

**CENTRO UNIVERSITÁRIO ESTÁCIO
POLO EAD GOIANA – PE**

RELATÓRIO TÉCNICO DO CURSO DE CIÊNCIA DE DADOS

**Modelo Preditivo para Crescimento de Receita da Magazine Luiza
(MGLU3): Uma Análise Crítica da Extrapolação Linear no E-commerce.**

LUCAS EDUARDO DE PONTES OLIVEIRA

PROF. ME. BRUNO MARTINS BARTOLOMEU

NOVEMBRO 2025

SUMÁRIO

1. CONTEXTO E OBJETIVO	3
2. METODOLOGIA	4
2.1. Aplicação da metodologia	4
2.1.1. Preparação dos dados.....	4
3. MODELAGEM PREDITIVA.....	7
3.1. Relacionamento chave para o E-commerce: MGLU3 vs. economia.....	7
3.2. Relacionamento interno: saúde financeira da MGLU3.....	8
3.3. Construção do dataset final para modelagem	8
3.4. Modelagem preditiva (Regressão Linear Múltipla).....	9
3.4.1. Projeção até 2030.....	10
4. DISCUSSÃO, DIAGNÓSTICO E LIMITAÇÕES DO MODELO.....	11
4.1. Resultado final do modelo de Regressão Linear.....	11
5. VALIDAÇÃO DO MODELO NÃO-LINEAR E RECOMENDAÇÕES FINAIS.....	12
5.1 Validação com o Modelo Random Forest.....	12
5.2. Recomendações Finais	13
6. REFERÊNCIAS.....	14
7. APÊNDICE.....	15

1. CONTEXTO E OBJETIVO

Este relatório técnico documenta as primeiras etapas do projeto de Previsão do Crescimento Anual da Receita Líquida da Magazine Luiza (MGLU3) para o horizonte de 2020 a 2030. O trabalho segue a metodologia CRISP-DM e utiliza dados históricos de 2011 a 2019 da MGLU3, combinados com indicadores macroeconômicos brasileiros de 2010 a 2020, para construir um modelo preditivo.

O objetivo principal é avaliar a sensibilidade do crescimento da receita do MGLU3, verificando variáveis internas (Margem Bruta, Retorno sobre o Patrimônio Líquido, Giro de Ativos) e externas (PIB, Consumo das Famílias), por fim, projetar o crescimento futuro.

2. METODOLOGIA

O CRISP-DM como a metodologia mais utilizada para projetos de Data Mining (Mineração de Dados) ou Data Science (ciência de Dados), foi o escolhido para este projeto.

2.1. Aplicação da metodologia

Foi feita a escolha dos dados do balanço do e-commerce Magazine Luiza, e para dar maior contexto e assertividade foi pego dados dos indicadores econômicos nacional.

Como ferramenta de manipulação e criação de scripts, foi utilizado a linhagem python.

2.1.1. Preparação dos dados

Foi realizando a padronização dos arquivos, convertendo-os em formato CSV, realizado o cálculo dos indicadores financeiros-chave de base trimestral para base anual realizando a padronização e granularidade temporal com os indicadores macroeconômicos e garantir a correta correlação na modelagem preditiva, e finalizado com a limpeza de dados, tratando de remover os valores vazios.

ANTES

The diagram illustrates the initial state of the data preparation process. It features three text labels positioned above a screenshot of a spreadsheet:

- "Campos vazios" (Empty fields) points to the first row of the spreadsheet, which contains metadata: "XLSWrite 3.34 Copyright(c) 1999,2000 Axolot Data".
- "Informações desnecessárias" (Unnecessary information) points to the header row of the spreadsheet, which includes columns labeled A, B, C, and D, and a header "Balanço Patrimonial - MAGAZ LUIZA".
- "Data trimestral (fora do padrão usado)" (Trimestral data (out of the standard used)) points to the dates in the B column: "30/06/2020", "31/03/2020", and "31/12/2019".

