

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Sarode, Mohit S.; Kumar, Anil; Prasad, Abhijit; Shetty, Abhishek

Article

Enhancing pricing strategies in the aftermarket sector with machine learning

Modern Supply Chain Research and Applications

Provided in Cooperation with: Emerald Publishing Limited

Suggested Citation: Sarode, Mohit S.; Kumar, Anil; Prasad, Abhijit; Shetty, Abhishek (2024) : Enhancing pricing strategies in the aftermarket sector with machine learning, Modern Supply Chain Research and Applications, ISSN 2631-3871, Emerald, Bingley, Vol. 6, Iss. 4, pp. 411-423, https://doi.org/10.1108/MSCRA-10-2023-0042

This Version is available at: https://hdl.handle.net/10419/314934

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.



https://creativecommons.org/licenses/by/4.0/

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

Enhancing pricing strategies in the aftermarket sector with machine learning

Mohit S. Sarode, Anil Kumar, Abhijit Prasad and Abhishek Shetty Daimler Truck Innovation Center India, Bangalore, India

Abstract

Purpose – This research explores the application of machine learning to optimize pricing strategies in the aftermarket sector, particularly focusing on parts with no assigned values and the detection of outliers. The study emphasizes the need to incorporate technical features to improve pricing accuracy and decision-making.

Design/methodology/approach – The methodology involves data collection from web scraping and backend sources, followed by data preprocessing, feature engineering and model selection to capture the technical attributes of parts. A Random Forest Regressor model is chosen and trained to predict prices, achieving a 76.14% accuracy rate.

Findings – The model demonstrates accurate price prediction for parts with no assigned values while remaining within an acceptable price range. Additionally, outliers representing extreme pricing scenarios are successfully identified and predicted within the acceptable range.

Originality/value – This research bridges the gap between industry practice and academic research by demonstrating the effectiveness of machine learning for aftermarket pricing optimization. It offers an approach to address the challenges of pricing parts without assigned values and identifying outliers, potentially leading to increased revenue, sharper pricing tactics and a competitive advantage for aftermarket companies.

Keywords Aftermarket, Machine learning, Price prediction **Paper type** Research paper

1. Introduction

In today's competitive business environment, pricing decisions are critical to an organization's financial success (Kalpana *et al.*, 2022). Businesses aiming for competitiveness and financial success must improve their pricing strategies. An emerging tool, machine learning-based predictive pricing, uses historical data to quickly forecast optimal prices, assisting in setting fair rates and adapting to changing market conditions (Banerjee and Bandyopadhyay, 2020).

This approach employs machine learning techniques to decipher massive amounts of data, revealing intricate patterns and determining the most profitable price points based on customer preferences, market dynamics, and corporate goals (Gupta and Pathak, 2014; Mantrala *et al.*, 2006). This transformative methodology enables data-driven pricing decisions, granting organizations a competitive edge in pricing strategies.

Several studies have explored the application of machine learning for pricing optimization in various fields. For example, traditional methods were compared against machine learning methods such as Random Forest, Gradient Boosted Machines, and Deep Learners in the insurance industry, highlighting the effectiveness of Gradient Boosting Methods (Spedicato *et al.*, 2018). Our research builds on this by applying a Random Forest Regressor model in the aftermarket sector.

Dynamic pricing in e-commerce has been explored using Gradient Boosting Machines (GBMs), showing superior performance in capturing complex non-linear pricing patterns

© Mohit S. Sarode, Anil Kumar, Abhijit Prasad and Abhishek Shetty. Published in *Modern Supply Chain Research and Applications*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http:// creativecommons.org/licences/by/4.0/legalcode Modern Supply Chain Research and Applications

411

Received 6 October 2023 Revised 24 January 2024 13 June 2024 10 September 2024 Accepted 13 October 2024



Modern Supply Chain Research and Applications Vol. 6 No. 4, 2024 pp. 411-423 Emerald Publishing Limited 2631-3871 DOI 10.1108/MSCRA-10-2023-0042 MSCRA 6.4

412

(Youbi *et al.*, 2023). Similarly, machine learning algorithms like Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), and Support Vector Regression (SVR) have been used for stock price prediction, with SVR achieving the highest accuracy (Chen, 2020). Further studies have demonstrated the effectiveness of Random Forests and Artificial Neural Networks (ANN) in predicting stock closing prices (Vijh *et al.*, 2020).

