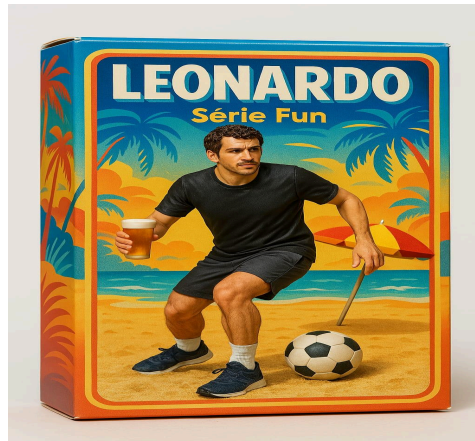


Trabalho Machine Learning para Negócios

MBA em IA, Data Science e Big Data para Negócios
Ibmec 2025.1

Grupo:

Leonardo Pazo - Matrícula: 2025.03.52088-7



Rafaela Filardi - Matrícula: 2025.03.52088-7

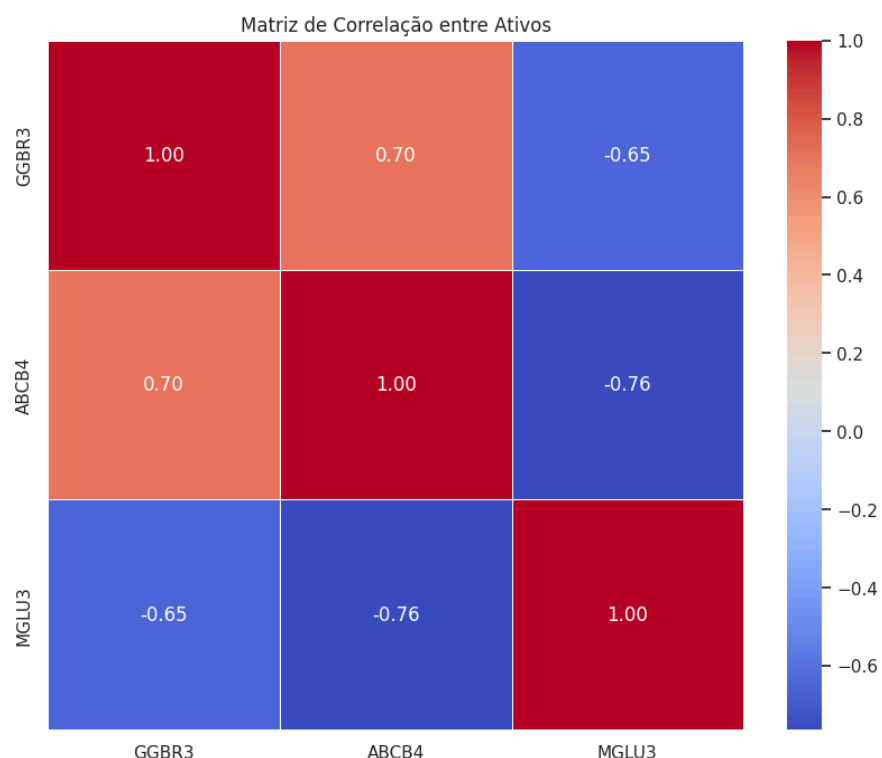


Silvio Mota - Matrícula: 2025.03.52728-8



Para realização do trabalho, escolhemos 3 empresas nacionais para analisar seus ativos financeiros com Regressão Linear. Para isso, escolhemos a Gerdau (GGBR3), empresa que opera no setor de siderurgia, sendo uma importante produtora de aço, ligada ao setor de *commodities*, seu desempenho pode ser influenciado por fatores macroeconômicos, preço de *commodities* e a demanda do setor de construção civil e indústria. A segunda empresa escolhida foi o Banco ABC Brasil (ABCB4), importante instituição do setor financeiro e seu desempenho pode ser influenciado por taxas de juros, crescimento econômico, concessão de crédito e regulação do setor financeiro. Por fim, escolhemos a Magalu (MGLU3), uma empresa do setor de varejo, com forte presença no *e-commerce* e que tem seu desempenho influenciado por fatores como consumo das famílias, taxas de juros que afetam o crédito do consumidor, concorrência em seu setor e inovações tecnológicas.

