# APPLICATION OF MACHINE LEARNING ALGORITHMS FOR DAILY SALES FORECASTING IN A BRAZILIAN BREWING INDUSTRY

## APLICAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA PREVISÃO DE VENDAS DIÁRIAS DE UMA INDÚSTRIA CERVEJEIRA BRASILEIRA

## APLICACIÓN DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO EN LA PREDICCIÓN DE VENTAS DIARIAS DE UNA INDUSTRIA CERVECERA BRASILEÑA

https://doi.org/10.56238/sevened2026.008-124

**Rodrigo Ribeiro Salgado Cézar[1], Allana Eduarda Braga Iquegami[2], Nelson Marinelli Filho[3], Gil Eduardo Guimarães[4], Geraldo Nunes Corrêa[5]**

**ABSTRACT**

Sales forecasting is a strategic element for the brewing industry, a sector characterized by high seasonality and significant fluctuations associated with weekends, holidays, and festive events. Traditional statistical models, although widely applied, present limitations in capturing nonlinear relationships and complex patterns inherent to demand time series. In this context, this study aims to evaluate the application of machine learning algorithms for daily sales forecasting in a Brazilian brewing industry. To this end, a historical dataset of daily sales covering the period from May 2024 to May 2025 was used, adopting a temporal split in which data from 2024 were employed for model training and data from 2025 were reserved for testing. Temporal feature engineering was performed, including calendar variables, holiday indicators, lagged variables, and moving statistics. Four tree-based algorithms were evaluated: Decision Tree, Random Forest, XGBoost, and LightGBM, using the coefficient of determination ($R^2$), mean absolute error (MAE), and root mean squared error (RMSE) as performance metrics. The results indicate that ensemble models significantly outperform the single decision tree, with XGBoost achieving the best predictive performance, explaining approximately 89.9% of the variance in daily sales. It is concluded that the application of machine learning algorithms combined with temporal feature engineering constitutes an

[1] Undergraduate student in Information Systems. Universidade do Estado de Minas Gerais (UEMG).
E-mail: rodrigo.2411093@discente.uemg.br Orcid: https://orcid.org/0009-0008-1119-3369
Lattes: http://lattes.cnpq.br/3527280447737544
[2] Undergraduate student in Information Systems. Universidade do Estado de Minas Gerais (UEMG).
E-mail: allana.2511001807@discente.uemg.br Orcid: https://orcid.org/0009-0005-5208-8758
Lattes: http://lattes.cnpq.br/6255602665290316
[3] Dr. in Mechanical Engineering. Instituto de Tecnologia e Educação Galileo da Amazônia (ITEGAM).
E-mail: nelson.marinelli@itegam.org.br Orcid: https://orcid.org/0009-0005-4362-0132
Lattes: http://lattes.cnpq.br/9077460245811758
[4] Dr. in Mechanical Engineering. Instituto de Tecnologia e Educação Galileo da Amazônia (ITEGAM).
E-mail: gil.guimaraes@itegam.org.br Orcid: https://orcid.org/0000-0003-2800-4620
Lattes: http://lattes.cnpq.br/8142709905105377
[5] Dr. in Production Engineering. Universidade do Estado de Minas Gerais (UEMG).
E-mail: geraldo.correa@uemg.br Orcid: https://orcid.org/0000-0001-5477-6953
Lattes: http://lattes.cnpq.br/6713336620979877

effective approach for demand forecasting in the brewing sector, providing relevant support for production, inventory, and logistics planning.