The use of machine learning for property price prediction has shown that Random Forest and GBM algorithms perform well (Ho *et al.*, 2020). Moreover, the application of machine learning for daily commodity price prediction has been highlighted, with ANNs showing effectiveness but suggesting the incorporation of domain knowledge and feature engineering for improvement (Amin, 2020).

Complex price prediction tasks involve numerous factors influencing price changes. Traditional methods often struggle to account for these complexities, while machine learning has shown promise in optimizing prices (Indira *et al.*, 2023). Additionally, a model-based pricing framework for machine learning models has been proposed to address gaps in data market pricing, demonstrating high revenue potential and low runtime costs (Chen *et al.*, 2019).

Existing literature also investigates micro-marketing pricing strategies based on supermarket scanner data (Montgomery, 1997), forecasts stock prices using various models and data representation techniques (Patel *et al.*, 2015), investigates pricing strategies in B2B aftermarkets based on firm size, industry, and location (Gunaydan, 2023), and optimizes prices in dynamic markets with limited information (Dodin *et al.*, 2021).

1.1 Gap in literature

Although machine learning is widely used in pricing, it is still necessary to apply it in certain industries, such as the aftermarket, which handles replacement parts and components for goods that have already been made. The field of machine learning in pricing has been the subject of extensive research in the literature. Spedicato *et al.* (2018) conducted a comparison between machine learning models and traditional pricing techniques. These studies do not take into account the unique features of the aftermarket sector, where technical attributes of parts have a significant impact on pricing, nor do they concentrate on specific methodological procedures. This drawback may be seen in research by Youbi *et al.* (2023), although Youbi's technique works well in dynamic contexts, it is not directly applicable to the aftermarket sector since it ignores technical factors that have a big influence on pricing decisions. Finding irregular observations that point to mistakes, poor data quality, or unusual pricing trends is the first step in detecting anomalies in pricing data. Similarly, outlier detection identifies data points that differ significantly from the majority, frequently representing extreme price points or unique market circumstances.

Furthermore, research by Indira *et al.* (2023) emphasizes the importance of incorporating industry-specific data for accurate price prediction. It highlights the need for more specialized approaches in sectors with unique characteristics, such as the aftermarket. Indira's work points out that existing models often fail to account for technical specifications, which are critical in determining prices for aftermarket parts. Our study directly addresses these gaps by focusing on the unique needs of the aftermarket industry, where the pricing of components like fuel tanks is determined not only by market dynamics but also by technical attributes such as material quality, size, and durability.

By offering a thorough, domain-specific technique for price prediction and anomaly identification, our research expands on the fundamental stages presented by Spedicato *et al.* (2018). Our study uses a Random Forest Regression model trained on a dataset supplemented with technical specifications of aftermarket parts, in contrast to the generic techniques in the literature currently in publication. This focus on technical elements distinguishes our research from earlier studies and offers a more accurate and relevant approach for the aftermarket sector.

1.2 Research objectives

Ensuring customer satisfaction is critical for significant players in the trucking industry. By ensuring parts availability at the right time, location, and price without sacrificing quality, downtime can be reduced, and meaningful business opportunities can be retained. The aftermarket industry, where pricing decisions directly impact revenue and profitability, can benefit significantly from accurately forecasting optimal parts prices, striking a balance between attracting customers and maximizing profit margins.

The primary objective of this research is to address the challenges of predicting prices for aftermarket parts that lack assigned values and to detect anomalies and outliers by leveraging machine learning algorithms. In line with the findings of Indira *et al.* (2023), our approach emphasizes incorporating industry-specific data—specifically the technical features of parts—to improve the accuracy of price predictions. This study aims to develop a robust methodology that accurately forecasts optimal prices and identifies irregular pricing patterns that could indicate errors, poor data quality, or extreme market scenarios.