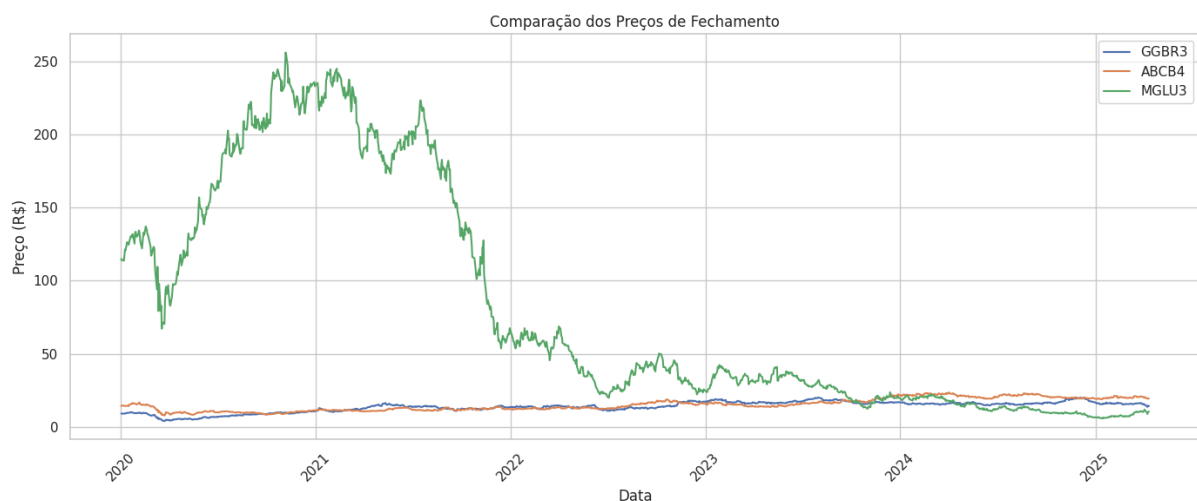
O critério de escolha para esses ativos foi a busca de diversificação setorial (*commodities*, financeiro e varejo), possibilitando uma análise mais crítica, além da relevância desses ativos no país. Somado a esses fatores, eles possuem um bom histórico de dados, acima de 5 anos, e volume diário acima de R\$1 milhão. Realizamos também a Matriz de Correlação entre ativos.



Utilizamos a regressão linear para prever o comportamento de ativos financeiros, esse modelo prevê a existência de uma relação linear entre as variáveis dependentes e independentes, porém existem riscos e limitações pois quando observamos fenômenos financeiros não necessariamente eles ocorrem dessa forma.

Além disso, se as variáveis independentes estiverem correlacionadas entre si, isso pode afetar a estabilidade do modelo e podemos não estar incluindo todas as variáveis relevantes que podem afetar o preço dos ativos, levando a uma previsão incompleta.

Quando trabalhamos com *Machine Learning*, temos também que estar atentos ao *Overfitting* e *Underfitting*, sendo assim, o modelo pode ter um sobreajuste ou sub ajuste em seu treinamento.



O Banco ABC e a Gerdau têm uma correlação histórica um pouco mais alta, pois são ativos com menor volatilidade, sendo um investimento mais seguro comparado a Magalu, que tem uma variância de preço elevada.

No gráfico acima, podemos ver essa diferença na volatilidade dos preços de fechamento dos três ativos.

Podemos verificar que quem investiu em 2020 na Magazine Luiza, dependendo do perfil provavelmente chegando em 2021, vendeu as ações porque, a partir dos meados de 2021 para 2022 a queda foi acentuada e provavelmente muitos tiveram um prejuízo considerável. O investidor que preferiu apostar no Banco ABC ou Gerdau, manteve um platô de valor nas ações.

Apesar da forte queda, a Magazine Luiza pode dar confiança para que mesmo com esse acontecimento, os investidores mais arrojados possam aproveitar o momento e comprar mais ações para ganhar lucros mais significativos, pois é uma empresa sólida. Enquanto os investimentos na Gerdau e no Banco ABC são mais conservadores, pois, não