Screenshot Description: The screenshot shows a table with 5 rows and 4 columns. The columns are labeled A, B, C, and D. Row 1 contains the text "XLSWrite 3.34 Copyright(c) 1999,2000 Axolot Data" in column A and "Balanço Patrimonial - MAGAZ LUIZA" in column C. Row 2 contains the date "30/06/2020" in column B and "31/03/2020" in column C. Row 3 contains the text "Ativo Total" in column A and the value "20.558.633" in column B. Row 4 contains the text "Ativo Circulante" in column A and the value "13.278.766" in column B. Row 5 contains the text "Caixa e Equivalentes de Caixa" in column A and the value "1.103.523" in column B. The values in columns C and D are partially cut off at the right edge of the screenshot.

	A	B	C	D
1	XLSWrite 3.34 Copyright(c) 1999,2000 Axolot Data	30/06/2020	Balanço Patrimonial - MAGAZ LUIZA	31/12/2019
2		31/03/2020		
3	Ativo Total	20.558.633	17.710.082	19.791.071
4	Ativo Circulante	13.278.766	10.589.868	12.841.161
5	Caixa e Equivalentes de Caixa	1.103.523	388.904	305.746

ANTES

Valores zerados



9	Ativos Biológicos		0	0	0	0	0
10	Tributos a Recuperar	748.944	877.448	864.144	745.693		
11	Despesas Antecipadas	0	0	0	0	0	
12	Outros Ativos Circulantes	963.199	869.713	506.316	384.307		
13	Ativo Realizável a Longo Prazo	1.337.940	1.861.255	1.748.703	1.856.321		
14	Aplicações Financeiras Avaliadas a Valor Justo	0	0	214	264		
15	Aplicações Financeiras Avaliadas ao Custo Amortizado	0	0	0	0		
16	Contas a Receber	0	0	0	0		

Planilha em PDF →

Tabela 2. Taxas de crescimento do PIB (2011/2020).

Anos	Variação real	PIB per capita
2011	3,97	3,07
2012	1,92	1,04
2013	3,00	2,13
2014	0,50	-0,35
2015	-3,55	-4,38
2016	-3,28	-4,07
2017	1,32	0,52
2018	1,78	0,96
2019	1,22	0,40
2020	-3,88	-4,60
Média anual	0,26	-0,56
Crescimento acumulado	2,67	-5,50

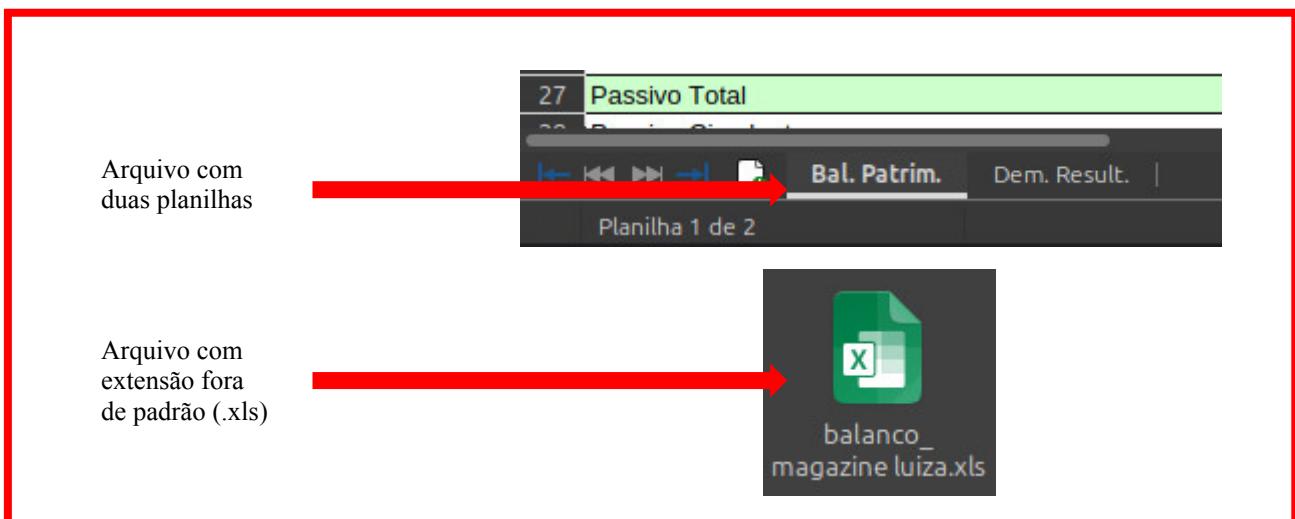
Fonte: IBGE (2022c; 2022d).