To achieve this, we compiled a comprehensive dataset focusing on fuel tanks due to their critical role in the trucking industry and their revenue potential. Given that existing methodologies, as highlighted by Spedicato *et al.* (2018) and Youbi *et al.* (2023), did not produce satisfactory results for our specific dataset, we adapted these techniques using our domain knowledge to develop a self-contained methodology that is readily applicable to real-world and industry-specific scenarios.

1.3 Research contributions and application

This study contributes to the existing body of literature by expanding on the methodologies presented by Spedicato *et al.* (2018), Indira *et al.* (2023), and Youbi *et al.* (2023), addressing their limitations by incorporating technical features that are critical for accurate pricing in the aftermarket sector. By tackling the research question of how machine learning algorithms can be utilized to predict the prices of aftermarket parts with no assigned values and detect pricing anomalies, this study provides a robust, data-driven framework that can be readily applied to real-world scenarios in the aftermarket industry. By addressing the gaps and limitations identified in previous studies, our research offers practical insights that enhance pricing strategies, profitability, and competitiveness in the aftermarket sector.

2. Methodology

The first step was to gather data. The dataset was then preprocessed, and key attributes were identified. Selected attributes were scaled, and different models were compared to determine which model was the best. The Random Forest Regressor model was chosen, and it was then trained and tested to predict the prices of the parts. Figure 1 depicts a block diagram of the entire



Source(s): Authors' own work

Figure 1. Block diagram of methodology

Modern Supply Chain Research and Applications methodology. Web scraping techniques, backend data, and part drawings were used to create the dataset. It had hundreds of thousands of data points. However, a few thousand-part numbers were chosen from the large dataset to focus on one specific truck assembly, the fuel tank.

To ensure data quality, attributes with insufficient data points or lacking relevant information were removed. Figure 2 shows a bar chart that was used to calculate the percentage of data covered in each of the remaining attributes. Attributes with less than 80% coverage in the data were removed. Furthermore, all attribute values were cleaned of noise and standardized to follow the same format. Numerical attributes with missing values were filled with the median value, whereas categorical attributes were filled with the most frequently occurring categories. The number of attributes was reduced by less than 60% as a result of this. Figure 3 depicts the missing value matrix for these attributes, where the white spaces between the matrix indicate the amount of missing data.

The frequency of the price points was observed using a histogram in the initial analysis. Based on this histogram, an average range of prices was determined, with prices outside of this range considered outliers. Figure 4 depicts the acceptable cost range and the outliers.

To investigate the impact on prices, feature engineering was carried out by calculating new attributes such as area, mass, and volume based on existing attributes such as diameter and length. To improve the model's performance, new attributes such as shape and vent were loaded and preprocessed into the dataset. Figure 5 shows a heat map created with the Seaborn library in Python to determine the significant attributes for price prediction.

Using the Random Forest Regressor model, various feature selection techniques were employed to evaluate the impact of attributes on price prediction. Variance Threshold, SelectKBest, and Recursive Feature Elimination with Cross-Validation (RFECV) were the 3 techniques used to evaluate the attributes.

The Variance Threshold method, represented by Equation (i), is effective at eliminating features with low variance, assuming they have a minimal contribution to the predictive model. During the selection process, the method automatically identifies and removes zero variance features.

$$selector = VarianceThreshold()$$
 (i)

Equation (ii) depicts the SelectKBest method, which uses a score function called $f_{regression}$ and a parameter k of 13. The parameter k is set to 13, indicating a preference to keep the top 13 features deemed most influential in predicting aftermarket part prices. This method assesses the statistical relationship between each feature and the target variable, ranking them according to their significance. The chosen parameters strike a balance between feature richness and model efficiency, taking into account the linear relationship between features and the target variable.

$$selector = SelectKBest(score_func = f_regression, k = 13)$$
 (ii)

The RFECV method, as shown in Equations (iii) and (iv), employs the Support Vector Regressor (SVR) as an estimator with a *linear kernel*. This method systematically evaluates feature relevance by recursively removing the least informative features. The use of a linear kernel corresponds to the assumed linear relationship between features and prices in our dataset. Step of 1 and cv of 10 were chosen as parameters to ensure thorough feature evaluation while maintaining computational efficiency.

$$estimator = SVR(kernel = linear)$$
(iii)

selector =
$$RFECV(estimator, step = 1, cv = 10)$$
 (iv)

Based on our dataset, RFECV elimination produced the most favorable results, and Table 1 shows the rankings of all RFECV-calculated attributes. This information was used to conduct trial and error tests on attributes to determine which contributed to the highest accuracy.