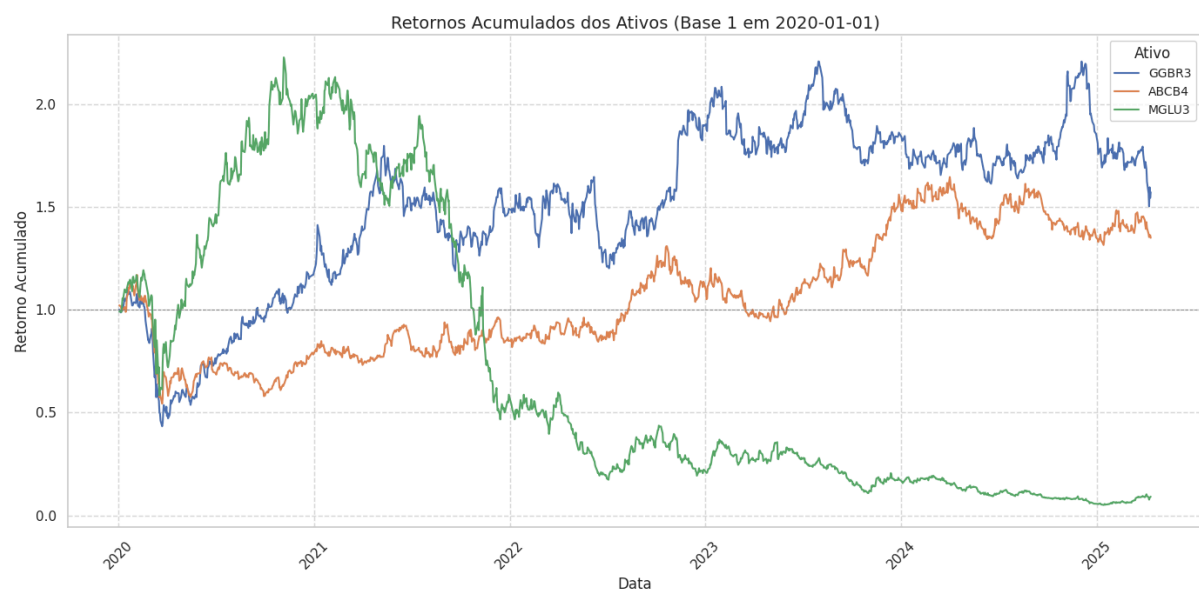
garantem quedas ou aumentos significativos, comparando inclusive com os investimentos mais simples como renda fixa e poupança.

Retorno dos ativos:

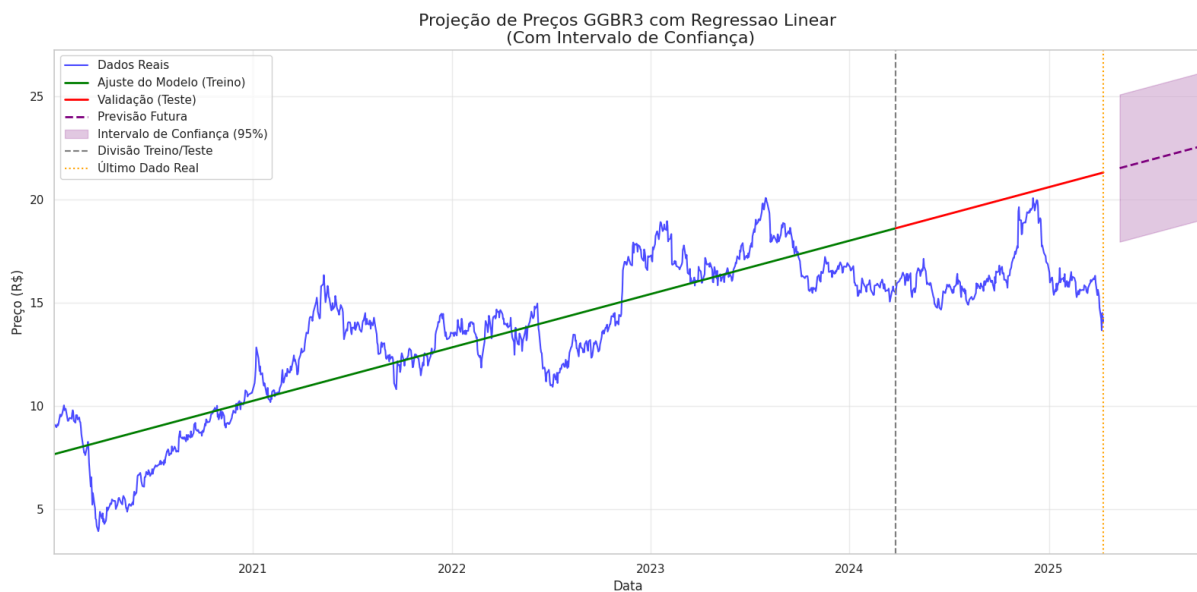
GGBR3 - Retorno total: 57.20%

ABCB4 - Retorno total: 35.03%

MGLU3 - Retorno total: -90.90%



O ativo com maior retorno no período foi Gerdau com 57,20%. Por isso, escolhemos ela para realizar as análises, utilizando regressão linear.



