Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness

Mohammad Anisur Rahman¹, Chinmoy Modak², Md Abu Sufian Mozumder³, Mohammed Nazmul Islam Miah⁴, Mehedi Hasan⁵, Md Murshid Reja Sweet⁶, Syeda Farjana Farabi⁷, Md Zikar Hossan⁸, Mani Prabha⁹ and Mahfuz Alam¹⁰

¹Department of Marketing & Business Analytics, Texas A&M University-Commerce, USA
²Information Technology, Craven Community College, USA
³College of Business, Westcliff University, Irvine, California, USA
⁴Department of Management Science and Quantitative Methods, Gannon University, USA
⁵Master of Science, Management- Business Analytics, St. Francis College, USA
⁶Doctor of Business Administration, Westcliff University, Irvine, California
⁷Master of Business Administration in Management Information System, International American University Los Angeles
⁸Master of Business Administration, International American University, Los Angeles, California

Corresponding Author: Md Abu Sufian Mozumder, E-mail: m.mozumder.158@westcliff.edu

ABSTRACT
Retail price optimization is essential for maximizing profitability and maintaining competitiveness in today's dynamic retail landscape. This study addresses retail price optimization as a regression problem, utilizing machine learning models to predict optimal price points for products. Leveraging factors such as product attributes, competitor pricing dynamics, and customer behaviors, regression analysis provides a structured approach to understanding the intricate relationships between variables. Among various regression techniques, the Random Forest Regressor emerges as a potent strategy, offering robustness and versatility in tackling complex pricing tasks. Results indicate that Random Forest outperforms Decision Tree and Logistic Regression models regarding accuracy, precision, recall, and overall predictive performance. With Random Forest achieving an accuracy of 94%, it demonstrates superior capability in capturing complex data patterns and making accurate predictions of retail prices. By leveraging advanced analytics and machine learning techniques, retailers can optimize pricing strategies, maximize profits, and maintain competitiveness in the market. This study underscores the importance of continuously analyzing and refining pricing strategies to gain a competitive edge in the retail landscape.

KEYWORDS
Retail Price Optimization; Machine Learning; Profitability; Competitiveness

ARTICLE INFORMATION
ACCEPTED: 23 May 2024
PUBLISHED: 23 May 2024
DOI: 10.32996/jbms.2024.6.3.11

1. Introduction
In today's dynamic retail landscape, pricing optimization stands as a cornerstone for maximizing profitability and maintaining competitiveness. Retailers are continually seeking innovative approaches to accurately forecast demand and set optimal prices for their products. With the advent of advanced analytics and machine learning techniques, the realm of retail price optimization has witnessed significant advancements, enabling retailers to leverage data-driven insights for informed decision-making. This article delves into the critical aspects of retail price optimization, focusing on the integration of machine learning models to predict optimal price points for retail products.

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.
Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness

Drawing from recent research studies and methodologies proposed by Ding H (2022) and Jewel et al. (2024), this article explores the multifaceted nature of retail price optimization and its implications for business management. The proposed method by Ding H (2022) harnesses the power of the long short-term memory (LSTM) deep learning algorithm to address the complexities of commodity price forecasting. By incorporating external factors and leveraging historical data, this method offers a comprehensive framework for accurate price prediction, promising higher accuracy and better stability compared to traditional approaches.

Additionally, the comparative analysis conducted by Jewel et al. (2024) sheds light on the performance of various machine learning models, such as Long Short-Term Memory (LSTM) and LightGBM (LGBM), in retail sales demand forecasting. Through meticulous evaluation and optimization, LGBM emerges as the preferred model for capturing and predicting sales patterns, emphasizing the significance of model selection in retail sales demand forecasting.

Furthermore, this article delves into the methodology of retail price optimization, elucidating the strategic processes involved in determining optimal prices to maximize profitability. Techniques such as competitor analysis, customer segmentation, and price testing are explored in detail, highlighting their role in shaping pricing strategies and driving sales.

Moreover, the article provides insights into dataset collection and processing for retail price optimization, elucidating the key features and metrics considered in analyzing retail products and sales performance. Through effective dataset preprocessing techniques, retailers can derive actionable insights and develop predictive models for forecasting sales and optimizing prices.

Lastly, the article presents the results of the study on retail price optimization, emphasizing the superiority of Random Forest Regressor in predicting optimal retail prices compared to Decision Tree and Logistic Regression models. With Random Forest achieving the highest accuracy and overall predictive performance, retailers can leverage this model to enhance profitability and maintain a competitive edge in the retail market.