MSCRA 6,4









Figure 3. Attributes included after preprocessing



Figure 4. Price distribution including acceptable range and outliers

Recursive Feature Elimination with Cross-Validation (RFECV) technique outperforms SelectKBest and Variance Threshold because it can capture intricate relationships and dependencies between features. By removing less relevant features during cross-validation, RFECV excels at evaluating the collective impact of features in our dataset, which includes technical attributes of automobile parts. The method retains the most relevant features, improving accuracy and interpretability.

SelectKBest, which is efficient at selecting top features based on individual metrics, does not consider feature interactions. Variance Threshold, which focuses on variance within individual features, may miss important associations required for accurate pricing predictions.



Source(s): Authors' own work

Figure 5. Heat map correlating all the attributes

Features	Ranking	Features	Ranking	
Length	2	Anti-Siphon	16	
Thickness	6	GROSS WT	3	
Outer diameter	1	TOTAL SALES DEMAND	1	
Inner diameter	1	Area	1	
Material	9	Volume	11	
Finish	13	Mass	8	
Mounting location	12	SHAPE	10	
Fuel tank capacity	4	INSTA_HEAT	14	
Filler neck to end length	1	VENT	5	
Baffles included	7	Integral Fuel Tank	15	
Source(s): Authors' own work		-		

Table 1.	Ranking	based	on	RFECV
Tuble I.	Running	buscu	on	ILL LO V

As a result, RFECV's consideration of feature interactions and dependencies makes it more suitable for selecting impactful features and improving predictive model accuracy.

Feature scaling was used to ensure that the range of values was uniform. Numerical values were normalized, and categorical data was encoded using labels. Normalization was used to adjust numerical values such as length, diameter, area, mass, and so on to a standard scale, eliminating potential biases caused by different measurement units or scales. Categorical data, on the other hand, such as material, finish, shape, and so on, was encoded using labels,

MSCRA 6.4

allowing for the representation of qualitative information in a numerical format suitable for computational analysis.

Given the large volume of available data, the dataset was divided into a 70% training set and a 30% test set. The validation dataset was assigned a nominal amount (10%) of the training dataset. To estimate the performance of the models, the error metric Root Mean Squared Error (RMSE) was chosen.

Because of its comprehensibility, ability to capture prediction accuracy, and emphasis on penalizing larger errors, the Root Mean Square Error (RMSE) metric stands out as a superior choice in various modeling scenarios. The average magnitude of the errors between predicted and actual values is measured by RMSE, providing a straightforward understanding of how far off the model's predictions are from the true values. Furthermore, because of its squared nature, RMSE gives significant weight to larger errors, making the metric more sensitive to outliers or extreme deviations. Furthermore, RMSE is well-suited for regression-type problems in which the goal is to minimize prediction errors.

A critical aspect of our methodology was the evaluation of various regression models, such as Random Forest Regression, AdaBoost Regression, Bagging Regression, Support Vector Regression (SVR), and K-nearest Neighbor Regression. The accuracy on the validation set, mean RMSE score, and standard deviation were all considered when selecting a model. Notably, the Random Forest Regressor emerged as the superior choice and this section investigates the limitations of alternative models while explaining why the Random Forest Regressor was chosen.

2.1 Limitations of alternative regression models

Other algorithms considered for this task had limitations, but Random Forest emerged as the best option. AdaBoost's sensitivity to noisy data and outliers could pose a problem for our dataset, which may have quality variations. Bagging, while effective in reducing variance, may impair interpretability due to its ensemble nature. Furthermore, Support Vector Regression (SVR) necessitates meticulous hyperparameter tuning, which increases the time required for implementation. For high-dimensional datasets, K-Nearest Neighbors (KNN) can suffer from the "curse of dimensionality," which can have an impact on accuracy.

