Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p27658-cd

BOOSTING WARRANTY PREDICTION IN THE AUTOMOTIVE SECTOR USING MACHINE LEARNING: A COMPARATIVE STUDY

O. EL BAKKALI
University of Angers, LARIS, SFR MATHSTIC, F-49000 Angers, France, oumaima.elbakkali@etud.univ-angers.fr

R. LARONDE
VALEO, F-63500 Issoire, France, remi.laronde@valeo.com

A. BELCAID
University of Abdelmalek Essaadi, ISI MA-93030 Tetouan, Morocco, a.belcaid@uae.ac.ma

K. REKLAOUI
University of Abdelmalek Essaadi, ISI MA-93030 Tetouan, Morocco, kreklaoui@uae.ac.ma

A. KOBI
University of Angers, LARIS, SFR MATHSTIC, F-49000 Angers, France, abdessamad.kobi@univ-angers.fr

This paper investigates the application of machine learning techniques for reliability forecasting in the automotive industry, focusing on the prediction of Months In Service (MIS), that is, the number of months between a vehicle's warranty start date and repair date. The accurate estimation of MIS is of crucial importance to improve the forecast of warranty costs, reliability planning, and maintenance strategies. Traditional reliability tools, such as Weibull analysis, work effectively under strict statistical assumptions but often fail to capture nonlinear relationships and interdependencies within complex high-dimensional field data. The aim of this paper is to compare two machine learning algorithms: Decision Trees (DT) and Gradient Boosting Decision Trees (GBDT). Their performance was compared based on real industrial warranty data from a Tier-1 automotive supplier. The dataset includes categorical, numerical, temporal, and operational variables describing vehicle characteristics and usage conditions. This comparison underlines the fact that the ensemble learning approach significantly outperforms the single-tree model in accuracy, generalization capability, and robustness when predicting MIS. Experimental results confirm the superiority of the boosting algorithm. These observations allow the conclusion that ensemble methods really improve the results of reliability forecasting due to modeling complex nonlinear dependencies. This work is concluded by verifying that boosting-based models provide a flexible, accurate, and data-driven approach; it opens new frontiers to improve current methodologies for reliability analysis and warranty management in the automotive sector.

Keywords: warranty prediction, machine learning, Weibull distribution, automotive industry, MIS, GBDT, DT

1. Introduction

Accurate estimation of warranty returns is a critical concern in industrial sectors, especially in automotive manufacturing, as the cost of warranty forms a significant share of after-sales expenses. It has been estimated that automobile warranties cost over $12 billion annually in North America alone and accounted for more than half of all U.S. manufacturer warranty costs (El Bakkali et al. (2025); Kleyner and Sandborn (2006)). Good warranty forecast estimation is important not only for effective financial planning and price competitiveness but also for supplier-OEM negotiations, lifecycle cost management, and customer satisfaction. Good forecasting prevents quality crises, provides better budgeting, and allows proactive risk reduction strategies. Traditional models, such as the Weibull distribution, introduced by Waloddi Weibull in "A Statistical Distribution Function of Wide Applicability" Weibull (1951), have been widely used over the last decades for warranty and reliability analysis. Weibull-based models provide simplicity, flexibility, and the possibility of mixed-Weibull extensions in order to model heterogeneous failure modes across product lifecycles Attardi et al. (2005). However, these parametric models count on strong assumptions regarding the distributions of failures, which may not reflect the complexity of real-world industrial data.

As pointed out by Kleyner and Sandborn Kleyner and Sandborn (2006), neglecting two-dimensional usage factors, such as mileage accumulation over time and component aging, leads to significant overestimation of warranty costs, which again underlines the need for more realistic and data-driven approaches. Although recent works from Stellantis, Renault, and Valeo experts contain multi-mode Weibull models, which give better modeling of high-mileage reliability and market heterogeneity, the models are still inherently parametric and can fail in case of complicated or irregular data structure Bonvin et al. (2022). This is why machine learning models, particularly tree-based models such as DT and GBDT, attract attention because of their interpretability, flexibility, and excellent performance in prediction. DT provides an intuitive hierarchical structure for failure behavior modeling, whereas GBDT iteratively combines a lot of weak learners to improve accuracy with robustness, hence capturing complex interactions in warranty data.

