**Aluno(a):** Cesar Ronai

**Orientador(a):** Renato Godoi Da Cruz

**Curso: MBA em Data Science e Analytics**

**Predição de Vendas no E-commerce com Séries Temporais: Um Estudo de Modelos Preditivos Aplicados à Gestão de Estoque e Demanda.**

**Introdução**

Nos últimos anos, o crescimento acelerado do comércio eletrônico transformou significativamente a forma como produtos e serviços são oferecidos e consumidos. Essa evolução tem imposto desafios substanciais à gestão de estoques e ao atendimento eficiente da demanda dos clientes. A previsão precisa de vendas em plataformas de e-commerce tornou-se, portanto, uma ferramenta essencial para mitigar o risco de falta ou excesso de produtos, o que pode impactar diretamente a eficiência operacional, os custos logísticos e a satisfação do cliente (Brownlee, 2018).

Métodos tradicionais de previsão, como o modelo Autorregressivo Integrado de Médias Móveis (ARIMA), embora úteis, possuem limitações na captura de padrões não lineares e comportamentos complexos que caracterizam o consumidor online moderno. As flutuações rápidas no mercado, influenciadas por fatores sazonais, tendências emergentes e eventos externos, requerem abordagens mais sofisticadas para a previsão de demanda.

Em resposta a essas necessidades, as redes neurais recorrentes, especialmente o modelo de Memória de Longo e Curto Prazo (LSTM), têm emergido como ferramentas promissoras para lidar com dados sequenciais e não lineares. As LSTM são capazes de manter e utilizar informações históricas relevantes a longo prazo, permitindo uma modelagem mais precisa de padrões temporais complexos (Hochreiter & Schmidhuber, 1997). No contexto do e-commerce, a aplicação de LSTM tem demonstrado melhorar significativamente a precisão das previsões, refletindo adequadamente as flutuações sazonais e eventos externos que influenciam o comportamento de compra (Bandara et al., 2019).

Além das LSTM, avanços recentes em técnicas de aprendizagem profunda, como modelos baseados em atenção e redes neurais recorrentes bidirecionais, oferecem novas oportunidades para aprimorar ainda mais a previsão de demanda no comércio eletrônico. Essas técnicas permitem que os modelos considerem tanto o contexto passado quanto influências futuras, proporcionando uma compreensão mais holística das tendências de consumo.

Este estudo tem como objetivo investigar o uso de modelos preditivos avançados, com ênfase nas redes LSTM, para prever a demanda de vendas no e-commerce, visando otimizar a gestão de estoques e aumentar a eficiência na cadeia de suprimentos. A pesquisa busca desenvolver modelos robustos que possam ser aplicados em ambientes de alta variabilidade, onde a precisão nas previsões é crucial para evitar custos elevados associados a excesso ou falta de estoque, além de perdas em oportunidades de venda.

Ao explorar e comparar diferentes métodos modernos de séries temporais e aprendizagem profunda, espera-se contribuir para a implementação de estratégias baseadas em dados para empresas de e-commerce. A adoção dessas técnicas pode levar a uma melhor alocação de recursos, redução de custos operacionais e aumento da satisfação do cliente, posicionando as empresas de forma mais competitiva no mercado global.

Em suma, a integração de modelos preditivos avançados na gestão de estoques representa um passo significativo para o aprimoramento das operações no comércio eletrônico, respondendo efetivamente aos desafios impostos pela dinâmica do mercado atual. A aplicação eficaz dessas ferramentas pode não apenas otimizar processos internos, mas também proporcionar insights valiosos sobre o comportamento do consumidor, permitindo que as empresas se adaptem rapidamente às mudanças do mercado e mantenham uma vantagem competitiva.

**Objetivo**

Investigar e desenvolver modelos preditivos de séries temporais, com ênfase em redes neurais como LSTM, para prever a demanda de vendas no e-commerce, visando otimizar a gestão de estoques e melhorar a eficiência no atendimento à demanda.

**Metodologia**

Este estudo caracteriza-se como uma pesquisa aplicada com abordagem quantitativa, visando desenvolver e avaliar modelos preditivos de séries temporais para a previsão de demanda de vendas no e-commerce. A metodologia adotada envolve a utilização de técnicas de aprendizado de máquina, com ênfase em redes neurais de Memória de Longo e Curto Prazo (LSTM), comparando seu desempenho com modelos tradicionais como o ARIMA.

**Tipo de Pesquisa**

A pesquisa enquadra-se como um **Estudo de Caso**, onde será analisado um conjunto de dados real de vendas de uma empresa de comércio eletrônico. O estudo de caso permite uma compreensão aprofundada do fenômeno em seu contexto real, possibilitando a aplicação prática dos modelos desenvolvidos e a avaliação de sua eficácia na gestão de estoques e demanda.

**Coleta de Dados**

Os dados utilizados neste estudo serão obtidos através de **levantamento de dados secundários** disponibilizados pela empresa de e-commerce em questão. Estes dados englobam históricos de vendas diárias ao longo dos últimos anos, incluindo informações como:

* Quantidade de produtos vendidos por dia.
* Datas e períodos promocionais.
* Dados sazonais (feriados, datas comemorativas).
* Informações sobre campanhas de marketing e eventos externos relevantes.

Para garantir a confidencialidade e atender às normas éticas, não serão divulgados o nome da empresa nem quaisquer detalhes que possam identificá-la.

