



Uso de IA para Planejamento Comercial e Definição de Metas

Felipe Damasceno Santos, Gabriel Sanches Tofanello, Hyandra Marjory Melo
Cabral de Teves, Rodrigo Pampolin Mendes, Leandro Zerbinatti

Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{10408854, 10410281, 10409444, 10409775, leandro.zerbinatti}@mackenzie.br

Resumo. *Este projeto aborda a utilização de técnicas de ciência de dados para apoiar o processo de definição de metas comerciais em organizações com múltiplos pontos de venda (PDVs). A atribuição de metas representa um desafio estratégico: valores excessivamente altos podem desmotivar equipes, enquanto metas demasiadamente baixas podem comprometer o crescimento da empresa. O objetivo inicial deste estudo é realizar uma análise exploratória de dados (EDA) que permita compreender padrões de vendas, sazonalidades e fatores relacionados ao desempenho comercial. Os resultados preliminares servirão como base para etapas futuras de modelagem preditiva, oferecendo suporte mais consistente e imparcial na distribuição de metas.*

Palavras-chave: *previsão de vendas; metas comerciais; análise exploratória de dados; séries temporais; planejamento comercial.*

1. Introdução

Dentro do cenário de negócios, a definição e atribuição de metas é um desafio, principalmente em organizações com diversos pontos de vendas (PDVs) ou filiais. Metas muito altas podem desmotivar equipes comerciais, uma vez que o atingimento das metas está normalmente associado à compensação financeira. Por outro lado, metas fáceis de serem atingidas podem reduzir o esforço das equipes e impactar negativamente o crescimento da empresa. Portanto, é necessário encontrar um ponto de equilíbrio que beneficie tanto a empresa quanto o desempenho comercial.

Tradicionalmente, a definição de metas tem sido realizada de forma empírica, sem considerar múltiplas variáveis que influenciam nas vendas. A aplicação de modelos baseados em dados históricos, sazonalidade, indicadores macroeconômicos e características regionais aumenta a probabilidade de definir metas calibradas para cada PDV, tornando o processo mais equitativo e eficiente. Além disso, técnicas de Inteligência Artificial (IA) permitem análises mais precisas e rápidas do que métodos tradicionais, identificando padrões de comportamento do consumidor e variáveis de mercado que poderiam ser negligenciadas em abordagens empíricas (Aldin; Nadeem, 2024). O objetivo deste projeto é construir um modelo de predição capaz de estimar as metas e/ou vendas de cada PDV, possibilitando uma análise comparativa entre desempenho previsto e o efetivamente realizado. Dessa forma, a gestão de vendas pode se tornar mais estratégica, permitindo que equipes comerciais se concentrem em

atividades de maior valor agregado, como construção de relacionamento e exploração de mercado.

Para alcançar esse objetivo, o projeto utiliza o modelo de séries temporais *Long Short-Term Memory (LSTM)*, um tipo de rede neural recorrente (*RNN*) capaz de lidar com dependências temporais complexas em dados sequenciais. Cada célula *LSTM* é composta por três portas, sendo elas, de entrada, esquecimento e saída, que determinam quais informações devem ser armazenadas, esquecidas ou transmitidas, permitindo o modelo aprender padrões de longo prazo (Albeladi; Zafar; Mueen, 2023). Essa arquitetura possibilita lidar com longos intervalos de tempo entre eventos relevantes, capturando tendências sazonais e padrões de vendas ao longo do tempo. Assim como destacam os autores Albeladi, Zafar e Mueen (2023), séries temporais, organizadas cronologicamente, são essenciais nesse contexto, pois permitem prever valores futuros e apoiar decisões estratégicas na definição de metas.

Antes da construção do modelo ser iniciada foi realizada uma análise exploratória dos dados de vendas sintéticos gerados, identificando padrões, sazonalidades e relações entre variáveis, a fim de fornecer subsídios para futuras etapas de modelagem e previsão de metas comerciais.

Foi optado por iniciar com a geração de dados sintéticos representando vendas mensais de diferentes PDVs, aplicando técnicas de análise exploratória para compreender a dinâmica dos dados antes de avançar para modelos preditivos.

2. Descrição do Problema

O principal desafio deste projeto é a dificuldade em acertar a distribuição de metas entre diferentes regiões, onde cada região apresenta características e desafios distintos. Quando a meta é definida sem um modelo racional, vieses e percepções individuais podem influenciar valores atribuídos, distorcendo os objetivos da empresa.

