


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

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

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

 <https://doi.org/10.56238/arev8n2-064>

Data de submissão: 12/01/2026

Data de publicação: 12/02/2026

Rodrigo Ribeiro Salgado César

Graduando em Sistemas de Informação

Instituição: 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>

Allana Eduarda Braga Iquegami

Graduanda em Sistemas de Informação

Instituição: 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>

Nelson Marinelli Filho

Doutor em Engenharia Mecânica

Instituição: 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>

Gil Eduardo Guimarães

Doutor em Engenharia Mecânica

Instituição: 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>

Geraldo Nunes Corrêa

Professor Orientador Bolsista – PAPq

Doutor em Engenharia de Produção

Instituição: 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>

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 superaram 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.

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 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.

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 Cervecer.

1 INTRODUÇÃO

A previsão de vendas desempenha papel central no planejamento operacional e estratégico das organizações, especialmente em setores caracterizados por elevada sazonalidade e forte variabilidade da demanda. No caso da indústria cervejeira, o consumo apresenta padrões cíclicos associados a fins de semana, feriados nacionais, eventos culturais e condições climáticas, o que impõe desafios significativos à gestão de produção, estoques e logística (ALVES; SILVA, 2023; DE OLIVEIRA DIAS; FALCONI, 2018). Erros na previsão podem resultar tanto em rupturas de estoque quanto em excesso de produtos armazenados, com impactos diretos nos custos operacionais e no nível de serviço ao mercado.

Tradicionalmente, a previsão de demanda tem sido realizada por meio de modelos estatísticos clássicos de séries temporais, como os modelos autorregressivos integrados de médias móveis (ARIMA). Embora esses modelos possuam sólida fundamentação teórica e ampla aplicação, sua eficácia depende fortemente de pressupostos de linearidade e estacionariedade, o que pode limitar sua capacidade de capturar relações complexas e não lineares frequentemente observadas em dados reais de vendas (BOX; JENKINS; REINSEL, 2015; HYNDMAN; ATHANASOPOULOS, 2018).

Nos últimos anos, técnicas de aprendizado de máquina têm sido amplamente empregadas em problemas de previsão de demanda, em virtude de sua flexibilidade para incorporar múltiplas variáveis explicativas e modelar padrões não lineares sem a necessidade de especificações paramétricas rígidas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Em particular, algoritmos baseados em árvores de decisão e métodos de ensemble, como Random Forest e técnicas de gradient boosting, têm demonstrado desempenho superior em aplicações de previsão no varejo e na indústria, sobretudo quando combinados com estratégias adequadas de engenharia de atributos temporais (BREIMAN, 2001; CHEN; GUESTRIN, 2016; KE et al., 2017).

Estudos recentes indicam que a incorporação de variáveis de calendário, indicadores de feriados, atributos defasados e estatísticas móveis contribui significativamente para o aumento da acurácia de modelos preditivos aplicados a séries temporais de vendas, ao permitir a representação explícita de padrões sazonais e dependências temporais de curto e médio prazo (BERGMEIR; BENÍTEZ, 2012; HYNDMAN; ATHANASOPOULOS, 2018). Apesar desses avanços, ainda são relativamente escassos, no contexto brasileiro, estudos empíricos que avaliem de forma comparativa o desempenho de diferentes algoritmos de aprendizado de máquina aplicados à previsão de vendas diárias no setor cervejeiro, utilizando dados reais e validação temporal adequada.

Diante desse contexto, o presente estudo tem como objetivo avaliar a aplicação de algoritmos de aprendizado de máquina na previsão de vendas diárias de uma indústria cervejeira brasileira,

comparando modelos baseados em árvores de decisão, a saber: Decision Tree, Random Forest, XGBoost e LightGBM. Para tanto, utiliza-se um conjunto de dados históricos de vendas diárias, com divisão temporal entre treinamento e teste, e emprega-se engenharia de atributos temporais para capturar padrões sazonais e dependências históricas.