**Procedimentos de Análise de Dados**

A análise dos dados será realizada em etapas:

1. **Pré-processamento dos Dados**: Consistirá na limpeza, organização e normalização dos dados coletados, preparando-os para a modelagem. Serão tratados valores ausentes, outliers e realizadas transformações necessárias para adequação aos modelos.
2. **Análise Exploratória**: Aplicação de técnicas estatísticas para compreender os padrões presentes nos dados, identificando tendências, sazonalidades e correlações relevantes que possam influenciar a demanda.
3. **Desenvolvimento dos Modelos Preditivos**:
   * **Modelo ARIMA**: Será implementado como representante dos métodos tradicionais de séries temporais, ajustando parâmetros de autorregressão, integração e média móvel conforme necessário.
   * **Rede Neural LSTM**: Será desenvolvida uma arquitetura de rede LSTM adequada para a previsão de séries temporais, ajustando hiperparâmetros como número de neurônios, camadas ocultas, taxa de aprendizado e função de ativação.
4. **Validação e Testes**: Os dados serão divididos em conjuntos de treinamento e teste. A validação cruzada será utilizada para avaliar o desempenho dos modelos, evitando overfitting e garantindo a generalização dos resultados.
5. **Comparação dos Modelos**: Serão utilizadas métricas de avaliação como o Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE) e o Mean Absolute Percentage Error (MAPE) para comparar a precisão dos modelos desenvolvidos.

**Ferramentas e Tecnologias**

Para a implementação e análise dos modelos, serão utilizadas as seguintes ferramentas:

* **Linguagem de Programação Python**: Devido à sua ampla utilização em ciência de dados e disponibilidade de bibliotecas especializadas.
* **Bibliotecas e Frameworks**:
  + **Pandas e NumPy**: Para manipulação e análise de dados.
  + **Matplotlib e Seaborn**: Para visualização gráfica dos dados.
  + **Scikit-learn**: Para técnicas de aprendizado de máquina e pré-processamento.
  + **Statsmodels**: Para implementação do modelo ARIMA.
  + **TensorFlow e Keras**: Para construção e treinamento da rede neural LSTM.

**Ética e Confidencialidade**

O estudo respeita os princípios éticos de pesquisa, garantindo a confidencialidade dos dados e informações da empresa analisada. Todos os dados serão utilizados exclusivamente para fins acadêmicos e de pesquisa, sem divulgação de informações sensíveis ou identificação da empresa.

**Limitações da Pesquisa**

* **Disponibilidade de Dados**: A qualidade e abrangência dos dados históricos podem influenciar os resultados dos modelos.
* **Fatores Externos**: Eventos inesperados, como mudanças econômicas ou sociais, podem afetar a demanda e não serem capturados pelos modelos.

**Considerações Finais**

A metodologia proposta permite uma abordagem estruturada para o desenvolvimento e avaliação de modelos preditivos aplicados ao e-commerce. Espera-se que os resultados obtidos contribuam para a otimização da gestão de estoques e para estratégias mais eficientes de atendimento à demanda, oferecendo insights valiosos para a empresa em estudo e para o setor como um todo.

**Resultados Esperados**

Espera-se que a aplicação dos modelos de séries temporais, como ARIMA e LSTM, no contexto de predição de vendas em e-commerce forneça insights práticos para melhorar a gestão de estoque e as estratégias de marketing. A partir da previsão mais precisa da demanda, o estudo visa minimizar ocorrências de excesso e falta de produtos, otimizando o fluxo de mercadorias e a alocação de recursos, além de aprimorar a satisfação do cliente.  
  
 A expectative é que o modelo LSTM, com capacidade de capturar padrões de longo prazo e identificar sazonalidades complexas, supere os modelos tradicionais em cenários de alta variabilidade. Dessa forma, o estudo poderá fornecer uma base robusta para decisões em tempo real, oferecendo uma vantagem competitiva significativa para as empresas de comércio eletrônico.

**Cronograma de Atividades**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Projeto de Pesquisa | Período | | | | | | | | | |
| 31 de outubro a 07 de novembro de 2024 |  |  |  |  |  |  |  |  |  |
| Resultados Preliminares | 30 de janeiro a 06 de fevereiro de 2025 |  |  |  |  |  |  |  |  |  |
| Entrega do TCC e Agendamento da Defesa | 15 a 29 de abril de 2025 |  |  |  |  |  |  |  |  |  |
| Defesa (datas a definir) | Previsão de junho a agosto de 2025 |  |  |  |  |  |  |  |  |  |

# 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.

Kotler, P., & Keller, K.L. (2012). Administração de Marketing. 14ª ed. Pearson Education do Brasil, São Paulo, SP, Brasil.

Corrêa, H.L., & Corrêa, C.A. (2017). Administração de Produção e Operações. 4ª ed. Atlas, São Paulo, SP, Brasil.

Project Management Institute [PMI] (2013). Um Guia do Conhecimento em Gerenciamento de Projetos. 5ª ed. Project Management Institute, Newtown Square, PA, EUA.

Yin, R. K. (2015). Estudo de Caso: Planejamento e Métodos. Bookman, Porto Alegre, RS, Brasil.

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.

Brownlee, J. (2018). Deep Learning for Time Series Forecasting. Machine Learning Mastery.

Corrêa, H.L., & Corrêa, C.A. (2017). Administração de Produção e Operações. 4ª ed. Atlas, São Paulo, SP, Brasil.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

Kotler, P., & Keller, K.L. (2012). Administração de Marketing. 14ª ed. Pearson Education do Brasil, São Paulo, SP, Brasil.

Project Management Institute [PMI] (2013). Um Guia do Conhecimento em Gerenciamento de Projetos. 5ª ed. Project Management Institute, Newtown Square, PA, EUA.

Yin, R. K. (2015). Estudo de Caso: Planejamento e Métodos. Bookman, Porto Alegre, RS, Brasil.