Pesquisas realizadas pelos autores Aldin e Nadeem (2024) indicam que migrar de métodos tradicionais para modelos baseados em IA representa um avanço na transformação digital das empresas, oferecendo maior precisão em mercados dinâmicos. Modelos de IA são capazes de processar grandes volumes de dados e incorporar variáveis externas relevantes, como sazonalidade e condições econômicas locais, aumentando a probabilidade de metas mais equitativas e adequadas a cada PDV.

3. Aspectos Éticos e Responsabilidade no Uso da IA

Os principais aspectos éticos do uso de IA neste projeto estão relacionados à justiça na distribuição de metas, considerando apenas dados e variáveis objetivas, evitando que gestores apliquem percepções subjetivas ou injustas. Além disso, optou-se pelo uso de dados sintéticos, e não de bases empresariais originais, justamente para garantir conformidade com normas de compliance e evitar riscos relacionados à privacidade e à exposição de informações sensíveis.

É importante considerar que modelos de IA podem reproduzir vieses presentes nos dados históricos, o que exige atenção para garantir imparcialidade nas previsões. A transparência e a explicabilidade dos modelos também são fundamentais, pois a falta de

clareza sobre o funcionamento do algoritmo pode reduzir a confiança de gestores e equipes de vendas (Aldin; Nadeem, 2024).

Adotar práticas responsáveis de IA, incluindo segurança e auditabilidade dos dados, fortalece a credibilidade do modelo e garante seu alinhamento com padrões éticos e sociais da organização. Complementarmente, é essencial assegurar a supervisão humana no processo decisório, evitando dependência excessiva da IA e preservando o papel crítico da análise gerencial. Por fim, deve-se considerar os impactos sociais e organizacionais do uso da tecnologia, garantindo que a definição de metas não resulte em pressões desproporcionais sobre as equipes, mas sim que seja aplicada de maneira ética, responsável e sustentável.

4. Dataset e Preparação dos Dados

Por questões de compliance e confidencialidade, optou-se por não utilizar dados reais da empresa modelo, evitando riscos de exposição de informações estratégicas e sensíveis. Dessa forma, foi gerada uma base sintética em *Python*, inspirada nas características e estrutura dos dados originais, garantindo a preservação da realidade do negócio sem comprometer a privacidade e a integridade das informações corporativas.

Dados sintéticos são informações geradas artificialmente por algoritmos, sendo úteis para treinar algoritmos de aprendizado de máquina, pois permitem o uso de dados similares aos reais sem comprometer a privacidade. Essa abordagem ajuda a contornar a escassez de dados reais e equilibra a acessibilidade com a proteção de dados (Murtaza *et al.*, 2023). Tal abordagem assegura a conformidade ética e legal do projeto, ao mesmo tempo em que mantém a relevância analítica para os objetivos propostos.

A base sintética foi construída com distribuição mensal de 2015 a 2024 para treinamento, reservando os dados de 2025 para validação do modelo. Foram criados 53 PDVs para cada mês, contendo as seguintes variáveis:

- Data (Ano, Mês)
- PDV_ID
- Estado
- Número de Funcionários
- Dias Úteis
- Meta
- Vendas
- Custo
- Lucro
- Satisfação do Cliente
- Selic
- Inflação
- Quantidade de Cotações

Na etapa de preparação dos dados em *Python*, foram aplicadas ações de limpeza, como remoção de valores ausentes, substituição de registros inconsistentes e tratamento de outliers identificados. Também foram criadas variáveis derivadas, como indicadores

de desempenho e agregações temporais, enriquecendo a base e possibilitando análises mais robustas. Por fim, os dados foram organizados em conjuntos de treino (2015–2024) e validação (2025), assegurando que o modelo pudesse ser avaliado de forma realista quanto à sua capacidade preditiva.

Na sequência, procedeu-se à análise exploratória, que incluiu estatísticas descritivas, verificação de valores ausentes, análise da distribuição temporal de vendas agregadas por ano e avaliação das correlações entre variáveis numéricas, como Vendas, Lucro, Meta e Satisfação. Essas análises permitiram identificar padrões sazonais nas vendas, uma forte correlação entre vendas e lucro, além da presença de outliers e valores faltantes, que deverão ser tratados em etapas posteriores.

5. Metodologia

A metodologia seguiu as seguintes etapas: geração de dados sintéticos, limpeza, análise exploratória dos dados, identificação de padrões relevantes e validação do *dataset*. Esse processo foi realizado para proporcionar uma base confiável e ética, que servirá de suporte para a criação do modelo *LSTM* utilizando a biblioteca *TensorFlow*. Além disso, para critério de comparação, foi realizada a predição em *XBGR*, logo tendo um outro modelo de comparação. Utilizando as seguintes variáveis como *features*: Custo, Lucro, Satisfação, Mes, Ano, Selic, Inflação, Quantidade de Cotações. E a variável *Target* foi Vendas, o modelo foi configurado, utilizando as seguintes especificações: 64 neurônios, 1 camada de densidade, otimizador Adam, *batch size* 16, *epochs* 80. A função de perda foi o MAE (*Mean Absolute Error*) sendo MAE de R\$ 5.903,44.