This work focuses on the prediction of MIS, using real industrial data from a Tier-1 automotive supplier. As shown in Fig.1, MIS corresponds to the time elapsed between the warranty start date and the repair date, representing a key indicator of component reliability. By applying DT and GBDT models, we seek to exploit such complexity of field data in order to enhance warranty cost estimation, early failure trend detection, and decision-making in reliability and main-

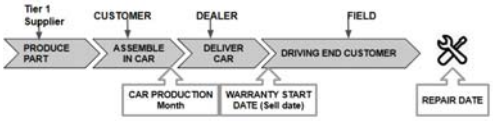

Fig. 1. Life Cycle of Automotive Component

tenance management. The paper is structured as follows: Section 2 describes the methodology, covering data preparation, the approach to modeling, and the metrics used in the evaluation. Section 3 discusses the data results and compares performances of models. Finally, Section 4 gives the conclusion and outlines future research directions focusing on integrating AI-driven methods into warranty estimation.

2. Methodology

In this work, the estimation of the MIS for automotive parts is done with the use of machine learning models, particularly DT and GBDT. The approach will be properly structured: preparation of the data, feature engineering, model development, tuning of hyperparameters, and evaluation of performances.

2.1. *Data Collection and Cleaning*

The data used in this research came from the internal warranty database of a Tier-1 automotive supplier, which tracks field returns, repair information, and in-service and operational details relative to customer vehicles. Access to this system was provided after formal authorization by the management, in full compliance with all internal data governance and confidentiality rules. The required records were extracted in a restricted-access environment by the company's reliability team, as any transfer or storage outside the company of warranty data is forbidden by corporate policy. During extraction, the dataset was directly filtered to retain only information within the study scope that was pre-defined according to the company's target, such as the selected customer, relevant product families, and the time window. Once the dataset was delivered, it was stored securely on the corporate laptop that was assigned for this project, and all machine-learning development has been carried out using the company's professional account, in strict adherence with the confidentiality constraints. The extracted dataset contains 54 columns and 17,883 rows. To ensure its quality and homogeneity, several filtering rules were applied to the data. The records were restricted to components whose repair dates were after January 1, 2022, since prior data were inconsistent. The logical sequencing was checked by retaining only the claims where the Inservice_Date occurs after the Production_Date. Target variable MIS was restricted within the warranty agreement period with the customer. These steps help to ensure high-quality input data and minimize any potential bias due to outliers or anomalous records. The filtered database contained 54 columns and 7,953 rows, as shown in Fig. 2.

2.2. *Feature engineering and preprocessing*

Feature engineering and preprocessing are important steps toward the enhance-

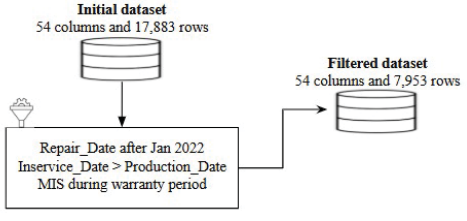

Fig. 2. Dataset filtering rules

ment of both model interpretability and predictive performance. Several transformation and selection operations were implemented to make sure the construction of meaningful representations of the underlying variables was performed accordingly. Fig. 3 describes the different steps to obtain the final dataset, including the imputation and encoding strategy used.