DEPOIS

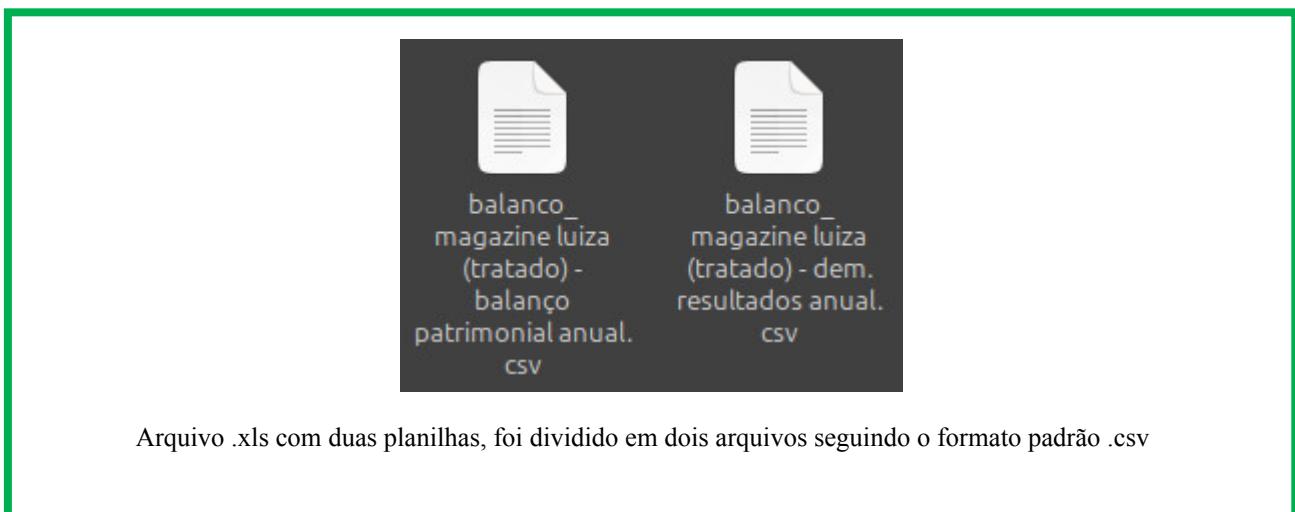
A	B	C	D	E	F	G	H	I	J	
1	Ano Vigente	2019	2018	2017	2016	2015	2014	2013	2012	2011
2	Receita Líquida de Vendas e/ou Serviços	19886309	15590444	11984250	9508746	8978259	9779385	22172922	7665112	6419371
3	Custo de Bens e/ou Serviços Vendidos	-14332350	-11053022	-8378239	-6586130	-6399630	-7086909	-15247125	-5146597	-4275084
4	Resultado Bruto	5553961	4537422	3606011	2922615	2578629	2692476	6925797	2518515	2144287
5	Despesas Com Vendas	-3444112	-2747447	-2119953	-1776258	-1720799	-1746258	-4365789	-1581049	-1270971
6	Despesas Gerais e Administrativas	-1188562	-759833	-679085	-615545	-584280	-556882	-1429211	-481965	-441566
7	Perdas pela Não Recuperabilidade de Ativos	-75993	-59737	-41921	-26074	-30462	-22547	-604662	-338519	-244931
8	Outras Receitas Operacionais	416662	53389	36486	13505	20233	24519	156292	31283	26841
9	Resultado da Equivalência Patrimonial	26607	57757	86156	62702	75605	99620	54464	0	0
10	Financeiras	-70413	-294688	-410821	-503849	-486055	-360742	-581982	-172297	-165723
11	Receitas Financeiras	640917	133929	110107	116655	130297	96469	176779	55695	51686
12	Despesas Financeiras	-711330	-428617	-520928	-620504	-616352	-457211	-758761	-227992	-217409
13	Resultado Antes Tributação/Participações	1218150	786863	476873	77096	-147129	130186	154909	-24032	47937
14	Lucro/Prejuízo do Período	921828	597429	389022	86565	-65605	128556	118727	-6745	11666
15										

Planilha estruturada de forma coesa com data em formato anual, linhas com valores zero e informações irrelevantes excluídos, e padronizado dados em formato .csv

