Predictive Maintenance for Heavy-Duty Truck Fleets: An Explainable ML Approach on the SCANIA Component X Dataset

Author: Irvin E. Hernández Vázquez
Advisor: Neliud Torres Batista, DBA
Master in Computer Science,
Graduate Project EXPO, May 2026

Abstract

This study develops a cost-optimized predictive maintenance system on the SCANIA Component X Dataset (33,641 heavy-duty trucks). A calibrated LightGBM model with class weighting achieved a 45% cost reduction and 86% failure recall on the test set. Ultimately, configuration-stratified SHAP analysis, revealed that vehicle specification groups exhibits mechanistically distinct failure pathways, with Spearman rank correlations between top features ranging from 0.38 to 0.96. These findings challenge the homogeneous fleet assumption common in prior work.

Background

Prior work on the SCANIA dataset has focused on cost optimization (Kim & Kim 2026 LightGBM) and architectural sophistication (CNN/LSTM, GNN, survival analysis). All prior studies treat the fleet as a single homogeneous population, leaving failure pathway heterogeneity across specification groups unaddressed.

Introduction

Component failures in industrial fleets generate costly downtime and safety risks.

Predictive maintenance using sensor data and machine learning offers a proactive solution, but real-world deployment is hindered by datasets that ignore operational realities.

Real fleets encounter class imbalance, heterogeneous configurations, and unreliable sensor readings. Consequently, degrading standard model performance.

To addresses this, we used a comprehensive truck dataset to predict Component X failures while simultaneously applying SHAP to identify critical failure-driving signals. By analyzing SHAP explanations across vehicle specification groups, we identify configuration-specific failure pathways.

Problem

(RQ1) Which classifier achieves the lowest cost under Scania's asymmetric cost matrix (50× penalty for missed failures vs false alarms)? (RQ2) Do failure pathways differ across truck specification groups? RQ2 is the novel knowledge-discovery contribution of this study.

Methodology

1.1M readouts → one row per truck (last readout). LightGBM trained with class weighting (scale_pos_weight=9.37), tuned via 100 Optuna trials with early stopping on Scania cost. Probability calibration uses temporal 5-fold OOF predictions. Cost-optimal threshold selected on validation set. SMOTE tested as ablation. SHAP TreeExplainer computed on test set, partitioned by specification category. Spearman rank correlations computed between top-10 feature rankings to quantify pathway divergence.