The first preprocessing step was the temporal feature transformation, meaning Production_Date and Inservice_Date had to be converted into numerical variables representing the number of days since a reference date. This allowed both models to capture potential linear temporal relationships with the target variable, MIS. Moreover, date decomposition was done to split Production_Date and Inservice_Date into their constituent elements, day, week, month, and year, to capture possible seasonality and batch-related effects. This step was informed by studies showing that calendar-month seasonality significantly impacts subsystem failure patterns and warranty expense forecasting Rai (2009). Sine and cosine transformations were therefore used to create continuous cyclic encodings and provide a better representation of the cyclical nature of time variables. This strategy will allow smooth transitions across boundary values, such as between December and January. Moreover, it has been shown in Bansal et al. (2025) that for time-dependent prediction tasks, this can lead to better results. Also, redundant or irrelevant columns like Production_Date and Inservice_Date were removed to reduce noise. Finally, a hybrid feature selection approach was performed.

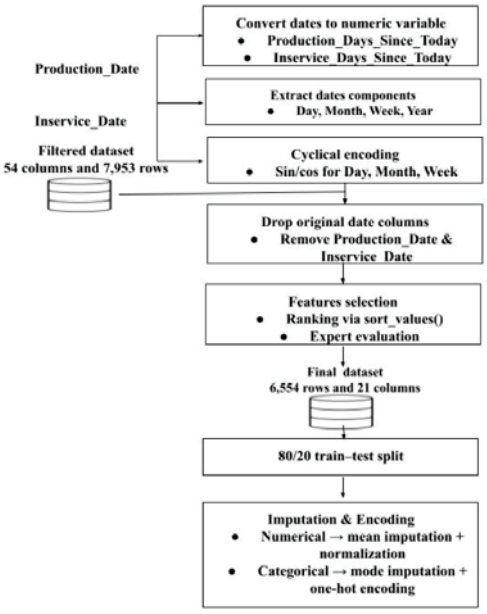

Fig. 3. Feature Engineering and preprocessing framework

In the ranking step, correlation analysis using Python sorted the features in descending order in respect to their relation with the MIS variable by the function *sort_values()* and identified the most predictive features. The second step of this analysis further refined these results by integrating inputs from the reliability expert in order to ensure that these features are relevant for the domain and represent real-life conditions. This two-tier process ensured only the most influential and meaningful features were retained for modeling, enhancing both model robustness and interpretability. The final dataset contained 6,554 rows and 21 columns. The data was then split into test and train sets (80/20). For numerical features, missing values were filled with the mean and then normalized,

and for categorical features, missing values were filled with the most common value and converted into one-hot vectors. All preprocessing steps were done using scikit-learn's ColumnTransformer to ensure consistency and reproducibility.

2.3. *Decision Tree*

Decision Trees (DT) Zhao and Nie (2021) is a flexible, interpretable model suited for regression and classification; it performs the partitioning of the feature space in a hierarchical manner using decision rules. For regression, its objective function is defined as

$$\begin{aligned}\text{obj}_{j,s} = \min \sum_{i \in R_1^{j,s}} (y_{\text{true}} - c_1)^2 \\ + \sum_{i \in R_2^{j,s}} (y_{\text{true}} - c_2)^2\end{aligned} \tag{1}$$

where c_1 is the mean target value in $R_1(j,s)$, and c_2 is the mean target value in $R_2(j,s)$. The subsets $R_1(j,s)$ and $R_2(j,s)$ can be the results from the split at the j-th feature with a threshold Zhu et al. (2024). The hyperparameters tuned for this model are the maximum tree depth *max_depth*, the minimum number of samples required at a leaf node to regulate tree growth *min_samples_leaf*, and the minimum number of samples required to split an internal node *min_samples_split*. The three hyperparameters were optimized to regulate tree growth, prevent overfitting, and extract meaningful rules from the data