ANTES



DEPOIS



3. MODELAGEM PREDITIVA

A etapa de modelagem iniciou-se com a escolha da aplicação de um modelo de Regressão Linear Múltipla para estabelecer as correlações preditivas, sedo os principais motivos de sua escolha, o fato deste modelo ser referência em projetos de previsão, tem uma formulação matemática fechada e extremamente eficiente. Sendo perfeito para a fase de modelagem do CRISP-DM, pois permite que a equipe de análise realize rapidamente os primeiros testes, validando as hipóteses de correlação e a avaliação da qualidade do dataset preparado. Também tendo em mente que a teoria econômica e financeira básica sugere que o crescimento da receita de uma empresa tem uma relação linear com indicadores macroeconômicos e indicadores internos.

3.1. Relacionamento chave para o E-commerce: MGLU3 vs. economia

O primeiro script gera um plotagem a ser utilizada na correlação entre o crescimento anual da receita da MGLU3 (Magazine Luiza), versus o consumo familiar no mesmo período.

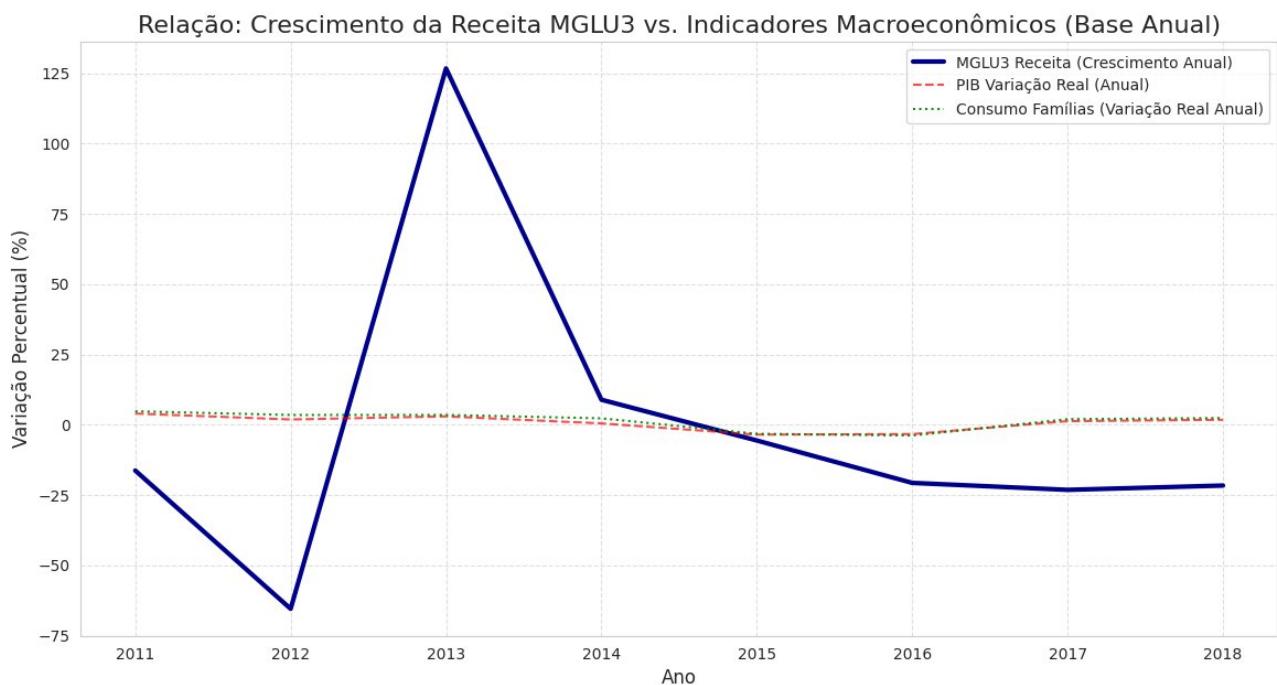


GRÁFICO 1

O gráfico ilustra a divergência do crescimento da receita da Magazine Luiza em relação aos indicadores macroeconômicos tradicionais no período de 2011 a 2014, o que sugere que o impulso do crescimento foi determinado por fatores internos ou específicos do setor.

3.2. Relacionamento interno: saúde financeira da MGLU3

O segundo script gera uma plotagem para nos mostrar a saúde financeira da MGLU3 (Magazine Luiza), observando o Retorno Sobre Ativos (ROA), da empresa e sua liquidez corrente.