**Keywords:** Demand Forecasting. Machine Learning. Time Series. Brewing Industry.

**RESUMO**
A previsão de vendas constitui um elemento estratégico para a indústria cervejeira, setor caracterizado por elevada sazonalidade e variações significativas associadas a fins de semana, feriados e eventos festivos. Modelos estatísticos tradicionais, embora amplamente utilizados, apresentam limitações na captura de relações não lineares e padrões complexos presentes em séries temporais de demanda. Diante desse contexto, o presente estudo objetiva avaliar a aplicação de algoritmos de aprendizado de máquina na previsão de vendas diárias de uma indústria cervejeira brasileira. Para tanto, utilizou-se um conjunto de dados históricos de vendas diárias referente ao período de maio de 2024 a maio de 2025, adotando-se uma divisão temporal, na qual os dados de 2024 foram empregados para treinamento dos modelos e os de 2025 para teste. Procedeu-se à engenharia de atributos temporais, incluindo variáveis de calendário, indicadores de feriados, defasagens e estatísticas móveis. Foram avaliados quatro algoritmos baseados em árvores de decisão: Decision Tree, Random Forest, XGBoost e LightGBM, utilizando como métricas de desempenho o coeficiente de determinação ($R^2$), o erro absoluto médio (MAE) e a raiz do erro quadrático médio (RMSE). Os resultados evidenciam que os modelos de ensemble superam significativamente a árvore de decisão isolada, com destaque para o XGBoost, que apresentou melhor desempenho preditivo, explicando aproximadamente 89,9% da variância das vendas diárias. Conclui-se que a aplicação de algoritmos de aprendizado de máquina, aliada à engenharia de atributos temporais, constitui uma abordagem eficaz para a previsão de demanda no setor cervejeiro, oferecendo subsídios relevantes para o planejamento de produção, estoques e logística.

**Palavras-chave:** Previsão de Demanda. Aprendizado de Máquina. Séries Temporais. Indústria Cervejeira.

**RESUMEN**
La previsión de ventas constituye un elemento estratégico para la industria cervecera, un sector caracterizado por una elevada estacionalidad y por variaciones significativas asociadas a fines de semana, feriados y eventos festivos. Los modelos estadísticos tradicionales, aunque ampliamente utilizados, presentan limitaciones para capturar relaciones no lineales y patrones complejos presentes en las series temporales de demanda. En este contexto, el presente estudio tiene como objetivo evaluar la aplicación de algoritmos de aprendizaje automático en la predicción de ventas diarias de una industria cervecera brasileña. Para ello, se utilizó un conjunto de datos históricos de ventas diarias correspondiente al período de mayo de 2024 a mayo de 2025, adoptándose una división temporal en la que los datos de 2024 se emplearon para el entrenamiento de los modelos y los de 2025 se reservaron para la prueba. Se llevó a cabo ingeniería de características temporales, incluyendo variables de calendario, indicadores de feriados, variables rezagadas y estadísticas móviles. Se evaluaron cuatro algoritmos basados en árboles de decisión: Decision Tree, Random Forest, XGBoost y LightGBM, utilizando como métricas de desempeño el coeficiente de determinación ($R^2$), el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). Los resultados evidencian que los modelos de tipo ensemble superan significativamente al árbol de decisión individual, destacándose el XGBoost, que presentó el mejor desempeño predictivo, explicando aproximadamente el 89,9% de la variancia de las ventas diarias. Se concluye que la aplicación de algoritmos de aprendizaje automático, combinada con la ingeniería de características temporales, constituye un

enfoque eficaz para la predicción de la demanda en el sector cervecero, aportando insumos relevantes para la planificación de la producción, los inventarios y la logística.

**Palabras clave:** Predicción de la Demanda. Aprendizaje Automático. Series Temporales. Industria Cervecera.

## 1 INTRODUCTION

Sales forecasting plays a central role in the operational and strategic planning of organizations, especially in sectors characterized by high seasonality and strong variability in demand. In the case of the brewing industry, consumption presents cyclical patterns associated with weekends, national holidays, cultural events and climatic conditions, which imposes significant challenges to the management of production, inventories and logistics (ALVES; SILVA, 2023; DE OLIVEIRA DIAS; FALCONI, 2018). Errors in forecasting can result in both stockouts and excess stored products, with direct impacts on operating costs and the level of service to the market.