A principal contribuição científica deste trabalho consiste em fornecer uma análise empírica comparativa do desempenho de diferentes algoritmos de aprendizado de máquina em um cenário real da indústria cervejeira brasileira, evidenciando os ganhos preditivos associados a modelos de ensemble e técnicas de gradient boosting. Adicionalmente, o estudo contribui para a literatura ao demonstrar que a utilização de atributos temporais relativamente simples, aliada a métodos de aprendizado de máquina robustos, pode produzir previsões consistentes e relevantes para apoiar decisões gerenciais relacionadas ao planejamento de produção, gestão de estoques e logística.

2 REFERENCIAL TEÓRICO

A previsão de vendas constitui um elemento fundamental para o planejamento operacional e estratégico das organizações, sobretudo em setores marcados por elevada sazonalidade e acentuada variabilidade da demanda. Na indústria cervejeira, o comportamento do consumo evidencia padrões cíclicos associados a fins de semana, feriados e eventos festivos, o que impõe desafios relevantes à gestão da produção, dos estoques e da logística. Diante desse cenário, a literatura especializada em previsão de demanda tem ressaltado a relevância da avaliação comparativa de métodos preditivos, apontando que abordagens mais flexíveis tendem a apresentar desempenho superior na modelagem de séries temporais com padrões complexos e não lineares, especialmente em problemas de previsão de curto prazo (MAKRIDAKIS; BAKAS, 2016).

No contexto da indústria cervejeira, estudos indicam que o consumo apresenta forte influência de fatores temporais, como dias da semana, períodos festivos e feriados prolongados, além de aspectos culturais e climáticos (ALVES; SILVA, 2020). Esses padrões tornam a previsão de vendas particularmente sensível à escolha do método e à incorporação adequada de variáveis explicativas capazes de representar tais efeitos.

2.1 MODELOS TRADICIONAIS DE PREVISÃO DE SÉRIES TEMPORAIS

Modelos estatísticos clássicos, como os modelos autorregressivos integrados de médias móveis (ARIMA), têm sido amplamente utilizados para previsão de séries temporais em diferentes contextos industriais. Esses modelos baseiam-se em pressupostos de linearidade e estacionariedade e buscam capturar dependências temporais por meio de componentes autorregressivos e de médias

móveis (BOX; JENKINS; REINSEL, 2015). Embora apresentem boa interpretabilidade e sólida fundamentação teórica, sua aplicação pode ser limitada quando a série apresenta padrões não lineares complexos ou quando múltiplas variáveis exógenas influenciam o comportamento da demanda.

Além disso, a seleção manual da estrutura do modelo e o ajuste de parâmetros podem tornar esses métodos sensíveis a ruídos e mudanças estruturais no comportamento das séries, fenômeno relativamente comum em mercados dinâmicos e sujeitos a alterações de hábitos de consumo (HYNDMAN; ATHANASOPOULOS, 2018). Tais limitações têm motivado a busca por abordagens mais flexíveis e adaptativas, especialmente em aplicações de previsão de vendas de curto prazo.

2.2 APRENDIZADO DE MÁQUINA NA PREVISÃO DE DEMANDA

Técnicas de aprendizado de máquina têm se destacado como alternativas promissoras para a previsão de demanda, em razão de sua capacidade de modelar relações não lineares e de incorporar simultaneamente um grande conjunto de variáveis explicativas, sem a necessidade de hipóteses paramétricas restritivas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Diferentemente dos modelos estatísticos tradicionais, métodos de aprendizado de máquina são orientados predominantemente pelo desempenho preditivo, o que os torna especialmente adequados para cenários nos quais a acurácia da previsão é prioritária (FILDES; MAKRIDAKIS; STELLA, 2019).