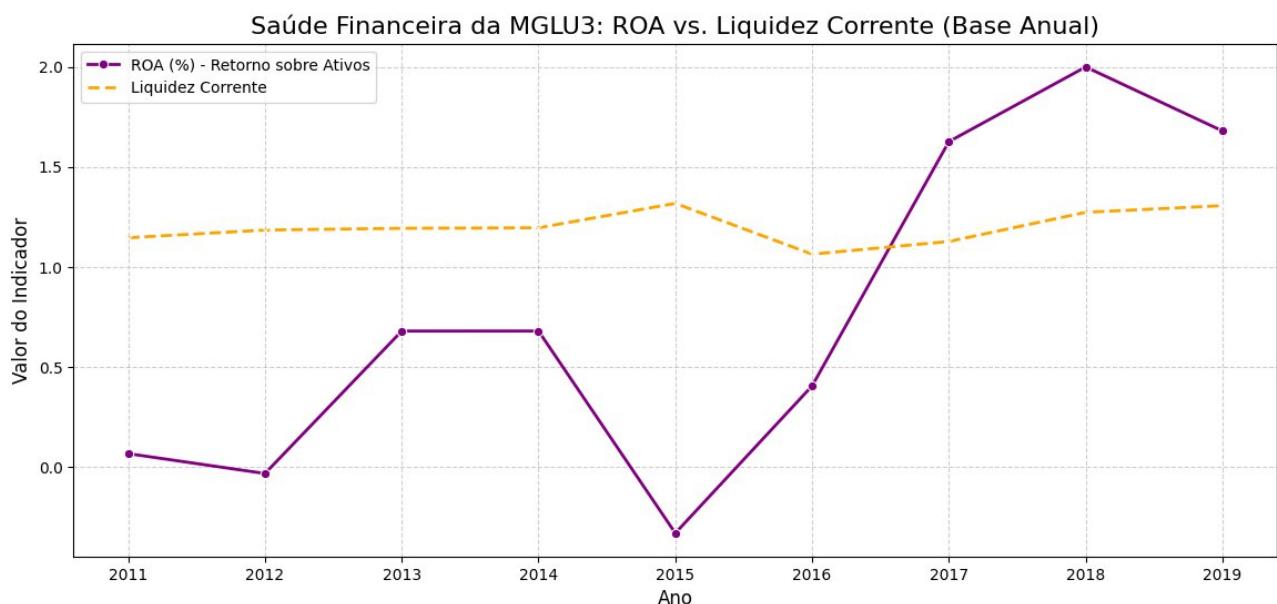


GRÁFICO 2

O gráfico mostra a MGLU3 saindo de uma crise de rentabilidade em 2015 para se tornar significativamente mais eficiente e lucrativa até 2018/2019, enquanto manteve sua saúde financeira (Liquidez Corrente) em um patamar seguro.

3.3. Construção do dataset final para modelagem

O terceiro script é responsável pelo processo de Extração, Transformação e Carga (ETL), culminando na exportação de um dataset unificado e pré-processado para a modelagem preditiva (Machine Learning/IA).

O processo integra três fontes de dados anuais – Demonstrativos Financeiros (DRE e BP da MGLU3) e Indicadores Macroeconômicos (PIB e Consumo das Famílias) – em um único DataFrame.

3.4. Modelagem preditiva (Regressão Linear Múltipla)

O quarto script gera a plotagem da representação visual da performance do modelo de Regressão Linear, observando o crescimento da receita e criando um comparativo do modelo real e preditivo.