Traditionally, demand forecasting has been performed using classical time-series statistical models, such as the autoregressive integrated moving average models (ARIMA). Although these models have a solid theoretical foundation and broad application, their effectiveness strongly depends on assumptions of linearity and stationarity, which may limit their ability to capture complex and nonlinear relationships often observed in real sales data (BOX; JENKINS; REINSEL, 2015; HYNDMAN; ATHANASOPOULOS, 2018).

In recent years, machine learning techniques have been widely employed in demand forecasting problems, due to their flexibility to incorporate multiple explanatory variables and model nonlinear patterns without the need for rigid parametric specifications (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). In particular, algorithms based on decision trees and ensemble methods, such as Random Forest and gradient boosting techniques, have demonstrated superior performance in predictive applications in retail and industry, especially when combined with appropriate temporal attribute engineering strategies (BREIMAN, 2001; CHEN; GUESTRIN, 2016; KE et al., 2017).

Recent studies indicate that the incorporation of calendar variables, holiday indicators, lagged attributes, and mobile statistics contributes significantly to the increased accuracy of predictive models applied to sales time series, by allowing the explicit representation of seasonal patterns and short- and medium-term temporal dependencies (BERGMEIR; BENÍTEZ, 2012; HYNDMAN; ATHANASOPOULOS, 2018). Despite these advances, empirical studies that comparatively evaluate the performance of different machine learning algorithms applied to the forecast of daily sales in the beer sector, using real data and adequate time validation, are still relatively scarce in the Brazilian context.

In this context, the present study aims to evaluate the application of machine learning algorithms in the daily sales forecast of a Brazilian brewing industry, comparing models based on decision trees, namely: Decision Tree, Random Forest, XGBoost and LightGBM. To this end, a set of historical data of daily sales is used, with a temporal division between training and testing, and temporal attribute engineering is used to capture seasonal patterns and historical dependencies.

The main scientific contribution of this work consists of providing a comparative empirical analysis of the performance of different machine learning algorithms in a real scenario of the Brazilian brewing industry, evidencing the predictive gains associated with ensemble models and gradient boosting techniques. Additionally, the study contributes to the literature by demonstrating that the use of relatively simple temporal attributes, combined with robust machine learning methods, can produce consistent and relevant predictions to support managerial decisions related to production planning, inventory management, and logistics.

## 2 THEORETICAL FRAMEWORK

Sales forecasting is a fundamental element for the operational and strategic planning of organizations, especially in sectors marked by high seasonality and marked variability in demand. In the brewing industry, consumption behavior shows cyclical patterns associated with weekends, holidays, and festive events, which imposes relevant challenges to the management of production, inventories, and logistics. In view of this scenario, the literature specialized in demand forecasting has highlighted the relevance of comparative evaluation of predictive methods, pointing out that more flexible approaches tend to have superior performance in modeling time series with complex and nonlinear patterns, especially in short-term forecasting problems (MAKRIDAKIS; BAKAS, 2016).

In the context of the brewing industry, studies indicate that consumption is strongly influenced by temporal factors, such as weekdays, festive periods and long holidays, as well as cultural and climatic aspects (ALVES; SILVA, 2020). These patterns make sales forecasting particularly sensitive to the choice of method and the appropriate incorporation of explanatory variables capable of representing such effects.

### 2.1 TRADITIONAL TIME SERIES FORECASTING MODELS

Classical statistical models, such as the autoregressive integrated moving average models (ARIMA), have been widely used for time series forecasting in different industrial contexts. These models are based on assumptions of linearity and stationarity and seek to capture temporal dependencies through autoregressive components and moving averages (BOX; JENKINS; REINSEL, 2015). Although they have good interpretability and solid theoretical foundation, their application may be limited when the series presents complex nonlinear patterns or when multiple exogenous variables influence the behavior of demand.

In addition, the manual selection of the model structure and the adjustment of parameters can make these methods sensitive to noise and structural changes in the behavior of the series, a relatively common phenomenon in dynamic markets and subject to changes in consumption habits (HYNDMAN; ATHANASOPOULOS, 2018). Such limitations have motivated the search for more flexible and adaptive approaches, especially in short-term sales forecasting applications.

