CAN Bus Anomaly Detection - K-Means Clustering

DATASET

2,804,975 Total Messages

94.48% Normal

5.52% Attacks

FEATURES (10)

data0-data7, dlc, iat

Preprocessing: StandardScaler (Mean=0, Std=1)

K-MEANS PARAMETERS

n_clusters: **3**

random_state: **42**

n_init: **10**

max_iter: **300**

ATTACK TYPES

119,627 Fuzzing

35,093 DoS

CLUSTER DISTRIBUTION

Cluster 0

1,171,636 (41.77%)

Normal: 93.47%

Attack: 6.53% (DoS + Fuzzing)

Cluster 1 (Safest)

1,005,403 (35.84%)

Normal: 98.56%

Attack: 1.44%

Cluster 2 (Highest Risk)

627,936 (22.39%)

Normal: 89.85%

Attack: 10.15%

QUALITY METRICS

0.223 Silhouette

2.025 Davies-Bouldin

610,692 Calinski-Harabasz

Key Findings and Conclusions

KEY FINDINGS

1. **Cluster 2 Concentrates Fuzzing Attacks**
   10.15% attack rate with 63,736 fuzzing attacks

2. **DoS Attacks Isolated in Cluster 0**
   All 35,093 DoS attacks (100%) in one cluster

3. **Attack Types Have Different Signatures**
   K-Means separates attacks without using labels

ATTACK DISTRIBUTION BY CLUSTER

	C0	C1	C2
Fuzzing	3.54%	1.44%	**10.15%**
DoS	**3.00%**	0%	0%

CONCLUSIONS

- ✓ K-Means groups CAN messages into distinct patterns without labels
- ✓ DoS and Fuzzing attacks show different data signatures
- ✓ Cluster membership can serve as a risk indicator

LIMITATIONS

Moderate silhouette score (0.223) indicates some cluster overlap. Unsupervised approach has accuracy limitations.

FUTURE WORK

Combine with supervised methods for hybrid detection systems

TOOLS

Python, pandas, NumPy, scikit-learn, Matplotlib, Seaborn