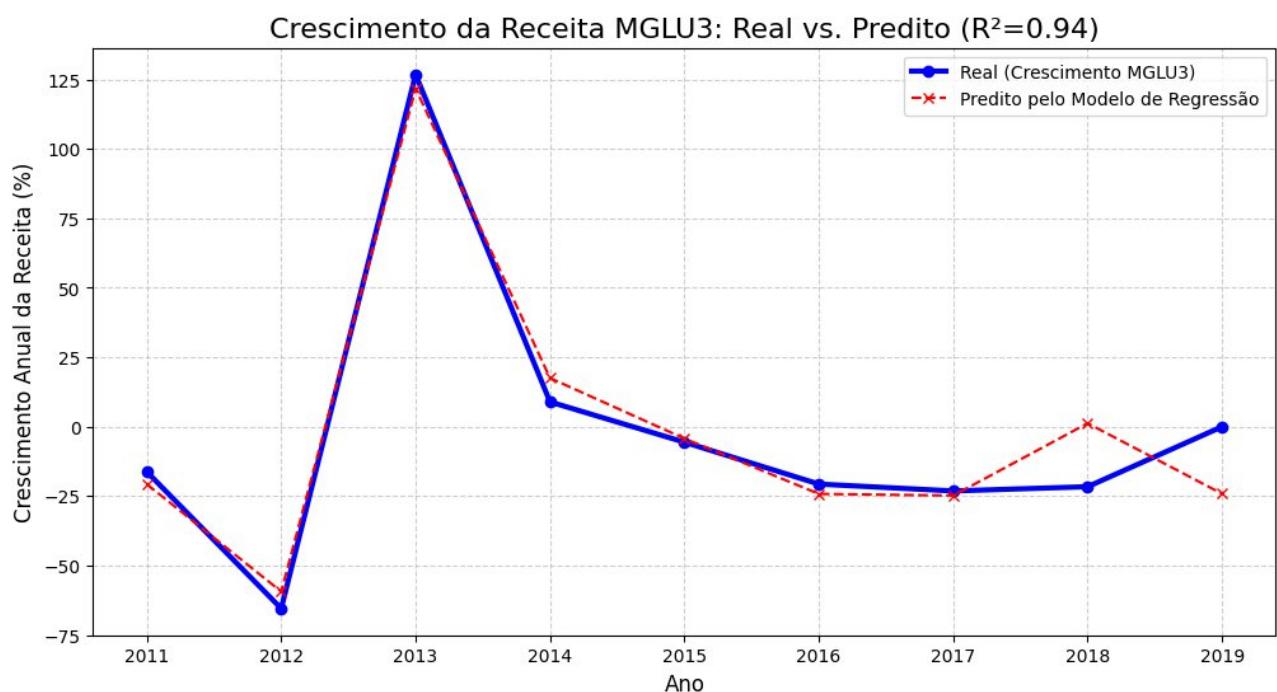


GRÁFICO 3

O gráfico demonstra que o modelo de Regressão Linear, utilizando os indicadores internos e macroeconômicos, é extremamente robusto e explicativo para a variável alvo "Crescimento Anual da Receita da MGLU3", com um poder de explicação de 94%.

3.4.1. Projeção até 2030

O quinto script gera a plotagem da tendência futura do crescimento da receita até 2030.