## 2.2 MACHINE LEARNING IN DEMAND FORECASTING

Machine learning techniques have stood out as promising alternatives for demand forecasting, due to their ability to model nonlinear relationships and to simultaneously incorporate a large set of explanatory variables, without the need for restrictive parametric hypotheses (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Unlike traditional statistical models, machine learning methods are predominantly predictive performance-driven, which makes them especially suitable for scenarios in which prediction accuracy is a priority (FILDES; MAKRIDAKIS; STELLA, 2019).

Among the widely used algorithms, the methods based on decision trees stand out, which partition the space of attributes into homogeneous regions, allowing the capture of complex interactions between variables. However, isolated decision trees tend to present a high risk of overfitting, which compromises their ability to generalize (LIU; WANG; ZHANG, 2012).

To mitigate this problem, ensemble techniques have been proposed, combining multiple models to produce more robust predictions. Random Forest, for example, uses the training of several trees on random subsets of the data and variables, reducing the variance of the model and improving its predictive stability (LIU; WANG; ZHANG, 2012). Gradient boosting methods, on the other hand, build models sequentially, so that each new tree seeks to correct the errors of the previous ones, resulting in high predictive power.

## 2.3 GRADIENT BOOSTING AND TEMPORAL ATTRIBUTE ENGINEERING

Tree-based gradient boosting algorithms, such as XGBoost and LightGBM, have achieved state-of-the-art performance in several regression and classification problems, including industrial demand forecasting applications. XGBoost stands out for incorporating regularization mechanisms and computational optimizations that contribute to greater efficiency and control of overfitting (CHEN; GUESTRIN, 2016). Similarly, LightGBM was developed with a focus on scalability and memory efficiency, maintaining high accuracy even in large volumes of data (KE et al., 2017).

The literature points out that the performance of these algorithms is strongly influenced by the quality of attribute engineering, especially in time series problems. The incorporation of calendar variables, holiday indicators, lags, and mobile statistics allows machine learning models to explicitly represent seasonal patterns and temporal dependencies, approximating the underlying structure of the series (BERGMEIR; BENÍTEZ, 2012; HYNDMAN; ATHANASOPOULOS, 2018).

In this sense, empirical studies have shown that the combination of gradient boosting methods with temporal attribute engineering results in significant gains in predictive performance when compared to both traditional statistical models and machine learning approaches without adequate treatment of the temporal dimension (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; KE et al., 2017).

Despite the advances observed in the international literature, there is still a gap in the national context regarding the comparative evaluation of machine learning algorithms applied to the prediction of daily sales in the Brazilian beer sector, especially in studies that use real data and rigorous time validation. As summarized in **Table 1**, different forecasting approaches have distinct characteristics, advantages, and limitations, highlighting the need for empirical analyses that consider both predictive performance and the challenges associated with time series modeling.

**Table 1**

*Comparison between time series forecasting approaches*

| Approach | Key features | Advantages | Limitations |
|---|---|---|---|
| Statistical Models (ARIMA) | Based on linearity and stationarity | Interpretability; solid theoretical foundation | Difficulty capturing nonlinearities |
| Decision trees | Recursive Partitioning of Attribute Space | Simplicity; Interpretability | High risk of overfitting |
| Random Forest | Tree ensemble with random sampling | Reduction of variance; Robustness | Lower overall interpretability |
| Gradient Boosting (XGBoost / LightGBM) | Sequential Tree Construction with Error Correction | High accuracy; Nonlinear Modeling | Increased computational complexity |

Source: Prepared by the authors, based on Breiman (2001), Hastie et al. (2009), Chen and Guestrin (2016) and Ke et al. (2017).

In addition, recent research highlights that the performance of these algorithms is strongly associated with the quality of temporal attribute engineering, including the use of calendar variables, lagged attributes, and mobile statistics, capable of effectively representing temporal dependencies and seasonal patterns (BANDARA et al., 2020). In this sense, the present study positions itself in this context by empirically analyzing the performance of different machine learning algorithms in a real industrial scenario, contributing to the advancement of knowledge applied to demand forecasting in the brewing sector.