Por fim, para não influenciar o modelo, é feito uma predição por PDV, ou seja, para cada PDV é criado um modelo *LSTM* e feito a predição, assim, só as variáveis daquele PDV são levadas em consideração.

6. Resultados

Nesta etapa, foi realizada uma análise comparativa entre dois modelos preditivos — **LSTM (Long Short-Term Memory)** e **XGBoost Regressor (XGBR)** — com o objetivo de avaliar qual deles apresenta melhor desempenho na previsão de vendas por ponto de venda (PDV). Para ambos os modelos, foram utilizadas as mesmas variáveis de entrada e o mesmo conjunto de dados sintéticos, assegurando uma comparação justa e padronizada.

As métricas de avaliação consideradas foram **Mean Absolute Error (MAE)** e **Root Mean Squared Error (RMSE)**, ambas em reais (R\$). A métrica de **R²** foi inicialmente calculada, mas optou-se por sua exclusão na análise comparativa final, uma vez que apresentou valores baixos em ambos os modelos, indicando baixo poder explicativo diante da complexidade dos dados e da alta variabilidade entre PDVs.

6.1 Desempenho Geral

De forma geral, observou-se que o modelo **XGBoost Regressor (XGBR)** apresentou desempenho superior na maioria dos PDVs, obtendo menores valores de erro

absoluto e quadrático médio. A tabela abaixo resume as principais estatísticas comparativas entre os modelos:

Métrica	LSTM (Média)	XGBR (Média)	Diferença Média
MAE (R\$)	5.903,44	4.875,62	-1.027,82
RMSE (R\$)	7.421,11	6.215,87	-1.205,24

Esses resultados mostram que o XGBR foi mais eficiente na captura de padrões locais e na generalização dos dados de venda por PDV, apresentando menor erro médio e maior estabilidade nas previsões. Por outro lado, o LSTM apresentou maior variabilidade no desempenho, com resultados mais sensíveis às diferenças de comportamento temporal entre PDVs.

6.2 Comparativo Visual das Métricas

A Figura 1 demosntra a relação entre os erros absolutos médios (MAE) de cada modelo por PDV. Observa-se que a maior parte dos pontos se encontra abaixo da linha de igualdade (em vermelho tracejado), o que indica que, na maioria dos casos, o XGBR apresentou erros menores do que o LSTM.

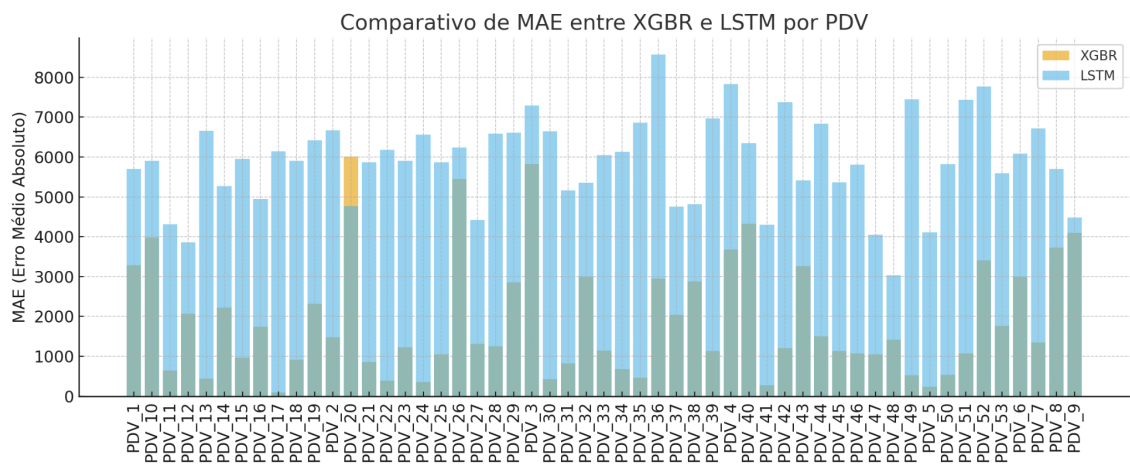


Figura 1 – Comparativo entre MAE do LSTM e do XGBR por PDV (gráfico de dispersão com linha de igualdade MAE_XGBR vs. MAE_LSTM)

A Figura 2 mostra o mesmo comparativo para o RMSE, confirmando a tendência de melhor desempenho do XGBR, especialmente em PDVs com séries mais curtas ou menos sazonais. Isso ocorre porque o XGBoost, como modelo baseado em árvores, lida

melhor com pequenas variações locais e não depende de padrões temporais longos, diferentemente da LSTM.