2.4. *Gradient Boosted Decision Trees*

Gradient Boosted Decision Trees (GBDT) Rizkallah (2025) is an ensemble model with a strong predictive capacity. Its principle relies on combining a large number of weak decision trees, where each tree is trained on the residual errors of the previous iteration, thereby focusing on the most difficult instances to predict Rizkallah (2025). The regression objective function is

$$\text{Obj}_{j,n} = \arg \min_{x_i \in R_j^n} L\big(y_{\text{true}}, F_{n-1}(x_i)\big) + \gamma \tag{2}$$

where $F_{n-1}(x_i)$ is the prediction value at the $(m-1)$ iterations, γ is the optimal adjustment value, and R_j^n is the region corresponding to the j-th leaf node in the n-th tree. GBDT can model complex nonlinear relationships and feature interactions, while regularization enhances its robustness. The key hyperparameters tuned in our experiments were the number of estimators *n_estimators*, the learning rate *learning_rate*, and the maximum tree depth *max_depth*.

2.5. *Hyperparameter optimization and model training*

For both DT and GBDT, hyperparameter tuning was performed using RandomizedSearchCV with **5-fold** cross-validation. All preprocessing steps and model fitting were done within a unified pipeline to ensure reproducibility. The randomized search effectively explored the hyperparameter space in a manner that balanced computational cost with model performance. Table 1 presents the hyperparameters and their corresponding ranges for the two models evaluated: DT and GBDT.

Table 1. Hyperparameter ranges for DT and GBDT models.

Model	Hyperparameter	Range
DT	max_depth	5 – 20
	min_samples_leaf	1 – 10
	min_samples_split	2 – 20
GBDT	n_estimators	50 – 500
	learning_rate	0.01 – 0.3 (uniform)
	max_depth	3 – 10

2.6. *Model evaluation*

Model performance was evaluated on several measures:

Train score & Test score R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

Mean Square Error:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

Mean Absolute Error

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

Mean Absolute Percentage Error :

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (7)$$

3. Data and Results Discussion

Both DT and GBDT models were applied to predict MIS. The randomized search with five-fold cross-validation was used for model training and hyperparameter tuning to find the best balance between model complexity and generalization. The optimization process was conducted using a company-issued laptop. Model development and execution of scikit-learn library was carried out within a secured computing environment. Per internal data-protection policies, the warranty dataset was strictly confidential and never uploaded or shared outside the company infrastructure; all computations were executed on the secured corporate device locally. The DT model was first evaluated to assess its capability to capture the relationship between warranty data features and MIS. After hyperparameter tuning, the optimal configuration, as shown in Table 2, included a maximum depth of 17, a minimum of 1 sample per leaf, and a minimum of 6 samples required for a split. These parameters provided the best trade-off between predictive power and overfitting control. The DT model achieved the performances shown in Table 3, an R^2 score of **0.7171** on the training set and **0.4989** on the test set, indicating limited generalization capability on unseen data. With respect to error metrics, the model produced an MSE of 38.8113, an RMSE of 6.2299 months, an MAE of 4.7132 months, and a MAPE of 0.5371. These findings highlight that, although the DT model was able to capture certain structure in the data, its predictive performance remained modest when

On the contrary, the GBDT model demonstrated strong prediction power in forecasting the MIS. After hyperparameter optimization with random search, the model was optimally obtained with the parameter values described in Table 2: a number of estimators equal to 355,

Table 2. Hyperparameters used for DT and GBDT models.

Model	Best Parameters
DT	max_depth = 17 min_samples_leaf = 1 min_samples_split = 6
GBDT	n_estimators = 355 learning_rate = 0.0833 max_depth = 8

a learning rate of 0.0833, and a maximum depth of 8. This configuration was found to achieve a balance between model complexity and generalization capability. The GBDT model reached the performance levels displayed in Table 3. On the training set, the model's R^2 score was 0.8868, indicating a very good fit. More importantly, the model was performing high on the test set, with an R^2 score of 0.7555, confirming its generalization capability on unseen data. In terms of error metrics, the MSE value was 18.9400, showing a good moderation between the predicted and actual MIS. Accordingly, the RMSE was 4.3520 months. The MAE was 3.2841 months, confirming that the model maintained a reasonable average error in practical terms. Finally, the MAPE reached 0.3106, which could be accepted as reasonable, considering the variations in field conditions and part usage across regions and different customer environments. These results are the evidence of the capability of the GBDT model to provide actionable insights regarding warranty failure timelines. Its capability to detect nonlinear interaction and heterogeneous feature contribution without much manual engineering is a significant leap over traditional reliability practice traditionally based on parametric lifetime distribution.