## 3 METHODOLOGY

The present research is characterized as a quantitative study, of applied nature, with an explanatory approach, whose objective is to evaluate the performance of different machine learning algorithms in the daily sales forecast of a Brazilian brewing industry. The methodological design was structured in order to ensure the replicability of the experiments and the validity of the results, respecting the temporal nature of the data analyzed.

### 3.1 DATASET AND EXPERIMENTAL DESIGN

The study used a set of historical data of daily beer sales, measured in volume, covering the period from May 1, 2024 to May 31, 2025. The data were organized in a continuous time series, without additional aggregations, preserving the daily granularity necessary for the analysis of short-term patterns.

Data division was performed temporarily, as recommended in the literature for time series problems, in order to avoid information leakage between the training and test sets (HYNDMAN; ATHANASOPOULOS, 2018; BERGMEIR; BENÍTEZ, 2012). Thus, the records referring to the year 2024 were used for training and adjustment of the models, while the data from 2025 were reserved exclusively for the final evaluation of the predictive performance.

### 3.2 TEMPORAL ATTRIBUTE ENGINEERING

In order to capture seasonal patterns and temporal dependencies inherent to daily sales, the time attribute engineering stage was carried out, considered fundamental in machine learning applications for time series. The predictor variables were organized into the following categories:

a) Calendar attributes: day of the week, day of the month, month of the year, binary indicator of weekend/working day and indicators of national holidays and relevant festive dates (such as Carnival, Christmas and New Year's Eve), aiming to represent weekly and annual cycles of consumption;
b) Lag attributes: values of sales observed in previous periods, including short-term (e.g., 1 day) and weekly (e.g., 7 days) lags, allowing modeling the direct temporal dependence between consecutive observations;
c) Moving statistics: moving averages and standard deviations calculated in 7-day and 30-day sliding windows, in order to smooth out one-off fluctuations and capture recent trends in the series.

Additional exogenous variables, such as climate data or information from promotional campaigns, were not incorporated due to unavailability in the analyzed data set, thus prioritizing temporal attributes widely used and validated in the literature.

## 3.3 ALGORITHMS EVALUATED

Four algorithms based on decision trees, widely used in regression and demand forecasting problems, were evaluated:

a) Decision Tree: regression model that partitions the attribute space recursively, serving as a baseline for its simplicity and interpretability;
b) Random Forest: ensemble method that combines multiple decision trees trained on random subsets of data and variables, with aggregation of predictions by mean, aiming to reduce variance and increase the robustness of the model (BREIMAN, 2001);
c) XGBoost: gradient boosting algorithm that builds trees sequentially to correct residual errors from previous models, incorporating regularization and computational optimizations to control overfitting and increase efficiency (CHEN; GUESTRIN, 2016);
d) LightGBM: histogram-based gradient boosting technique, designed for efficient training and reduced memory usage, while maintaining high predictive accuracy (KE et al., 2017).

## 3.4 TRAINING, VALIDATION, AND EVALUATION OF MODELS

The training of the models was carried out using exclusively the 2024 dataset. For the adjustment of hyperparameters, cross-validation with temporal division was adopted, respecting the chronological order of the observations and avoiding the use of future data in the learning process.

The final evaluation of the predictive performance was conducted on the test set (2025 data), using metrics widely consolidated in the literature for regression problems:

a) Coefficient of determination ($R^2$), which measures the proportion of variance explained by the model;
b) Mean Absolute Error (MAE), which expresses the mean error in absolute terms, facilitating practical interpretation;
c) Root mean squared error (RMSE), which penalizes larger errors and provides a sensitive measure of extreme deviations.

These metrics allow for a consistent comparative analysis between the models, considering both the accuracy and stability of the forecasts.