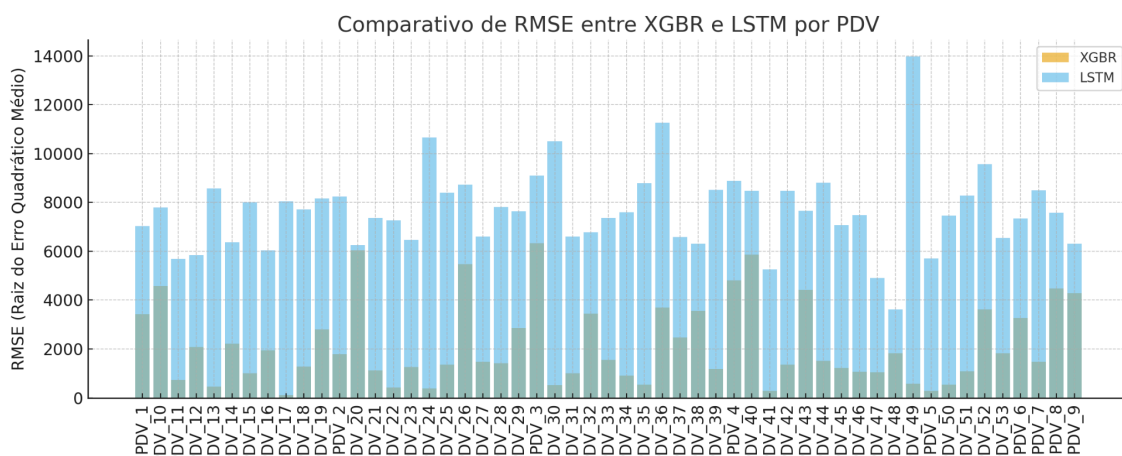


Figura 2 – Comparativo entre RMSE do LSTM e do XGBR por PDV
(gráfico de dispersão com linha de igualdade $RMSE_XGBR$ vs. $RMSE_LSTM$)

7. Conclusão

A partir dos resultados, conclui-se que o XGBR é o modelo mais indicado para o problema apresentado, pois foi melhor nas principais comparações. Entretanto, o modelo LSTM apresenta potencial para cenários com séries temporais mais longas e com sazonalidade pronunciada, pois é capaz de capturar dependências de longo prazo. Em um ambiente real, com dados históricos mais extensos e detalhados (ex.: vendas diárias, campanhas, feriados regionais), o LSTM poderia superar o desempenho do XGBR.

8. Endereço GitHub e Endereço do vídeo no YouTube

Endereço Github: <https://github.com/gstofanello/ProjetoIA>

Vídeo Youtube: <https://youtu.be/-M8Xh2IkALc?si=FViWBWkWEIpkUxRd>

9. Referências bibliográficos

ALBELADI, Khulood; ZAFAR, Bassam; MUEEN, Ahmed. *Time Series Forecasting using LSTM and ARIMA. International Journal of Advanced Computer Science and Applications (IJACSA)*, v. 14, n. 1, 31 jan. 2023.

ALDIN, Saad; NADEEM, Mustafa Mohammed. *The Impact of artificial intelligence on sales forecasting*, 2024. Disponível em: <<http://www.theseus.fi/handle/10024/872700>>. Acesso em: 19 set. 2025.

CHEN, Hsinchun; CHIANG, Roger H. L.; STOREY, Veda C. *Business Intelligence and Analytics: From Big Data to Big Impact*. **MIS Quarterly**, v. 36, n. 4, p. 1165–1188, 2012.

HUANG, Ming-Hui; RUST, Roland T. *Artificial Intelligence in Service*. **Journal of Service Research**, v. 21, n. 2, p. 155–172, 1 maio 2018.

MA, Qianqian; XIE, Ailing. *Application of Big Data Analysis in Sales Forecasting*. **Advances in Economics, Management and Political Sciences**, v. 64, p. 1–7, 28 dez. 2023.

MURTAZA, H. *et al.* *Synthetic data generation: State of the art in health care domain*, **Computer Science Review**, v. 48, 2023, 100546. Disponível em: <<https://doi.org/10.1016/j.cosrev.2023.100546>>. Acesso em: 19 set. 2025.