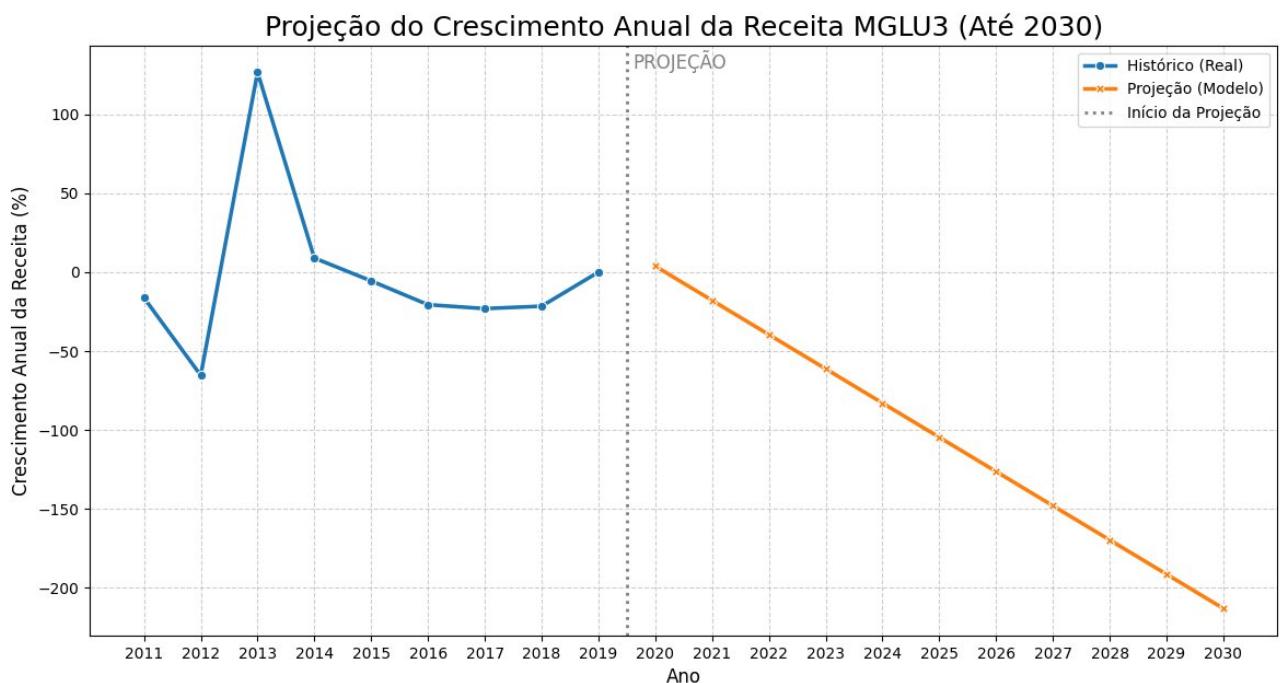


GRÁFICO 4

O gráfico de "Projeção do Crescimento Anual da Receita MGLU3 (Até 2030)" ilustra a previsão do modelo de Regressão Linear desenvolvido, contrastando o histórico real com o cenário futuro, também nos mostra uma tendência de crescimento anual caindo drasticamente, chegando a -200% em 2030, o que é um resultado irreal para qualquer cenário de negócios.

4. DISCUSSÃO, DIAGNÓSTICO E LIMITAÇÕES DO MODELO

Tendo em vista o objetivo inicial de avaliar a sensibilidade do crescimento da receita do MGLU3 e projetar o crescimento futuro, vemos que o objetivo foi alcançado, com algumas ressalvas.

4.1. Resultado final do modelo de Regressão Linear

O Gráfico 4 ("Projeção do Crescimento Anual da Receita MGLU3 (Até 2030)'), mostra uma resultado irreal tendo em vista o histórico da empresa.

Apesar de ter um excelente desempenho no período histórico (treinamento) com um coeficiente de determinação (R^2), com poder de explicação de 94%, existe uma clara falha na projeção futura, sendo a causa Provável o modelo de regressão linear. Sendo ela linear, ao ter uma série histórica (2011–2019) possuindo pontos extremos (picos e vales) e uma tendência geral de alta nos últimos anos, o modelo pode ter ajustado a inclinação baseada nos dados iniciais, levando a uma extrapolação insustentável. O modelo não consegue capturar a não-linearidade e os limites de crescimento de uma empresa.

5. VALIDAÇÃO DO MODELO NÃO-LINEAR E RECOMENDAÇÕES FINAIS

O crescimento de uma empresa de e-commerce é um fenômeno não-linear. O crescimento é exponencial em certas fases, mas depois se estabiliza (Lei dos Grandes Números). Seguindo este conhecimento e tendo em vista a falta de confiança na geração futura gerada, o indicado é buscar modelos alternativos.

As soluções envolvem mudar o foco para modelos mais robustos, como, séries temporais, dados não-lineares e modelos de curvas de crescimento. Alguns exemplos de modelos a se seguir com o projeto são, ARIMA/SARIMA, Random Forest Regressor, XGBoost e Prophet.

A recomendação mais forte para o novo modelo é o Random Forest Regressor, para superar a limitação da extração do modelo linear inicial.

5.1 Validação com o Modelo Random Forest

Ao seguir a recomendação de um novo modelo foi usado o Random Forest Regressor, gerando uma nova plotagem.