When dealing with large and high-dimensional datasets, Random Forest excels at predictive analysis, especially when there are complex interactions between features or nonlinear relationships with the target variable. It performs well even without extensive hyperparameter tuning and effectively handles noisy data. Random Forest appeared as the best option among the mentioned algorithms because the goal is to obtain robust predictions while dealing with diverse types of data and maintaining good interpretability, as shown in Table 2.

It is critical to accurately assess model performance during the development process. Validation techniques help with this by assessing how well the Random Forest Regressor (rfr) model generalizes to new data. In this study, the model was trained using the expected price range, and then its validation score was calculated. This step involved determining how well the model predicted prices on data that had not been previously trained on. However, more

Models	Accuracy	Mean RMSE score	Standard deviation
RandomForestRegressor	0.717507	61.800927	14.306699
AdaBoostRegressor	0.451401	70.588170	11.112890
BaggingRegressor	0.673993	64.175979	14.116304
SVR	0.464940	113.950125	16.046542
KNeighborsRegressor	0.622910	100.102034	16.436865
Source(s): Authors' own work	ζ.		

Table 2. Comparison of different models

sophisticated techniques such as Grid Search Cross-Validation (GridSearchCV) and RandomizedSeach Cross-Validation (RandomizedSearchCV) were used to improve the model's performance by tuning the hyperparameters even further.

The code snippet, Equation (v), encapsulates the GridSearchCV process. The GridSearchCV systematically explores a predefined grid of hyperparameters for the Random Forest Regressor, assessing the model's performance across various combinations through cross-validation. The selected parameters, specified in the *param_grid* variable, include a range of values for tuning the Random Forest Regressor, such as the number of estimators and the tree depth. The choice of these parameters aligns with the need to strike a balance between model complexity and performance, ensuring optimal generalization to unseen data. The scoring metric, set as *neg_mean_squared_error*, aims to minimize the mean squared error during cross-validation, guiding the grid search toward configurations that yield the most accurate predictions.

$$grid_search = GridSearchCV(rfr, param_grid = param_grid, cv$$

$$= 10$$
, scoring = neg_mean_squared_error) (v)

The code snippet Equation (vi) encapsulates the RandomizedSearchCV process. The use of RandomizedSearchCV over GridSearchCV allows for a more randomized exploration of the hyperparameter space, making it computationally less intensive while still yielding robust results.

random_search = RandomizedSearchSearchCV(rfr, param_grid = param_grid, cv

$$= 10$$
, scoring = neg_mean_squared_error) (vi)

The optimal hyperparameter values for the Random Forest Regressor model were determined using these techniques to be a tree depth of 15 and an estimator size of 200. Following that, various permutations and combinations were investigated based on the feature importance determined by recursive feature elimination. Table 3 highlights the final list of attributes that provided the best accuracy of 76.14% on the Random Forest Regressor model, as shown in Table 4.

3. Results and discussion

The Random Forest Regressor model, chosen for its superior performance, predicted aftermarket part prices with an impressive accuracy rate of 76.14%. This robust predictive capability stands out in the context of the dataset, demonstrating its effectiveness in dealing with the complex interactions and nonlinear relationships inherent in automotive part pricing. The accuracy of the model in capturing underlying patterns and relationships was further

Sl. No.	Attribute	Sl. No.	Attribute
1	Length	9	Thickness
2	Volume	10	Baffles Included
3	Mass	11	TOTAL_SALES_DEMAND
4	SHAPE	12	Outer Diameter
5	Fuel Tank Capacity	13	Inner Diameter
6	Finish	14	Material
7	Area	15	VENT
8	GROSS_WT		
Source(s): Au	thors' own work		

Table 3. Final list of attributes selected

Modern Supply Chain Research and Applications

MSCRA	Table 4. Best output given by random forest regressor model			
6,4	Models	RandomForestRegressor		
	Accuracy Mean RMSE score	0.761468 54.192732		
	Standard deviation	11.211734		
420	Source(s): Authors' own work			

validated by comparing predicted prices to actual prices, as shown in Figure 6. The scatter plot shows a strong correlation between predicted and actual prices, indicating that the model can provide accurate estimates.

