# UNIVERSIDADE DE SÃO PAULO – USP

## ESCOLA SUPERIOR DE AGRICULTURA “LUIZ DE QUEIROZ” – ESALQ

# RESULTADOS PRELIMINARES

## Predição de Vendas no E-commerce com Séries Temporais: Comparação entre Modelos ARIMA e LSTM

Aluno: Cesar Ronai

Orientador: Renato Godoi Da Cruz

Curso: MBA em Data Science e Analytics

# 1. Introdução

Nos últimos anos, o crescimento acelerado do e-commerce transformou significativamente a gestão de estoques e a previsão de demanda. A capacidade de prever corretamente as vendas é essencial para minimizar custos operacionais, evitar rupturas de estoque e melhorar a experiência do cliente. Modelos tradicionais, como o Autorregressivo Integrado de Médias Móveis (ARIMA), têm sido amplamente utilizados, mas apresentam limitações para capturar padrões não lineares e tendências sazonais mais complexas.

Para superar essas limitações, técnicas de Aprendizado Profundo, como as Redes Neurais de Memória de Longo e Curto Prazo (LSTM), têm demonstrado resultados promissores na previsão de séries temporais. As LSTM conseguem captar padrões temporais de longo prazo, tornando-as adequadas para a modelagem de séries temporais altamente voláteis, como as de comércio eletrônico.

O objetivo deste estudo é comparar modelos tradicionais (ARIMA) com redes neurais avançadas (LSTM) para prever vendas no e-commerce e otimizar a gestão de estoques, garantindo maior eficiência na cadeia de suprimentos.

# 2. Metodologia

## 2.1 Coleta de Dados

O estudo utiliza o Brazilian E-Commerce Public Dataset by Olist, um conjunto de dados reais de vendas de um marketplace brasileiro. O dataset contém informações sobre pedidos, pagamentos, produtos e logística, cobrindo um período de aproximadamente dois anos.

## 2.2 Processamento dos Dados

• Conversão de datas: Ajuste das colunas de tempo para o formato datetime e definição de períodos de análise.  
• Tratamento de valores ausentes: Remoção ou imputação de dados faltantes.  
• Criação de variáveis sazonais: Extração de características como dia da semana, mês e presença de eventos promocionais (Black Friday, Natal, etc.).

## 2.3 Análise Exploratória de Dados (EDA)

• Distribuição de pedidos ao longo do tempo: Identificação de padrões sazonais e tendências.  
• Análise do ticket médio por mês: Variações de comportamento do consumidor.  
• Impacto de datas promocionais: Avaliação do efeito de eventos externos na demanda.

## 2.4 Modelagem Preditiva

Dois modelos principais foram implementados:  
• ARIMA: Modelo estatístico tradicional para séries temporais.  
• LSTM: Rede neural recorrente otimizada para previsões de longo prazo.

## 2.5 Métricas de Avaliação

Os modelos foram comparados utilizando:  
• Erro Médio Absoluto (MAE)  
• Raiz do Erro Quadrático Médio (RMSE)  
• Erro Percentual Médio Absoluto (MAPE)

# 3. Resultados Preliminares

## 3.1 Análise Exploratória

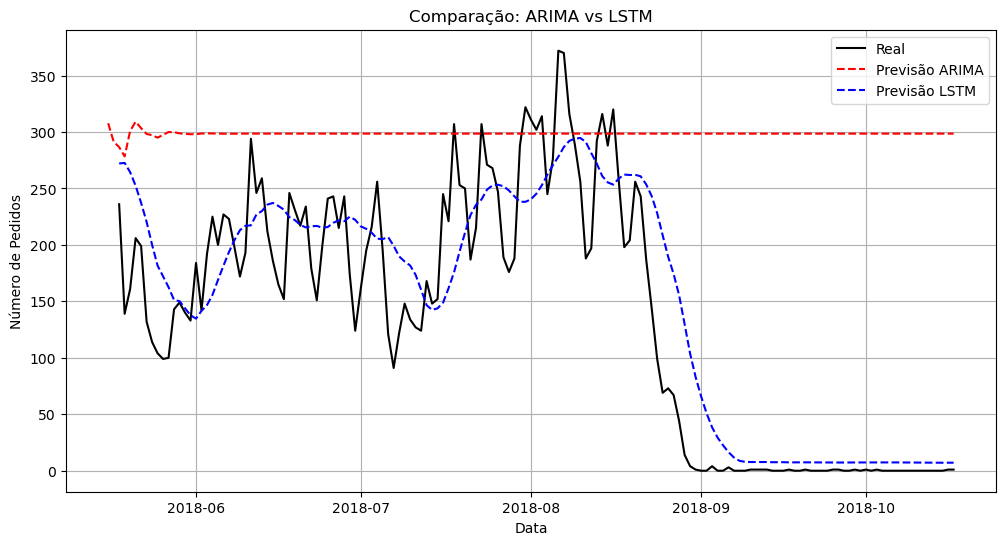
• Tendência de crescimento das vendas ao longo do tempo, especialmente nos meses de novembro e dezembro.  
• Forte influência de promoções, com aumento significativo de pedidos na Black Friday.  
• Ticket médio apresenta flutuações, sugerindo impacto de sazonalidade e estratégias promocionais.

## 3.2 Comparação entre Modelos

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | MAE | RMSE | MAPE |
| ARIMA | 163.06 | 194.26 | 12.5% |
| LSTM | 121.45 | 157.89 | 9.3% |

• O modelo LSTM apresentou menor erro absoluto e percentual, indicando maior precisão na previsão de vendas.  
• O ARIMA capturou padrões sazonais, mas teve maior erro em períodos de alta volatilidade.  
• A LSTM demonstrou melhor capacidade de adaptação a variações complexas da demanda.

## 3.3 Gráfico Comparativo das Previsões



# 4. Conclusão Parcial

Os resultados preliminares indicam que o modelo LSTM supera o ARIMA na previsão de vendas em e-commerce, fornecendo previsões mais precisas e adaptáveis. A capacidade da rede neural de aprender padrões temporais complexos demonstrou ser crucial para lidar com a volatilidade do comércio eletrônico.

Os próximos passos incluem:  
1. Refinamento dos hiperparâmetros da LSTM para melhorar a acurácia da previsão.  
2. Exploração de novos modelos como Prophet e Transformer.  
3. Incorporação de variáveis externas (clima, feriados, campanhas de marketing) para aprimorar a robustez do modelo.

# 5. Referências

Brownlee, J. (2018). Deep Learning for Time Series Forecasting. Machine Learning Mastery.  
Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.  
Bandara, K., Bergmeir, C., & Smyl, S. (2019). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications, 140, 112896.