Entre os algoritmos amplamente utilizados, destacam-se os métodos baseados em árvores de decisão, que particionam o espaço de atributos em regiões homogêneas, permitindo capturar interações complexas entre variáveis. Contudo, árvores de decisão isoladas tendem a apresentar elevado risco de sobreajuste, o que compromete sua capacidade de generalização (LIU; WANG; ZHANG, 2012).

Para mitigar esse problema, técnicas de ensemble têm sido propostas, combinando múltiplos modelos para produzir previsões mais robustas. O Random Forest, por exemplo, utiliza o treinamento de várias árvores em subconjuntos aleatórios dos dados e das variáveis, reduzindo a variância do modelo e melhorando sua estabilidade preditiva (LIU; WANG; ZHANG, 2012). Já os métodos de gradient boosting constroem modelos de forma sequencial, de modo que cada nova árvore busca corrigir os erros das anteriores, resultando em elevado poder preditivo.

2.3 GRADIENT BOOSTING E ENGENHARIA DE ATRIBUTOS TEMPORAIS

Algoritmos de gradient boosting baseados em árvores, como o XGBoost e o LightGBM, têm alcançado desempenho de estado da arte em diversos problemas de regressão e classificação, incluindo aplicações industriais de previsão de demanda. O XGBoost destaca-se por incorporar

mecanismos de regularização e otimizações computacionais que contribuem para maior eficiência e controle do sobreajuste (CHEN; GUESTRIN, 2016). De forma semelhante, o LightGBM foi desenvolvido com foco em escalabilidade e eficiência de memória, mantendo elevada acurácia mesmo em grandes volumes de dados (KE et al., 2017).

A literatura aponta que o desempenho desses algoritmos é fortemente influenciado pela qualidade da engenharia de atributos, especialmente em problemas de séries temporais. A incorporação de variáveis de calendário, indicadores de feriados, atributos defasados (lags) e estatísticas móveis permite que modelos de aprendizado de máquina representem explicitamente padrões sazonais e dependências temporais, aproximando-se da estrutura subjacente da série (BERGMEIR; BENÍTEZ, 2012; HYNDMAN; ATHANASOPOULOS, 2018).

Nesse sentido, estudos empíricos têm demonstrado que a combinação de métodos de gradient boosting com engenharia de atributos temporais resulta em ganhos significativos de desempenho preditivo quando comparada tanto a modelos estatísticos tradicionais quanto a abordagens de aprendizado de máquina sem tratamento adequado da dimensão temporal (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; KE et al., 2017).

Apesar dos avanços observados na literatura internacional, ainda se identifica uma lacuna no contexto nacional quanto à avaliação comparativa de algoritmos de aprendizado de máquina aplicados à previsão de vendas diárias no setor cervejeiro brasileiro, especialmente em estudos que utilizam dados reais e validação temporal rigorosa. Conforme sintetizado na **Tabela 1**, diferentes abordagens de previsão apresentam características, vantagens e limitações distintas, evidenciando a necessidade de análises empíricas que considerem tanto o desempenho preditivo quanto os desafios associados à modelagem de séries temporais.

Tabela 1. Comparação entre abordagens de previsão de séries temporais

Abordagem	Características principais	Vantagens	Limitações
Modelos estatísticos (ARIMA)	Baseados em linearidade e estacionariedade	Interpretabilidade; base teórica sólida	Dificuldade em capturar não linearidades
Árvores de decisão	Particionamento recursivo do espaço de atributos	Simplicidade; interpretabilidade	Alto risco de sobreajuste
Random Forest	Ensemble de árvores com amostragem aleatória	Redução de variância; robustez	Menor interpretabilidade global
Gradient Boosting (XGBoost / LightGBM)	Construção sequencial de árvores com correção de erros	Alta acurácia; modelagem não linear	Maior complexidade computacional

Fonte: Elaborada pelos autores, com base em Breiman (2001), Hastie et al. (2009), Chen e Guestrin (2016) e Ke et al. (2017).

