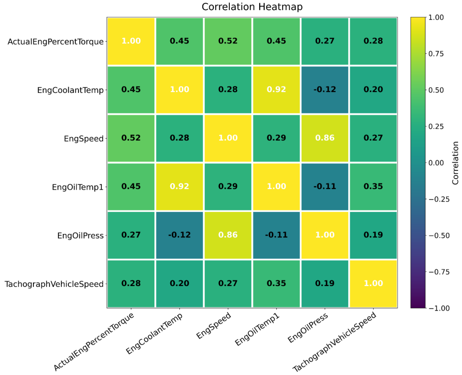


Figure: Correlation Heatmap of Vehicle Health Features.

- We selected engine-health signals that show meaningful relationships ($|\rho| > 0.3$)
- **Selected (6):** *ActualEngPercentTorque*, *EngSpeed*, *TachographVehicleSpeed*, *EngCoolantTemp*, *EngOilTemp1*, *EngOilPress*.

Anomaly Detection Algorithms

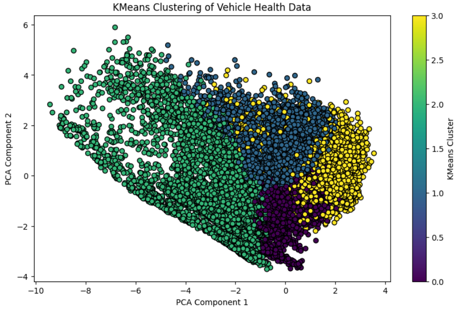
- **K-Means** (operational-mode clustering)
- **Isolation Forest** (outlier scoring)
- **Z-Score** *(within K-Means clusters)*
- **GMM** (probabilistic clustering)
- **DBSCAN** (density-based clustering)

Domain-Specific Threshold Validation

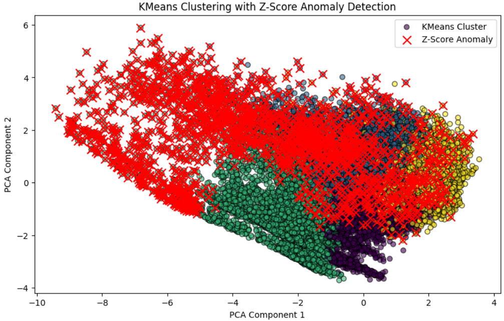
- We validated detected anomalies against domain-defined physical operating limits:
  - **Engine Torque (ActualEngPercentTorque)**: 90% (high-load operation).
  - **Engine Speed (EngSpeed)**: 2500 rpm (high-speed operating region).
  - **Engine Coolant Temperature (EngCoolantTemp)**: 100°C (overheating risk).
  - **Engine Oil Temperature (EngOilTemp1)**: 110°C (anti-wear additive degradation risk).
  - **Engine Oil Pressure (EngOilPress)**: 600 kPa (upper operating bound).
  - **Vehicle Speed (TachographVehicleSpeed)**: 90 km/h (safety / compliance barrier).

Benchmarking Results: Clustering and Anomaly Overlay (PCA View)

- Clusters capture operational modes; Z-score highlights extreme deviations mainly near sparse/boundary regions.



K-Means clusters



Z-score anomalies (red)

Benchmarking Results

Method	Silhouette Score	Calinski-Harabasz Index
Z-Score (within K-Means clusters)	**0.569**	7777.99
GMM	0.513	6379.11
Isolation Forest	0.473	671
K-Means	0.33	**19281.7**
DBSCAN	0.297	5.916

Table: Comparative performance of unsupervised methods.

- Scores computed on cluster labels derived from K-Means; anomaly flags used as overlay.
- Applying Z-Score filtering on K-Means clusters yielded the highest structural separability (Silhouette: 0.569), indicating a clear separation between typical and atypical operational patterns.
- K-Means exhibited strong cluster cohesion but was less sensitive to subtle anomalies.

Conclusion and Future Directions

- Z-Score analysis provides a transparent and precise per-signal sanity check.
- An ensemble approach effectively handles the complexities of real-world telemetry.
- **Future Work**:
  - Transitioning from offline batch processing to online/streaming anomaly detection on edge hardware.
  - Exploring temporal modeling and change-point detection on multivariate telemetry.
  - Integrating maintenance logs for weak supervision and threshold calibration.
  - Expanding analysis to larger fleets and seasonal operational variability.

Questions

Thank you for your attention!

Aslihan Reyhanoglu

areyhanoglu@ku.edu.tr