Embora a regressão linear possa fornecer insights iniciais e uma base para a modelagem, é importante ser cauteloso ao interpretá-la e usá-la para previsões financeiras. Considerar técnicas adicionais, como modelos não lineares, modelos de séries temporais (como ARIMA), ou técnicas de aprendizado de máquina, pode ajudar a capturar padrões mais complexos e melhorar a precisão das previsões. Além disso, sempre é bom complementar a análise com uma compreensão sólida dos fundamentos econômicos e do mercado.

Segundo o gráfico, o teste do modelo fica acima dos dados reais, não capturando a queda que existe no início de 2025.

O modelo pode ser útil para investimentos em um período mais de curto prazo, mas ainda assim não é perfeito. Ele falha em períodos de alta volatilidade explicados por eventos internos e externos, como pandemia, eleições, alta concorrência e taxa de juros. Em 2020, por exemplo, as ações tiveram uma grande queda, resultado do início da pandemia do Covid-19. Assim como em 2021 começou a ter uma recuperação deste vale. Entre 2022 e 2023 também tivemos o início da guerra entre Ucrânia e Rússia e também a guerra entre Israel e o Hamas. Além de que no final de 2022 tivemos as eleições presidenciais, que é mais um fator de altíssima complexidade para o mercado, que aguarda essa decisão e em cima dela define o futuro da economia.

A regressão linear é útil para identificar tendências gerais, mas possui algumas limitações:

- Assunção de linearidade: Pressupõe que as relações entre variáveis são lineares, o que pode não capturar de forma adequada dinâmicas mais

complexas ou não lineares dos preços dos ativos e isso pudemos observar nos gráficos dos modelos.

- Sensibilidade a outliers: A Magalu por exemplo obteve valores extremos que distorcem os resultados da regressão linear, causando desconfiança.
- Retrospectividade: A análise foi baseada em dados históricos e pode não prever corretamente alguns movimentos no futuro, especialmente em mercados que tenham volatilidade.

Portanto, para se obter uma análise mais robusta dos modelos, recomenda-se que façamos combinações entre outros como regressões não lineares, outras técnicas de *Machine Learning* ou indicadores técnicos adaptados às características de cada ativo. Além disso, podem ser incluídas mais variáveis determinantes no preço de fechamento dos ativos, como taxa de juros e inflação.

Explicando sobre o R^2 , a variação entre os ativos, não é significativa para a empresa Gerdau, como pode ser verificado, visto que seus preços não possuem altas mudanças de patamar, diferente do que aconteceria com a Magazine Luiza, que tem uma mudança significativa de preços principalmente na virada entre 2022 e 2023, o que afeta a confiança no modelo.

Sobre as métricas RMSE e MAE, elas indicam o desempenho do modelo no trabalho que consiste em realizar a previsão dos preços, seja em R\$ ou em USD, conforme a indicação.

Avaliando as métricas abaixo:

RMSE (Erro Quadrático Médio) - 3,90

O RMSE mede o erro médio das previsões do modelo, penalizando de forma mais significativa, grandes desvios devido a elevação ao quadrado dos erros. Um RMSE de 3,90 significa que em média, as previsões do modelo difere do valor real em aproximadamente 3,90 unidades monetárias. Valores menores indicam melhor desempenho do modelo.

MAE (Erro Absoluto Médio) - 3,66

O MAE mede o erro absoluto entre as previsões e os valores reais, sem penalizar tanto os grandes desvios quanto o RMSE. Um MAE de 3,66 indica que, em média, o erro absoluto das previsões é de 3,66 unidades monetárias. Bem como o RMSE, quanto menor o MAE, melhor o desempenho do modelo.

Interpretando a análise acima, a proximidade entre o RMSE (3,90) e o MAE (3,66) sugere que não há muitos *outliers* ou erros extremos nas previsões, pois, quando há erros significativos, o RMSE tende a ser de forma relevante maior que o MAE.

Os valores absolutos dessas métricas devem ser avaliados em relação à escala dos dados. Se os preços previstos estão em uma faixa muito maior que 3,90, esses erros podem ser considerados baixos, mas se acontecer ao contrário, podem indicar necessidade de melhorias no modelo.

CONSIDERAÇÕES FINAIS

O modelo de regressão linear apresentou bom desempenho nos dados de treinamento, mas falhou em generalizar para os dados de teste, como evidenciado pelo aumento dos erros e pelo R^2 negativo. Isso significa que o modelo está super ajustado e precisa ser revisado para melhorar sua capacidade preditiva em dados novos. É recomendável explorarmos ajustes no modelo e comparar com abordagens alternativas para obter melhores resultados em produção.