Overall, this article aims to provide a comprehensive overview of retail price optimization, integrating insights from recent research studies and methodologies to offer valuable insights for retailers seeking to optimize their pricing strategies and maximize profitability in today's dynamic retail landscape.

2. Literature Review
The main concept of the proposed (ding H 2022) method for commodity price prediction lies in its utilization of the long short-term memory (LSTM) deep learning algorithm to address the complexities of price forecasting. Recognizing that commodity prices are influenced by various factors beyond just time series data, the method incorporates external factors through a multifactor approach. By leveraging the memory capabilities of LSTM to analyze historical data, coupled with the integration of external factors via a fully connected layer, the method offers a comprehensive framework for price prediction. This approach not only considers the intrinsic time-dependent patterns in price data but also accounts for the impact of external variables, such as market trends, supply-demand dynamics, and economic indicators. Moreover, the method employs a strategy of leveraging data from similar commodities to enhance the training set, thereby improving the accuracy and stability of the price prediction model. By combining these elements, the proposed method provides a novel solution to the challenge of commodity price forecasting, offering higher accuracy and better stability compared to traditional methods like the backpropagation (BP) neural network. Ultimately, this method offers a promising avenue for macroeconomic decision-making and micro-level management by providing reliable insights into future price movements of commodities.

This in article the author (Jewel er al 2024) revolves around a comparative analysis of sales forecasting models, specifically Long Short-Term Memory (LSTM) and LightGBM (LGBM), using retail sales data. Through a systematic approach involving memory optimization, feature engineering, and parameter tuning, the study meticulously evaluates the performance of both models. By employing evaluation metrics such as RMSE, MAE, WMAPE, and WRMSEE, the study concludes that LGBM consistently outperforms LSTM in accurately capturing and predicting sales patterns. This finding underscores the significance of model selection in retail sales demand forecasting, with LGBM emerging as the preferred choice based on its superior performance. The study’s contribution lies in providing practical insights into machine learning applications for retail sales forecasting, thereby highlighting LGBM as an effective and reliable model for this purpose.

3. Methodology
Retail price optimization is the strategic process through which retailers determine the most effective prices for their products to maximize profitability. This involves identifying the ideal price point that not only attracts customers and drives sales but also ensures optimal profit margins. To achieve this, retailers employ a range of techniques, including competitor analysis, customer segmentation, and price testing. Competitor analysis entails monitoring the pricing strategies of rival businesses offering similar products and adjusting prices accordingly to remain competitive. Customer segmentation involves categorizing shoppers into distinct groups based on their purchasing habits and tailoring prices to suit each segment. Price testing is another crucial method
wherein different price points are trialed to identify the most profitable option. By leveraging these techniques, retailers can enhance their profitability and strengthen their position in the market. Successful price optimization requires a deep understanding of consumer behavior, market dynamics, and pricing tactics, as well as the capability to gather and analyze data related to sales and pricing trends. Retailers who effectively optimize their prices stand to gain a significant competitive edge, while also ensuring that customers receive fair and reasonable pricing for the products they desire.

### 3.1 Dataset collection and processing
In the process of collecting and preprocessing datasets for retail price optimization, various features of existing products are considered. These features encompass not only competitor prices and ratings but also temporal factors reflecting product sales over time. Analyzing competitors’ pricing strategies is a crucial aspect of this process, as it helps retailers identify opportunities for competitive pricing adjustments based on their positioning and overall strategy. By monitoring and benchmarking against competitor prices, retailers can determine whether to price their products below or above the competition, aligning with their specific objectives. Additionally, temporal features such as sales trends provide valuable insights into product demand and may require the application of demand forecasting techniques. Metrics like total prices, customer count, and product quantity sold are also pivotal in determining the optimal price point where retailers can maximize profits. These datasets serve as foundational elements for regression techniques aimed at identifying the most effective pricing strategies.