Ademais, pesquisas recentes destacam que o desempenho desses algoritmos está fortemente associado à qualidade da engenharia de atributos temporais, incluindo a utilização de variáveis de calendário, atributos defasados e estatísticas móveis, capazes de representar dependências temporais e padrões sazonais de forma eficaz (BANDARA et al., 2020). Nesse sentido, o presente estudo posiciona-se nesse contexto ao analisar empiricamente o desempenho de diferentes algoritmos de aprendizado de máquina em um cenário industrial real, contribuindo para o avanço do conhecimento aplicado à previsão de demanda no setor cervejeiro.

3 METODOLOGIA

A presente pesquisa caracteriza-se como um estudo quantitativo, de natureza aplicada, com abordagem explicativa, cujo objetivo é avaliar o desempenho de diferentes algoritmos de aprendizado de máquina na previsão de vendas diárias de uma indústria cervejeira brasileira. O delineamento metodológico foi estruturado de forma a garantir a replicabilidade dos experimentos e a validade dos resultados, respeitando a natureza temporal dos dados analisados.

3.1 CONJUNTO DE DADOS E DELINEAMENTO EXPERIMENTAL

O estudo utilizou um conjunto de dados históricos de vendas diárias de cerveja, medidos em volume, abrangendo o período de 1º de maio de 2024 a 31 de maio de 2025. Os dados foram organizados em uma série temporal contínua, sem agregações adicionais, preservando a granularidade diária necessária para a análise de padrões de curto prazo.

A divisão dos dados foi realizada de forma temporal, conforme recomendado na literatura para problemas de séries temporais, a fim de evitar vazamento de informação entre os conjuntos de treinamento e teste (HYNDMAN; ATHANASOPOULOS, 2018; BERGMEIR; BENÍTEZ, 2012). Assim, os registros referentes ao ano de 2024 foram utilizados para treinamento e ajuste dos modelos, enquanto os dados de 2025 foram reservados exclusivamente para a avaliação final do desempenho preditivo.

3.2 ENGENHARIA DE ATRIBUTOS TEMPORAIS

Com o objetivo de capturar padrões sazonais e dependências temporais inerentes às vendas diárias, foi realizada a etapa de engenharia de atributos temporais, considerada fundamental em aplicações de aprendizado de máquina para séries temporais. As variáveis preditoras foram organizadas nas seguintes categorias:

- a) Atributos de calendário: dia da semana, dia do mês, mês do ano, indicador binário de fim de semana/dia útil e indicadores de feriados nacionais e datas festivas relevantes (como Carnaval, Natal e Réveillon), visando representar ciclos semanais e anuais do consumo;
- b) Atributos defasados (lags): valores das vendas observadas em períodos anteriores, incluindo defasagens de curto prazo (por exemplo, 1 dia) e de natureza semanal (por exemplo, 7 dias), permitindo modelar a dependência temporal direta entre observações consecutivas;
- c) Estatísticas móveis: médias móveis e desvios padrão calculados em janelas deslizantes de 7 e 30 dias, com o intuito de suavizar flutuações pontuais e capturar tendências recentes da série.

Variáveis exógenas adicionais, como dados climáticos ou informações de campanhas promocionais, não foram incorporadas por indisponibilidade no conjunto de dados analisado, priorizando-se, assim, atributos temporais amplamente utilizados e validados na literatura.