## 3.5 ETHICAL CONSIDERATIONS AND METHODOLOGICAL LIMITATIONS

The data used in this study do not contain personal or sensitive information, and are restricted to aggregated sales records, which does not require formal ethical consent procedures. As methodological limitations, the absence of relevant exogenous variables, such as climate and promotional actions, and the analysis restricted to a single brewing industry stand out, which may limit the generalization of the results. Even so, the adopted design allows for a robust evaluation of the potential of machine learning algorithms in a real industrial application scenario.

# 4 RESULTS

This section presents the results obtained from the application of the machine learning algorithms proposed for the daily sales forecast of a Brazilian brewing industry. The models were evaluated based on the predictive performance on the test set, corresponding to the data from the year 2025, preserving the temporal integrity of the series.

## 4.1 PERFORMANCE OF FORECASTING MODELS

Table 2 presents the quantitative results of the four algorithms evaluated — Decision Tree, Random Forest, XGBoost and LightGBM — considering the metrics coefficient of determination ($R^2$), mean absolute error (MAE) and root mean squared error (RMSE).

**Table 2**

*Comparison of the performance of daily sales forecasting models*

| Model | $R^2$ | MOTHER | RMSE |
|---|---|---|---|
| Decision Tree | 0,700 | 380,5 | 480,2 |
| Random Forest | 0,850 | 250,0 | 320,0 |
| LightGBM | 0,880 | 212,3 | 275,6 |
| XGBoost | 0,899 | 198,4 | 263,5 |

Source: Prepared by the authors.

It is observed that the ensemble models presented superior performance than the isolated decision tree. XGBoost obtained the highest value of $R^2$, explaining approximately 89.9% of the variance of daily sales, in addition to presenting the lowest values of MAE and RMSE. LightGBM showed close performance, followed by Random Forest, while Decision Tree showed the highest prediction errors.

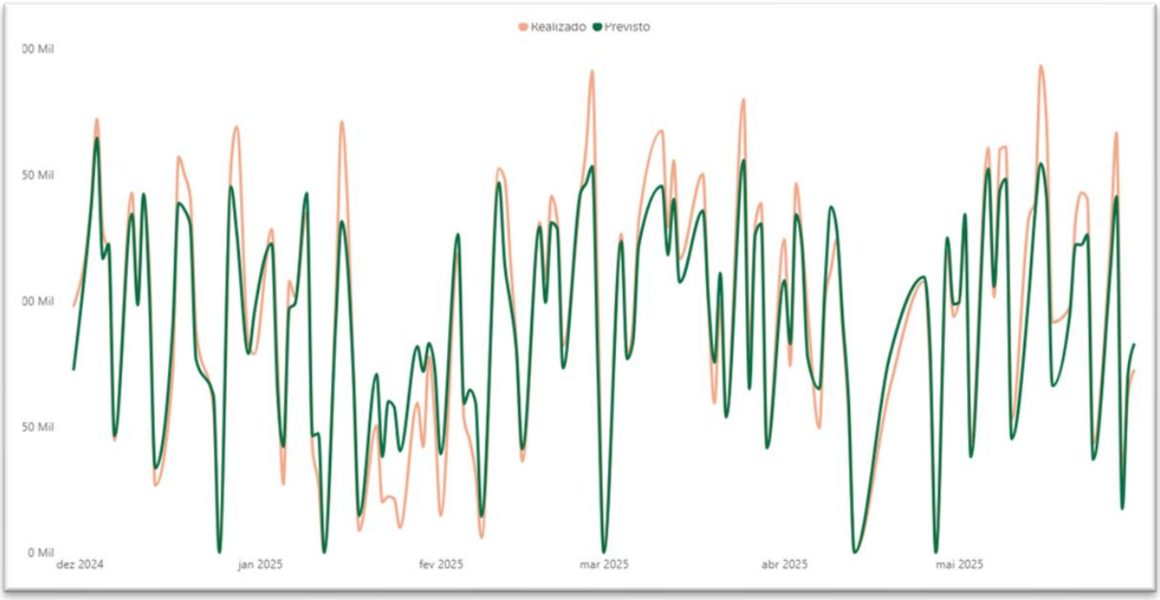
## 4.2 GRAPHICAL ANALYSIS OF FORECASTS

To complement the quantitative analysis, Figure 1 presents a visual comparison between the actual sales values and the forecasts generated by the best performing model (XGBoost), over the test period.