Figure 1: Methodology Pipeline

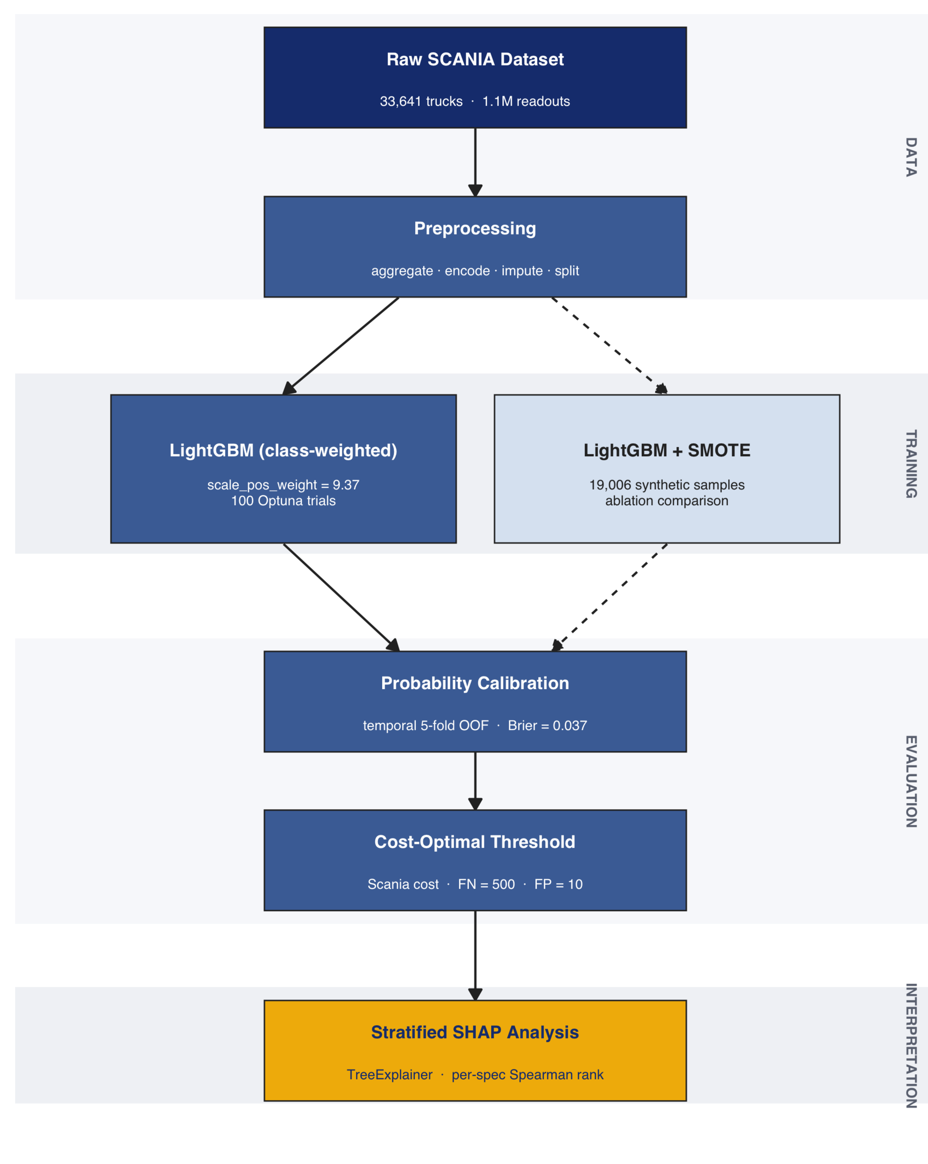

Figure 1: Methodology pipeline. The class-weighted LightGBM is the primary model; SMOTE is trained only as an imbalance-strategy ablation (dashed branch). Both pass through identical probability calibration and cost-optimal thresholding before stratified SHAP interpretation across vehicle specification groups.

Table 1: Model Performance Comparison

Metric	Baseline	LightGBM	LightGBM + SMOTE
Scania Cost	71,000	39,030	42,760
Cost Reduction		45.0%	39.8%
ROCAUC	0.500	0.6858	0.6426
Recall	0.000	0.8592	0.7958
Brier Score		0.0368	0.034
True Positives	0	122	113
False Positives	0	2,903	2,826
False Negatives	142	20	29
True Negatives	4,903	2,000	2,077
Trees		33	49
Threshold		0.095	0.095
Cost per Truck	14.07	7.74	8.48

Table 1: Test-set performance comparison across the three approaches. The class-weighted LightGBM achieves the lowest Scania cost (39,030 vs. 71,000 baseline, 45% reduction) and highest recall (85.9%), outperforming the SMOTE ablation by 3,730 cost units despite SMOTE training on a larger augmented dataset. Brier scores below 0.04 confirm well-calibrated probability outputs for both models.

Results and Discussion

Class-weighted LightGBM: Scania cost 39,030 (45.0% reduction from 71,000 baseline), 85.9% failure recall (122/142). SMOTE underperformed at 42,760 (39.8% reduction), with lower ROC-AUC (0.643 vs. 0.686). Global SHAP top features: 158_9, 167_1, 167_6. Stratified analysis reveals divergent failure pathways: Spec_3 Cat1 vs Cat3 Spearman $\rho = 0.383$ (fewer than 40% top-feature overlap); Spec_0 Cat0 vs Cat1 $\rho = 0.719$, with 158_9 dominant for one configuration and 167_6 for the other. Across all 13 specification pairs, ρ ranges from 0.383 to 0.964 (Figure 3). Notably, different vehicle configurations exhibit mechanistically distinct failure pathways.