3.3 ALGORITMOS AVALIADOS

Foram avaliados quatro algoritmos baseados em árvores de decisão, amplamente empregados em problemas de regressão e previsão de demanda:

- a) Decision Tree (Árvore de Decisão): modelo de regressão que particiona o espaço de atributos de forma recursiva, servindo como baseline por sua simplicidade e interpretabilidade;
- b) Random Forest: método ensemble que combina múltiplas árvores de decisão treinadas em subconjuntos aleatórios dos dados e das variáveis, com agregação das previsões por média, visando reduzir a variância e aumentar a robustez do modelo (BREIMAN, 2001);
- c) XGBoost: algoritmo de gradient boosting que constrói árvores sequencialmente para corrigir erros residuais dos modelos anteriores, incorporando regularização e otimizações computacionais para controle do sobreajuste e aumento da eficiência (CHEN; GUESTRIN, 2016);
- d) LightGBM: técnica de gradient boosting baseada em histogramas, projetada para treinamento eficiente e uso reduzido de memória, mantendo elevada acurácia preditiva (KE et al., 2017).

3.4 TREINAMENTO, VALIDAÇÃO E AVALIAÇÃO DOS MODELOS

O treinamento dos modelos foi realizado utilizando exclusivamente o conjunto de dados de 2024. Para o ajuste de hiperparâmetros, adotou-se validação cruzada com divisão temporal, respeitando a ordem cronológica das observações e evitando o uso de dados futuros no processo de aprendizado.

A avaliação final do desempenho preditivo foi conduzida sobre o conjunto de teste (dados de 2025), utilizando-se métricas amplamente consolidadas na literatura para problemas de regressão:

- a) Coeficiente de determinação (R^2), que mede a proporção da variância explicada pelo modelo;
- b) Erro absoluto médio (MAE), que expressa o erro médio em termos absolutos, facilitando a interpretação prática;
- c) Raiz do erro quadrático médio (RMSE), que penaliza erros maiores e fornece uma medida sensível a desvios extremos.

Essas métricas permitem uma análise comparativa consistente entre os modelos, considerando tanto a acurácia quanto a estabilidade das previsões.

3.5 CONSIDERAÇÕES ÉTICAS E LIMITAÇÕES METODOLÓGICAS

Os dados utilizados neste estudo não contêm informações pessoais ou sensíveis, restringindo-se a registros agregados de vendas, o que dispensa procedimentos formais de consentimento ético. Como limitações metodológicas, destacam-se a ausência de variáveis exógenas relevantes, como clima e ações promocionais, e a análise restrita a uma única indústria cervejeira, o que pode limitar a generalização dos resultados. Ainda assim, o delineamento adotado permite avaliar de forma robusta o potencial dos algoritmos de aprendizado de máquina em um cenário real de aplicação industrial.

4 RESULTADOS

Esta seção apresenta os resultados obtidos a partir da aplicação dos algoritmos de aprendizado de máquina propostos para a previsão de vendas diárias de uma indústria cervejeira brasileira. Os modelos foram avaliados com base no desempenho preditivo sobre o conjunto de teste, correspondente aos dados do ano de 2025, preservando-se a integridade temporal da série.

4.1 DESEMPENHO DOS MODELOS DE PREVISÃO

A Tabela 2 apresenta os resultados quantitativos dos quatro algoritmos avaliados — Decision Tree, Random Forest, XGBoost e LightGBM — considerando as métricas coeficiente de determinação (R^2), erro absoluto médio (MAE) e raiz do erro quadrático médio (RMSE).

Tabela 2. Comparação do desempenho dos modelos de previsão de vendas diárias

Modelo	R²	MAE	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

Fonte: Elaborada pelos autores.