The Random Forest Regressor was found to be the most accurate of the regression models tested, including AdaBoost, Bagging, Support Vector Regression (SVR), and K-nearest Neighbor Regression. The Random Forest Regressor was chosen because of its resilience in handling diverse data types, avoidance of overfitting, and effective management of noisy data. While each alternative model has advantages, they all have limitations that make them unsuitable for the dataset's specific characteristics.

The random forest regression model was also tested for its ability to detect anomalies. The model predicted the price of each data point at zero cost at first. Figure 7 depicts a scatter plot illustrating the relationship between predicted prices and zero-cost parts. Notably, the predicted price range for all zero-cost components is within the acceptable range.

The machine learning insights gained from feature importance analysis and Recursive Feature Elimination with Cross-Validation (RFECV) add significantly to existing aftermarket pricing knowledge. The RFECV technique, which has been identified as superior in selecting relevant features, excels in evaluating complex relationships and dependencies among features in the dataset. This understanding is critical for the automotive industry, where feature interactions play a critical role in pricing decisions. RFECV's ability to eliminate less informative features iteratively improves model accuracy and interpretability, providing a more nuanced understanding of how technical attributes influence pricing.



Source(s): Authors' own work

Figure 6. Scatter plot of actual vs predicted prices



Modern Supply Chain Research and Applications

421

Figure 7. Scatter plot of zero priced parts vs predicted price

As shown in Figure 8, the methodology successfully identified outliers representing unusual pricing scenarios. The model correctly predicted prices for these outliers within the acceptable range, demonstrating its ability to deal with unusual market conditions. When faced with atypical pricing situations, this anomaly detection capability is critical for aftermarket businesses, allowing them to mitigate potential revenue loss and make informed decisions.



Figure 8. Scatter plot of actual prices of lower and higher outliers vs predicted prices

MSCRA 4. Conclusion

6,4

422

In conclusion, the application of machine learning techniques in the aftermarket sector yielded key findings with significant implications for pricing strategies in this study. The Random Forest Regressor model predicted aftermarket part prices with an impressive 76.14% accuracy, providing a robust tool for businesses navigating the complex landscape of pricing decisions. The methodology's success in feature selection via RFECV and anomaly detection increases its practical applicability even further.

The identified technical pricing attributes, as well as the model's ability to handle outliers and unique market circumstances, add valuable insights to the existing knowledge gap in aftermarket pricing strategies. The findings of the study provide decision-makers with a more nuanced understanding of the factors influencing pricing decisions, as well as a reliable framework for optimizing revenue and competitiveness.

4.1 Limitations and further work

While the study has made significant progress in leveraging machine learning for pricing optimization in the aftermarket sector, it is important to recognize some limitations that may impact the generalizability and robustness of the findings. One limitation is the dataset's narrow focus, which was primarily on fuel tanks within a single truck assembly. This may necessitate additional research to validate and extend the approach to a more diverse range of components.

Furthermore, while the Random Forest Regressor model produced noteworthy accuracy, it may be subject to overfitting or may not be the best choice for all aftermarket pricing scenarios. Model performance can vary depending on the characteristics of the data, so experimenting with alternative models or ensemble approaches may provide a more robust understanding of the predictive capabilities across diverse datasets. Ensemble methods, such as stacking or blending, could be investigated to capitalize on the strengths of multiple models, reducing the risk of relying solely on one algorithm.

Moreover, the anomalies discovered during the study, which represent irregular pricing scenarios, can be used for more than just identification. These anomalies can help refine and improve the machine-learning code. By assessing and understanding the underlying causes of these anomalies, the model can be adjusted and improved, resulting in a more refined and accurate predictive tool. An iterative process of detecting anomalies, refining the model, and relearning from the updated dataset can help to create a constantly improving machine learning system for aftermarket pricing.

In the future, potential research directions in this domain could include the incorporation of advanced machine-learning models or ensemble techniques to improve predictive accuracy. Furthermore, researching the dynamic nature of market conditions and how machine learning can adapt to real-time changes in the aftermarket sector is an exciting avenue to pursue. Furthermore, investigating the applicability of the methodology to other technical parts of the field could open new avenues for research and practical applications. Continuous technological and data analysis tool advancements provide a rich landscape for future endeavors, ensuring that the aftermarket sector remains at the forefront of innovative pricing strategies.