Table 2: Imbalance Strategy Ablation

Strategy	Training Rows	Positive Cases	Test Cost	Cost Reduction	Missed Failures
Naive baseline			71,000		142
Class weighting	23,550	2,272	39,030	45.0%	20
SMOTE	42,556	21,278	42,760	39.8%	29

Table 2. Comparison of class imbalance handling strategies on the SCANIA Component X test set (5,045 trucks, 142 failures). Class weighting achieves the lowest Scania cost of 39,030 (45.0% reduction from the 71,000 naive baseline) while missing only 20 failures. SMOTE, despite expanding the training set with 19,006 synthetic failure cases, underperforms at 42,760 cost with 29 missed failures. Cost computed using the asymmetric Scania cost matrix (FN = 500, FP = 10).

Table 3 SHAP RESULT

Spec Column	Category A	Category B	Trucks A	Trucks B	Spearman Corr	Top Divergent Features	Interpretation
Spec_3	Cat1	Cat3	1,059	431	**0.383**	158_8, Spec_3, 309_0	Strongest divergence
Spec_3	Cat0	Cat3	3,077	431	0.417	158_8, Spec_3, 309_0	Strong divergence
Spec_3	Cat2	Cat3	478	431	0.583	Spec_3, 309_0, 167_2	Moderate divergence
Spec_0	Cat0	Cat1	4,130	910	0.719	158_8, 309_0, 666_0	Moderate divergence
Spec_6	Cat1	Cat4	1,603	590	0.730	158_8, 309_0, 666_0	Moderate divergence
Spec_7	Cat4	Cat1	1,001	901	0.765	167_2, 459_14, 158_8	Mild divergence
Spec_6	Cat0	Cat4	2,176	590	0.772	158_8, 309_0, 167_2	Mild divergence
Spec_7	Cat0	Cat4	1,629	1,001	0.804	167_2, 666_0, 158_8	Mild divergence
Spec_3	Cat1	Cat2	1,059	478	0.828	158_8, 167_2, 459_14	Low divergence
Spec_3	Cat0	Cat2	3,077	478	0.873	167_6, 167_1, 158_8	Low divergence
Spec_6	Cat0	Cat1	2,176	1,603	0.955	158_9, 167_2, 158_6	Similar pathways
Spec_3	Cat0	Cat1	3,077	1,059	0.964	167_1, 167_6, 167_2	Similar pathways
Spec_7	Cat0	Cat1	1,629	901	0.964	158_9, 397_3, 167_1	Similar pathways

Table 3. Pairwise Spearman rank correlations between top-10 SHAP feature importance rankings across truck specification category pairs. Lower correlation indicates greater divergence in failure pathways. Spec_3 Cat1 vs Cat3 ($\rho = 0.383$) exhibits the strongest divergence, with differential markers 158_8, Spec_3, and 309_0 driving risk attribution distinctly between configurations. Pairs with $\rho \geq 0.95$ show near-identical failure pathways. (**) denotes the headline knowledge-discovery finding.

Conclusions

(1) Class-weighted LightGBM achieves 45.0% cost reduction with 85.9% failure recall. (2) Class weighting outperforms SMOTE by 3,730 cost units. (3) Stratified SHAP reveals distinct failure pathways across specification groups (Spearman 0.38-0.96), challenging the homogeneous fleet assumption. (4) Configuration-aware alerting could improve maintenance efficiency.

Future Work

1) Re-engineer binary labels into 5-class severity to enable the full IDA 2024 cost matrix. (2) Apply survival analysis (Cox, DeepSurv) for time-to-failure modeling. (3) Use conformal prediction for uncertainty intervals. (4) Extend stratified explainability to high-frequency CAN-bus streams.

Acknowledgements

Thanks to Dr. Neliud Torres Batista for advisory guidance throughout this Design Project and to Joann Casillas for reviewing the article. Additional thanks to the Polytechnic University of Puerto Rico for institutional support and to the 2024 Industrial Challenge organizers for making this dataset publicly available to the research community.

References

For references, please scan the QR code.