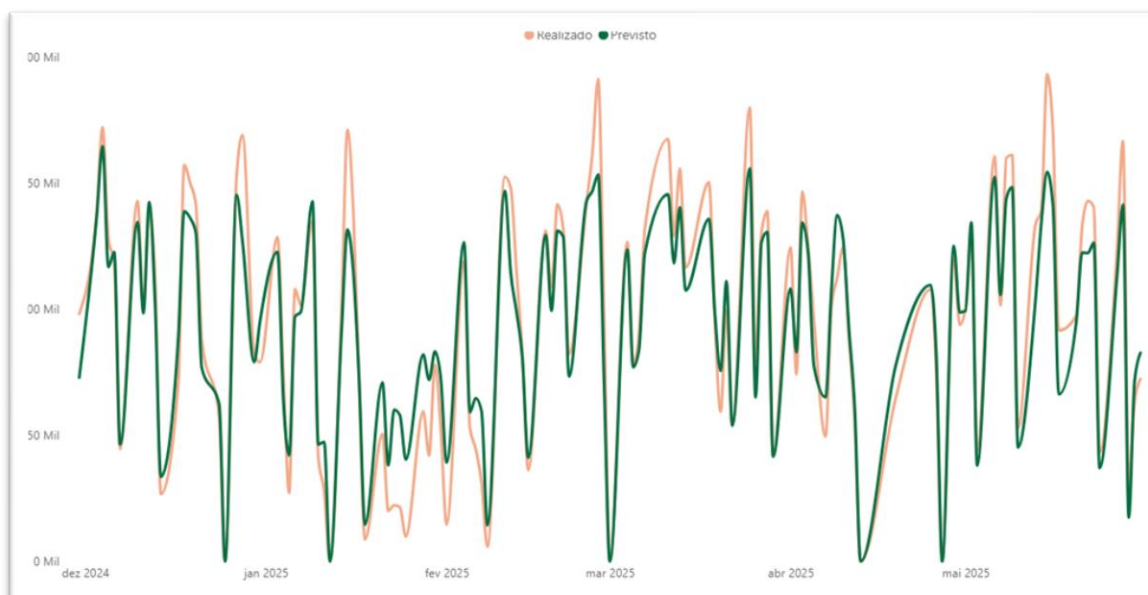
Observa-se que os modelos do tipo ensemble apresentaram desempenho superior ao da árvore de decisão isolada. O XGBoost obteve o maior valor de R², explicando aproximadamente 89,9% da variância das vendas diárias, além de apresentar os menores valores de MAE e RMSE. O LightGBM apresentou desempenho próximo, seguido pelo Random Forest, enquanto a Decision Tree apresentou os maiores erros de previsão.

4.2 ANÁLISE GRÁFICA DAS PREVISÕES

Para complementar a análise quantitativa, a Figura 1 apresenta uma comparação visual entre os valores reais de vendas e as previsões geradas pelo modelo de melhor desempenho (XGBoost), ao longo do período de teste.

A inspeção visual indica que o modelo XGBoost acompanha a dinâmica da série temporal ao longo do período analisado, capturando padrões sazonais semanais e oscilações ao longo do período analisado. Diferenças pontuais entre valores observados e previstos concentram-se em dias atípicos, possivelmente associados a fatores não modelados, como promoções específicas ou variações climáticas.

Figura 1. Vendas reais e vendas previstas pelo modelo XGBoost no período de teste.



Fonte: Elaborada pelos autores.

5 CONCLUSÃO

O presente estudo teve como objetivo avaliar a aplicação de algoritmos de aprendizado de máquina na previsão de vendas diárias de uma indústria cervejeira brasileira, comparando o desempenho de modelos baseados em árvores de decisão em um cenário real de dados históricos. Para tanto, empregou-se um delineamento experimental que respeitou a natureza temporal da série, aliado à engenharia de atributos temporais para capturar padrões sazonais e dependências históricas.

Os resultados obtidos demonstraram que os modelos do tipo ensemble apresentaram desempenho superior à árvore de decisão isolada, evidenciando maior capacidade de generalização e precisão preditiva. Entre os algoritmos avaliados, o XGBoost destacou-se ao explicar aproximadamente 89,9% da variância das vendas diárias, além de apresentar os menores valores de erro absoluto médio e raiz do erro quadrático médio. O LightGBM também apresentou desempenho elevado, reforçando a adequação de métodos de gradient boosting para problemas de previsão de demanda com elevada sazonalidade.