References

- Amin, M. (2020), "Predicting price of daily commodities using machine learning", *Conference: 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies*, Vol. 3 No. 11, pp. 1-6, doi: 10.1109/3ICT51146.2020.9312012.
- Banerjee, P. and Bandyopadhyay, S. (2020), "Prediction of price of product by machine leaning method through web based system", *International Journal of Trend in Research and Development*, Vol. 7, pp. 2394-9333.

Chen,	L. (2020),	"Using machine	learning	algorithms	on predictio	on of stocl	k price", Journal of	
	Modeling of	and Optimization	, Vol. 12	No. 2, pp.	84-99, doi:	10.32732/	/jmo.2020.12.2.84.	

- Chen, L., Koutris, P. and Kumar, A. (2019), "Towards model-based pricing for machine learning in a data marketplace", *International Conference on Management of Data*. doi: 10.1145/ 3299869.3300078.
- Dodin, P., Xiao, J., Adulyasak, Y., Alamdari, N.E., Gauthier, L., Grangier, P., Lemaitre, P. and Hamilton, W. (2021), "Bombardier aftermarket demand forecast with machine learning", *INFORMS Journal on Applied Analytics*, Vol. 53, doi: 10.2139/ssrn.3957452.
- Gunaydan, I. (2023), "An analysis of pricing strategies within business-to-business companies in the aftermarket domain", *The Impact of Company Size, Industry, and Location on Pricing Strategies*, available at: https://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-61015
- Gupta, R. and Pathak, C. (2014), "A machine learning framework for predicting purchase by online customers based on dynamic pricing", *Procedia Computer Science*, Vol. 36, pp. 599-605, doi: 10.1016/j.procs.2014.09.060.
- Ho, W., Tang, B. and Wong, S. (2020), "Predicting property prices with machine learning algorithms", *Journal of Property Research*, Vol. 38, pp. 1-23, doi: 10.1080/09599916.2020.1832558.
- Indira, L., Suchetan, T.C.K., Iqbal, S.M., Rohith, A. and Shinde, P. (2023), "Pricing optimization using machine learning", *International Research Journal of Engineering and Technology* (*IRJET*), Vol. 10.
- Kalpana, G., Durga, A.K., Reddy, A. and Karuna, G. (2022), "Predicting the price of pre-owned cars using machine learning and data science", *International Journal of Recent Advances in Engineering and Technology (IJRAET)*, Vol. 10 No. 7, pp. 1027-1032, doi: 10.22214/ ijraset.2022.45469.
- Mantrala, M., Seetharaman, P.B., Kaul, R., Gopalakrishna, S. and Stam, A. (2006), "Optimal pricing strategies for an automotive aftermarket retailer", *Journal of Marketing Research*, Vol. 43 No. 4, pp. 588-604, doi: 10.1509/jmkr.43.4.588.
- Montgomery, A.L. (1997), "Creating micro-marketing pricing strategies using supermarket scanner data", *Marketing Science*, Vol. 16 No. 4, pp. 315-337, doi: 10.1287/mksc.16.4.315.
- Patel, J., Shah, S., Kotecha, K. and Thakkar, P. (2015), "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques", *Expert Systems with Applications*, Vol. 42 No. 1, pp. 259-268, doi: 10.1016/j.eswa.2014.07.040.
- Spedicato, G.A., Dutang, C. and Petrini, L. (2018), "Machine learning methods to perform pricing optimization", *A Comparison with Standard GLMs*, Vol. 12, pp. 69-89.
- Vijh, M., Chandola, D., Tikkiwal, V. and Kumar, A. (2020), "Stock closing price prediction using machine learning techniques", *Procedia Computer Science*, Vol. 167, pp. 599-606, doi: 10.1016/ j.procs.2020.03.326.
- Youbi, R.E., Messaoudi, F. and Loukili, M. (2023), "Machine learning-driven dynamic pricing strategies in E-commerce", *Conference*, pp. 1-5, doi: 10.1109/ICICS60529.2023.10330541.

Corresponding author

Mohit S. Sarode can be contacted at: mohit.s_sarode@daimlertruck.com

Modern Supply Chain Research and Applications