Table 3. Error metrics of DT and GBDT models.

Metric	DT	GBDT
R^2 (Train)	0.7171	0.8868
R^2 (Test)	0.4989	0.7555
MSE	38.8113	18.94
RMSE	6.2299	4.352
MAE	4.7132	3.2841
MAPE	0.5371	0.3106

4. Conclusion

This paper examines the performance of two machine learning techniques, DT and GBDT, in predicting MIS, which is considered one of the most critical indicators of early warranty return forecasting for the automotive industry. Based on real industrial data and a strict experimental framework using randomized search with five-fold cross-validation, the analysis underlined the contrasting capabilities of both models. Although the DT model identified some useful patterns, its predictive performance remained limited on unseen data. The higher error values and lower R^2 scores show that DTs alone cannot fully capture the complexity and nonlinear relationships found in real warranty data. Conversely, the GBDT model showed a robust predictive performance and robust generalization capability with a good score for all metrics, including R^2, RMSE, MAE, MSE, and MAPE. These results confirm that GBDT models can effectively learn complex feature interactions and heterogeneous usage patterns without the need for extensive manual feature engineering. This superior performance further points out the relevance of

advanced ensemble methods toward complementing or surmounting some traditional parametric reliability models.

Overall, the GBDT model outperformed the DT model across all performance metrics, therefore confirming its suitability for warranty return forecasting. In fact, compared with traditional reliability analysis methods based on parametric lifetime distributions, it is capable of modeling nonlinear interactions and heterogeneous feature contributions without requiring extensive manual feature engineering. The predictive accuracy and generalization potential of the model make it particularly useful in proactive assessments of the reliability of the product and optimization of warranty costs within the automotive industry.

Advanced models using AI, such as Random Forests and XGBoost, will be explored further to improve the predictive performance and better capture complex nonlinear relationships in warranty data.

References

Attardi, L., M. Guida, and G. Pulcini (2005). A mixed-weibull regression model for the analysis of automotive warranty data. *Reliability Engineering & System Safety 87*, 265–273.

Bansal, A., K. Balaji, and Z. Lalani (2025, mar). Temporal encoding strategies for energy time series prediction. arXiv. Working paper.

Bonvin, L., C. Ramus Serment, N. Forissier, and R. Laronde (2022, may). Modeling market incident levels (warranty & over) using weibull distribution (part 2). Technical report, Société des Ingénieurs de l'Automobile. Working paper.

El Bakkali, O., R. Laronde, A. Belcaid, K. Reklaoui, and A. Kobi (2025). Warranty prediction of automotive components using machine learning: Case study of a tier-1 supplier. In *Proceedings of ICAT 2025*.

Kleyner, A. and P. Sandborn (2006). Forecasting the cost of unreliability for products with two-dimensional warranties. In *Proceedings of the European Safety and Reliability Conference*, pp. 1903–1908.

Rai, B. K. (2009). Warranty spend forecasting for subsystem failures influenced by calendar month seasonality. *IEEE Transactions on Reliability 58*, 649–657.

Rizkallah, L. (2025). Enhancing the performance of gradient boosting trees on regression problems. *Journal of Big Data 12*, 35.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics 18*, 293–297.

Zhao, X. and X. Nie (2021, oct). Splitting choice and computational complexity analysis of decision trees. *Entropy 23*(10), 1241.

Zhu, T., X. Mei, J. Zhang, and C. Li (2024). Prediction of thermal conductivity of eg–al_2o_3 nanofluids using six supervised machine learning models. *Applied Sciences 14*, 6264.