Do ponto de vista científico, a principal contribuição deste trabalho consiste em fornecer uma análise empírica comparativa do desempenho de diferentes algoritmos de aprendizado de máquina aplicados à previsão de vendas diárias no contexto da indústria cervejeira brasileira, área ainda pouco explorada na literatura nacional. Adicionalmente, os resultados indicam que a utilização de atributos temporais relativamente simples, quando combinada a modelos robustos de aprendizado de máquina, pode produzir previsões consistentes, mesmo na ausência de variáveis exógenas mais complexas.

Sob a perspectiva gerencial, os achados sugerem que a adoção de técnicas de aprendizado de máquina pode apoiar decisões relacionadas ao planejamento de produção, gestão de estoques e logística, contribuindo para a redução de rupturas e de excessos de estoque. Ainda que o estudo se restrinja a uma única indústria, os procedimentos metodológicos adotados são replicáveis e podem ser estendidos a outros contextos industriais com características semelhantes.

Como limitações, destacam-se a ausência de variáveis exógenas relevantes, como dados climáticos e informações sobre campanhas promocionais, bem como a análise centrada em um único conjunto de dados. Como trabalhos futuros, recomenda-se a incorporação de novas fontes de dados, a avaliação de algoritmos adicionais — como CatBoost e redes neurais — e a ampliação do estudo para múltiplas linhas de produtos ou diferentes empresas do setor, de modo a ampliar a generalização dos resultados e aprofundar a compreensão dos fatores que influenciam a demanda no mercado cervejeiro.

AGRADECIMENTOS

Os autores agradecem à Universidade do Estado de Minas Gerais (UEMG), por meio do Programa de Apoio à Pesquisa (PAPq/UEMG), pelo apoio institucional concedido ao desenvolvimento desta pesquisa.

REFERÊNCIAS

ALVES, Ricardo Marques; SILVA, Luiz Fernando. Padrões temporais no consumo de bebidas alcoólicas no Brasil. *Revista de Saúde Pública*, São Paulo, v. 57, p. 17, 2023. Disponível em: <https://www.scielo.org/article/rsp/2023.v57/17/pt/>. Acesso em: [26/01/2026].

BANDARA, Kasun; BERGMEIR, Christoph; SMITH, Rob J. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, v. 140, 112896, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417419306128>. Acesso em: [26/01/2026].

BERGMEIR, Christoph; BENÍTEZ, José Manuel. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, v. 191, p. 192–213, 2012.

BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. *Time series analysis: forecasting and control*. 5. ed. Hoboken: Wiley, 2015.

BREIMAN, Leo. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>. Acesso em: 26 jan. 2026.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016. p. 785–794.

DE OLIVEIRA DIAS, Murillo; FALCONI, Davi. The evolution of craft beer industry in Brazil. *Journal of Economics and Business*, v. 1, n. 4, p. 618-626, 2018.

MAKRIDAKIS, Spyros; BAKAS, Nikolas. Forecasting and uncertainty: A survey. *Risk and Decision Analysis*, v. 6, n. 1, p. 37–64, 2016.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The elements of statistical learning: data mining, inference, and prediction*. 2. ed. New York: Springer, 2009.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. *Forecasting: principles and practice*. 2. ed. Melbourne: OTexts, 2018.

KE, Guolin; MENG, Qi; FINLEY, Thomas; WANG, Taifeng; CHEN, Wei; MA, Weidong; YE, Qiwei; LIU, Tie-Yan. LightGBM: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems*. Red Hook: Curran Associates, 2017. p. 3146–3154.

LIU, Yanli; WANG, Yourong; ZHANG, Jian. New machine learning algorithm: Random Forest. In: LIU, Bing; MA, Min; CHANG, Jie (eds.). *Information Computing and Applications. Lecture Notes in Computer Science*, v. 7473. Berlin, Heidelberg: Springer, 2012. p. 246–252. Disponível em: https://doi.org/10.1007/978-3-642-34062-8_32. Acesso em: [26/01/2026].