Visual inspection indicates that the XGBoost model follows the dynamics of the time series over the analyzed period, capturing weekly seasonal patterns and oscillations over the analyzed period. Punctual differences between observed and predicted values are concentrated on atypical days, possibly associated with unmodeled factors, such as specific promotions or climatic variations.

**Figure 1**

*Actual sales and forecasted sales of the XGBoost model in the test period*




Source: Prepared by the authors.

## 5 CONCLUSION

The present study aimed to evaluate the application of machine learning algorithms in the daily sales forecast of a Brazilian brewing industry, comparing the performance of models based on decision trees in a real scenario of historical data. To this end, an experimental design was used that respected the temporal nature of the series, combined with the engineering of temporal attributes to capture seasonal patterns and historical dependencies.

The results obtained showed that the ensemble models presented superior performance to the isolated decision tree, showing greater generalization capacity and predictive accuracy. Among the algorithms evaluated, XGBoost stood out by explaining approximately 89.9% of the variance of daily sales, in addition to presenting the lowest values of mean absolute error and root mean squared error. LightGBM also showed high performance, reinforcing the adequacy of gradient boosting methods for demand forecasting problems with high seasonality.

From a scientific point of view, the main contribution of this work is to provide a comparative empirical analysis of the performance of different machine learning algorithms applied to the prediction of daily sales in the context of the Brazilian brewing industry, an area still little explored in the national literature. Additionally, the results indicate that the use of relatively simple temporal attributes, when combined with robust machine learning models, can produce consistent predictions, even in the absence of more complex exogenous variables.

From a managerial perspective, the findings suggest that the adoption of machine learning techniques can support decisions related to production planning, inventory management, and logistics, contributing to the reduction of stockouts and overstocks. Although the study is restricted to a single industry, the methodological procedures adopted are replicable and can be extended to other industrial contexts with similar characteristics.

As limitations, the absence of relevant exogenous variables, such as climate data and information on promotional campaigns, as well as the analysis centered on a single set of data, stand out. As future works, it is recommended to incorporate new data sources, evaluate additional algorithms — such as CatBoost and neural networks — and expand the study to multiple product lines or different companies in the sector, in order to broaden the generalization of the results and deepen the understanding of the factors that influence demand in the beer market.


## ACKNOWLEDGMENTS

The authors would like to thank the State University of Minas Gerais (UEMG), through the Research Support Program (PAPq/UEMG), for the institutional support granted to the development of this research.


## REFERENCES


Alves, R. M., & Silva, L. F. (2023). Padrões temporais no consumo de bebidas alcoólicas no Brasil. Revista de Saúde Pública, 57, Article 17. https://www.scielosp.org/article/rsp/2023.v57/17/pt/

Bandara, K., Bergmeir, C., & Smith, R. J. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications, 140, Article 112896. https://doi.org/10.1016/j.eswa.2019.112896

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

de Oliveira Dias, M., & Falconi, D. (2018). The evolution of craft beer industry in Brazil. Journal of Economics and Business, 1(4), 618–626. https://doi.org/10.31014/aior.1992.01.04.55

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts. https://otexts.com/fpp2/

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (pp. 3146–3154). Curran Associates.

Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In B. Liu, M. Ma, & J. Chang (Eds.), Information computing and applications (Lecture Notes in Computer Science, Vol. 7473, pp. 246–252). Springer. https://doi.org/10.1007/978-3-642-34062-8_32

Makridakis, S., & Bakas, N. (2016). Forecasting and uncertainty: A survey. Risk and Decision Analysis, 6(1), 37–64. https://doi.org/10.3233/RDA-150114