The provided dataset encompasses various features related to retail products and their sales performance, along with additional contextual factors such as competitor prices and ratings. Here’s a comprehensive discussion on the dataset:

1. **Product Features**: The dataset includes essential attributes of products, such as quantity sold, total price, unit price, product name length, product description length, product photos quantity, product weight, and product score. These features offer insights into the characteristics and perceived value of the products being sold.
2. **Customer Metrics**: Customer-related metrics like the number of customers indicate the level of consumer interest and engagement with the products. Understanding customer behavior is crucial for devising effective marketing and pricing strategies.
3. **Temporal Factors**: Time-related variables such as weekday, weekend, holiday, month, and year provide temporal context to the sales data. Analyzing sales patterns across different time periods can reveal seasonality trends and help retailers anticipate demand fluctuations.
4. **Competitor Analysis**: The dataset includes competitor-related metrics such as competitor prices (comp_1, comp_2, comp_3), competitor product scores (ps1, ps2, ps3), and competitor freight prices (fp1, fp2, fp3). Monitoring competitor pricing strategies allows retailers to adjust their own prices strategically to remain competitive in the market.
5. **Derived Metrics**: Additionally, the dataset features derived metrics like lag_price, which may indicate the historical pricing trends of the products. Understanding how prices have evolved over time can inform pricing decisions and strategies.
6. **Volume and Revenue Metrics**: Metrics like volume and total revenue provide an overview of the overall sales performance of the products. These metrics are essential for assessing the financial health of the retail business and identifying opportunities for revenue growth.

Overall, this dataset offers a comprehensive view of various factors influencing retail sales and pricing decisions. By analyzing these features collectively, retailers can gain valuable insights into customer behavior, market dynamics, and competitor positioning, enabling them to optimize their pricing strategies and maximize profitability. Additionally, machine learning models and statistical techniques can be applied to this dataset to develop predictive models for forecasting sales and optimizing prices further.

Furthermore, efficient dataset preprocessing techniques play a vital role in preparing the data for analysis. Apart from competitor pricing and temporal sales data, customer information can be leveraged through clustering methods to segment customers into distinct groups. This segmentation facilitates targeted marketing strategies tailored to specific audience preferences for individual products. By organizing and refining the dataset through preprocessing steps, retailers can ensure that the data fed into regression models is accurate, relevant, and conducive to deriving actionable insights. Overall, effective dataset collection and preprocessing lay the groundwork for successful retail price optimization, enabling retailers to make informed pricing decisions and ultimately enhance their profitability.

This bar chart 1 snippet efficiently creates an interactive visualizing the distribution of total prices in the dataset, providing valuable insights into the pricing structure and variability of the retail products.
This scatter plot visualizes the relationship between the quantity of products sold and the total price of those products. By plotting these variables against each other, retailers can gain insights into how changes in quantity impact total revenue. The trendline, generated using OLS regression, provides additional information about the overall trend in the data. If the trendline slopes upward, it indicates a positive correlation between quantity and total price, suggesting that increasing the quantity sold leads to higher total revenue. Conversely, a downward-sloping trendline suggests a negative correlation, indicating that increasing the quantity sold may lead to lower total revenue. Analyzing this plot can help retailers make informed decisions about pricing strategies, inventory management, and sales forecasting.
4. Result

In the realm of retail price optimization, addressing the challenge as a regression problem offers a structured approach to predict the ideal price point for a given product. This entails leveraging a variety of factors including product attributes, competitor pricing dynamics, and customer behaviors. Regression, a fundamental statistical technique, serves as the bedrock for this endeavor, allowing for the systematic analysis of relationships between multiple variables. Essentially, regression seeks to establish a predictive model that captures the intricate interplay among these variables, facilitating the estimation of a dependent variable based on the values of independent variables. The essence of regression lies in identifying the line or curve that best encapsulates the data, enabling accurate predictions for new data points.

In this context, the utilization of Random Forest Regressor emerges as a potent strategy. This regression technique harnesses the power of the Bagging method within the domain of Ensemble learning, specifically leveraging the Random Forest algorithm. By aggregating the predictions of multiple decision trees, Random Forest not only mitigates overfitting but also enhances the overall predictive performance. Its robustness and versatility make it a favored choice for tackling complex regression tasks in retail price optimization. Through an amalgamation of data-driven insights and predictive modeling, Random Forest Regressor equips retailers with the tools to navigate the dynamic landscape of pricing strategies, empowering them to make informed decisions that optimize profitability and enhance competitiveness in the market.
Overall, this chart 4 provides a visual assessment of how well a predictive model performs in predicting retail prices, comparing the predicted values with the actual observed values. It helps in evaluating the model’s accuracy and identifying any discrepancies or patterns in the predictions.