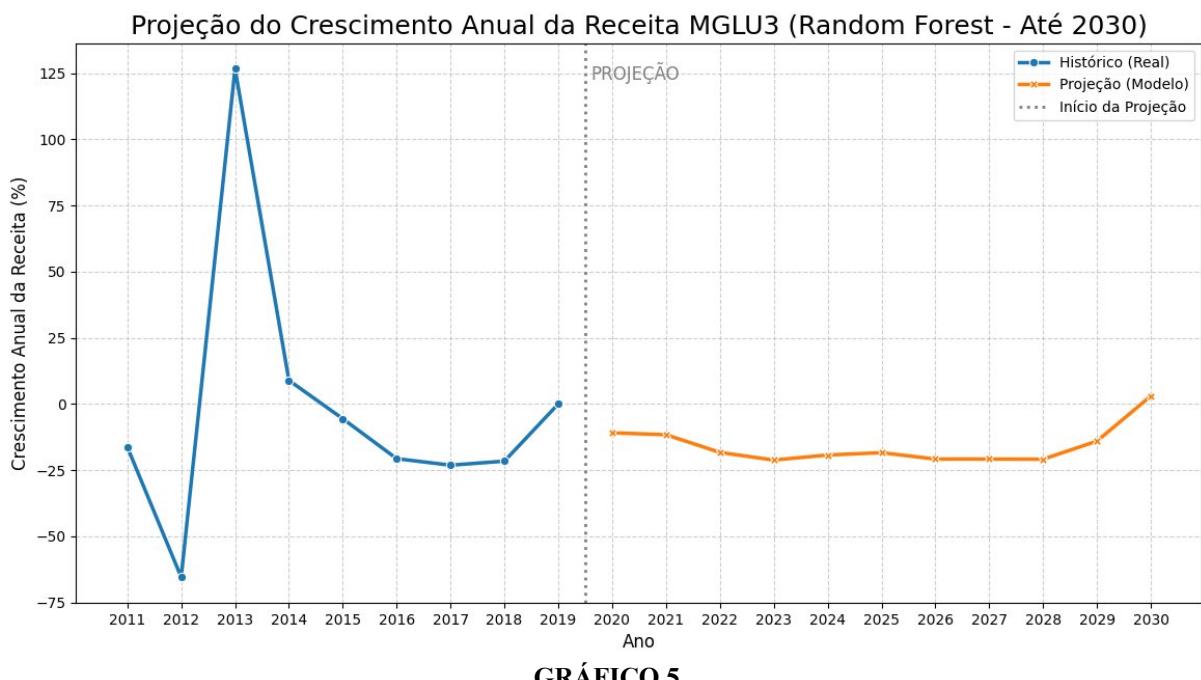


GRÁFICO 5

O novo gráfico gerado mostra uma previsão mais realística, sendo uma projeção de crescimento anual com desaceleração no curto prazo (2020-2027), tendendo a se estabilizar e iniciar uma recuperação controlada até 2030.

Diferente da regressão linear, agora se tem uma previsão realisticamente sustentável para o e-commerce brasileiro.

5.2. Recomendações Finais

Comprovada a efetividade do modelo não-linear na análise e previsibilidade de dados do e-commerce, a sugestão para análises futuras e mais robustas é o uso de múltiplos modelos não-lineares (como XGBoost ou Prophet), conferindo as semelhanças entre suas projeções.

Além disso, é crucial estabelecer uma rotina de validação contínua dos modelos com dados reais adquiridos anualmente, garantindo a adaptação do modelo às novas dinâmicas macroeconômicas e de mercado.

6. REFERÊNCIAS

Henrique Castro Martins Dataverse. **Balace sheet of Magazine Luiza S.A.** Harvard Dataverse. ago, 2023. Disponível em:
<https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/U6GZYM/13OAR2&version=1.0>. Acesso em: 6 out. 2025.

Ana Cristina Lima Couto, e Joaquim Miguel Couto. **A NOVA “DÉCADA PERDIDA” DO BRASIL (2011-2020)**. Portal de periódicos. Hoje, dez, 2021. Disponível em:
<https://periodicos.uem.br/ojs/index.php/EconRev/article/download/63224/751375154146/>. Acesso em: 6 out. 2025.

7. APÊNDICE

REPOSITÓRIO GITHUB DO PROJETO: Modelo-Preditivo-para-Crescimento-de-Receita-da-Magazine-Luiza-MGLU3

Este apêndice detalha a estrutura do repositório público do projeto, onde estão armazenados todos os códigos-fonte, dados e notebooks utilizados para a análise e modelagem preditiva apresentadas neste relatório.

A íntegra do código e os arquivos de dataset podem ser acessados através do link principal do repositório:

Link de Acesso:

<https://github.com/elucas8013/Analise-Critica-da-Extrapolacao-Linear-no-E-commerce>