Retail price optimization plays a pivotal role in the profitability of retailers, as it directly impacts their bottom line. Leveraging advanced analytics and machine learning techniques, retailers can enhance their pricing strategies to maximize profits while maintaining competitiveness in the market. The results obtained from various models, including Random Forest, Decision Tree, and Logistic Regression, provide valuable insights into the effectiveness of different approaches to price optimization. With Random Forest achieving the highest accuracy of 94%, followed by Decision Tree at 88% and Logistic Regression at 80%, retailers can utilize these models to make informed decisions about pricing strategies. By understanding customer behavior, market trends, and competitive dynamics, retailers can fine-tune their pricing models to strike the optimal balance between profitability and customer satisfaction. Through continuous analysis and refinement of pricing strategies, retailers can gain a competitive edge in the dynamic retail landscape while maximizing their revenue potential.

In the table 1 as we can see Random Forest outperformed compared to other regression model. Based on the results obtained from our study on Retail Price Optimization using various machine learning models, it is evident that Random Forest outperforms both Decision Tree and Logistic Regression in terms of accuracy, precision, recall, and F1 score. With an impressive accuracy of 94%, Random Forest demonstrates superior performance in predicting optimal retail prices compared to the other models. This high accuracy indicates that Random Forest is highly effective in capturing the complex relationships and patterns present in the data, leading to more accurate predictions of retail prices.

Moreover, Random Forest also exhibits high precision (93%) and recall (93%), indicating its ability to minimize false positives and false negatives, respectively. This balanced performance is crucial in the context of retail price optimization, as it ensures that the model not only identifies optimal price points accurately but also avoids overpricing or underpricing products, thereby maximizing profitability.

In contrast, Decision Tree, while achieving a respectable accuracy of 88%, falls short in terms of precision (86%) and recall (87%) compared to Random Forest. This suggests that Decision Tree may be more prone to misclassifying retail prices, leading to suboptimal pricing decisions and potentially lower profitability for retailers.

<table>
<thead>
<tr>
<th>Models</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>94%</td>
<td>93%</td>
<td>93%</td>
<td>92%</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>88%</td>
<td>86%</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>80%</td>
<td>78%</td>
<td>79%</td>
<td>79%</td>
</tr>
</tbody>
</table>

Similarly, as we can see in the chart 5 Logistic Regression, with an accuracy of 80%, demonstrates the lowest performance among the three models. Although Logistic Regression provides a reasonable level of precision (78%) and recall (79%), its overall predictive power is not as robust as Random Forest. As a result, Logistic Regression may not be as reliable in accurately determining optimal retail prices, potentially leading to missed revenue opportunities or reduced competitiveness in the market.
Chart 5: Performance of different machine learning algorithm

On the results of our study, Random Forest emerges as the preferred model for Retail Price Optimization, offering superior accuracy, precision, recall, and overall predictive performance compared to Decision Tree and Logistic Regression. By leveraging Random Forest for pricing decisions, retailers can enhance their profitability, optimize inventory turnover, and maintain a competitive edge in the dynamic retail landscape.

5. Conclusion and Discussion

Through the lens of regression analysis, we explored the multifaceted factors influencing optimal pricing strategies, ranging from product attributes to competitor dynamics and customer behaviors. Our investigation revealed that Random Forest Regressor stands out as a potent tool for predicting optimal retail prices, surpassing Decision Tree and Logistic Regression models in terms of accuracy, precision, recall, and overall predictive performance.

The superiority of Random Forest lies in its ability to capture complex data patterns and make accurate predictions, empowering retailers to navigate the competitive landscape with confidence. By embracing Random Forest and other machine learning techniques, retailers can refine their pricing strategies, maximize profits, and maintain a competitive edge in the market.

However, the journey towards effective retail price optimization doesn't end here. Continuous analysis, refinement, and adaptation of pricing strategies are essential to stay ahead in the ever-evolving retail landscape. As new technologies emerge and consumer preferences evolve, retailers must remain agile and responsive, leveraging data-driven insights to drive informed decision-making.

In conclusion, this study underscores the critical role of advanced analytics and machine learning in retail price optimization. By harnessing the power of predictive modeling, retailers can unlock new avenues for profitability, driving sustainable growth and competitiveness in an increasingly dynamic retail environment.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Authors’ Contributions: All the authors read and approved the final manuscript.

Reference


Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness


