

FATEC RUBENS LARA - BAIXADA SANTISTA

CURSO DE CIÊNCIA DE DADOS

3º ciclo

**ANÁLISE DE AÇÕES BRASILEIRAS
UTILIZANDO REGRESSÃO LINEAR COM
GRADIENTE DESCENDENTE: UM
ESTUDO EMPÍRICO DAS EMPRESAS
ESTATAIS PETR4.SA, BBAS3.SA E
ELET3.SA**

Eduardo Miguel Ribeiro Cordeiro

**Brasil
2025**

RESUMO

Este trabalho apresenta uma análise quantitativa de três ações brasileiras do setor estatal: Petrobras (PETR4.SA), Banco do Brasil (BBAS3.SA) e Eletrobras (ELET3.SA), utilizando técnicas de regressão linear implementadas através do algoritmo de gradiente descendente. O objetivo principal foi desenvolver um modelo preditivo capaz de identificar tendências de preços e realizar previsões de curto prazo. Foram analisados 251 dias de dados históricos de fechamento, compreendendo o período de um ano. A metodologia incluiu a implementação manual do algoritmo de gradiente descendente para otimização dos parâmetros da regressão linear, cálculo de métricas de qualidade do modelo através do coeficiente de determinação (R^2), e análise estatística descritiva completa. Os resultados indicaram que a ELET3.SA apresentou a maior capacidade preditiva ($R^2 = 0,2491$), seguida pela PETR4.SA ($R^2 = 0,0929$) e BBAS3.SA ($R^2 = 0,0433$). As análises revelaram tendência de baixa para PETR4.SA (-0,008189 por dia) e tendências de alta para BBAS3.SA (+0,004695 por dia) e ELET3.SA (+0,020870 por dia). O estudo demonstra a aplicabilidade de técnicas de machine learning na análise do mercado financeiro brasileiro e fornece insights quantitativos para tomada de decisões de investimento.

Palavras-chave: Regressão Linear. Gradiente Descendente. Análise Financeira. Mercado de Ações. Aprendizado de Máquina.

Sumário

1	INTRODUÇÃO	3
2	METODOLOGIA	3
2.1	Problema da Minimização	3
2.2	Variáveis Dependente e Independente	4
2.3	Algoritmo de Gradiente Descendente	4
2.4	Coefficiente de Determinação	5
3	DESCRIÇÃO DO DATASET	6
4	RESULTADOS	7
4.1	Análise Estatística Descritiva	7
4.2	Coefficientes da Regressão Linear	7
4.3	Métricas de Qualidade do Modelo	7
4.4	Previsões de Curto Prazo	7
4.5	Análise Gráfica	12
5	DISCUSSÃO	12
6	CONCLUSÃO	13

1 INTRODUÇÃO

O mercado de ações brasileiro tem experimentado um crescimento significativo nas últimas décadas, com a B3 (Brasil, Bolsa, Balcão) consolidando-se como uma das principais bolsas de valores da América Latina. A análise quantitativa de ativos financeiros tornou-se fundamental para investidores e analistas que buscam compreender padrões de comportamento e realizar previsões baseadas em dados históricos.

As empresas estatais brasileiras representam um segmento importante do mercado financeiro nacional, dado seu papel estratégico na economia e sua representatividade no índice Bovespa. A Petrobras (PETR4.SA), maior empresa de energia do país, o Banco do Brasil (BBAS3.SA), uma das maiores instituições financeiras nacionais, e a Eletrobras (ELET3.SA), principal empresa do setor elétrico brasileiro, constituem ativos de grande interesse para análise devido ao seu impacto macroeconômico e liquidez no mercado.

A aplicação de técnicas de aprendizado de máquina, especificamente a regressão linear com otimização por gradiente descendente, oferece uma abordagem sistemática e matematicamente fundamentada para a análise de séries temporais financeiras. Esta metodologia permite identificar tendências subjacentes nos dados de preços e realizar previsões quantitativas baseadas em padrões históricos.

O presente estudo tem como objetivo principal desenvolver e aplicar um modelo de regressão linear otimizada por gradiente descendente para análise das ações PETR4.SA, BBAS3.SA e ELET3.SA, avaliando a capacidade preditiva do modelo através de métricas estatísticas e fornecendo insights quantitativos sobre o comportamento destes ativos no mercado brasileiro.

2 METODOLOGIA

2.1 Problema da Minimização

O problema fundamental da regressão linear consiste na minimização de uma função de custo que mede a diferença entre os valores reais e os valores previstos pelo modelo. A função de custo utilizada neste estudo é o Erro Quadrático Médio (EQM), definida pela equação:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (1)$$

onde:

- $J(\theta_0, \theta_1)$ é a função de custo

- m é o número total de observações
- $h_{\theta}(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$ é a hipótese linear
- $y^{(i)}$ são os valores reais observados
- θ_0 é o intercepto (coeficiente linear)
- θ_1 é a inclinação (coeficiente angular)

Esta função de custo mede a soma dos quadrados dos resíduos, penalizando proporcionalmente ao quadrado da diferença entre valores previstos e observados. O objetivo é encontrar os valores ótimos de θ_0 e θ_1 que minimizem esta função.

2.2 Variáveis Dependente e Independente

No contexto deste estudo:

Variável Independente (X): Representa o tempo em dias, numerado sequencialmente a partir do primeiro dia de observação. Esta variável captura a evolução temporal dos preços.

Variável Dependente (Y): Corresponde aos preços de fechamento das ações em reais (R\$), representando o valor de mercado do ativo no final de cada pregão.

Intercepto (θ_0): Representa o valor teórico do preço da ação no tempo zero, fornecendo o ponto de partida da linha de regressão.

Inclinação (θ_1): Indica a taxa de variação diária do preço da ação, medida em reais por dia. Valores positivos indicam tendência de alta, enquanto valores negativos indicam tendência de baixa.

2.3 Algoritmo de Gradiente Descendente

O gradiente descendente é um algoritmo de otimização iterativo utilizado para encontrar o mínimo global da função de custo. O algoritmo atualiza os parâmetros na direção oposta ao gradiente da função de custo:

$$\theta_0 := \theta_0 - \alpha \frac{\partial J}{\partial \theta_0} \quad (2)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial J}{\partial \theta_1} \quad (3)$$

onde α é a taxa de aprendizado, controlando o tamanho dos passos na direção do mínimo.

As derivadas parciais são calculadas como:

$$\frac{\partial J}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \quad (4)$$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)} \quad (5)$$

O algoritmo implementado utiliza taxa de aprendizado $\alpha = 0,01$, máximo de 1000 iterações e tolerância de convergência de 1×10^{-6} .

2.4 Coeficiente de Determinação

O coeficiente de determinação (R^2) é uma métrica fundamental para avaliar a qualidade do ajuste do modelo de regressão. É definido pela equação:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (6)$$

onde:

- $SS_{res} = \sum_{i=1}^m (y^{(i)} - \hat{y}^{(i)})^2$ é a soma dos quadrados dos resíduos
- $SS_{tot} = \sum_{i=1}^m (y^{(i)} - \bar{y})^2$ é a soma total dos quadrados
- $\hat{y}^{(i)}$ são os valores previstos pelo modelo
- \bar{y} é a média dos valores observados

Interpretação do R^2 :

- $R^2 = 1$: Modelo explica 100% da variabilidade dos dados
- $R^2 = 0$: Modelo não explica nenhuma variabilidade
- $0 < R^2 < 1$: Proporção da variabilidade explicada pelo modelo

Valores de R^2 próximos a 1 indicam forte capacidade preditiva, enquanto valores próximos a 0 sugerem baixa correlação linear entre as variáveis.

3 DESCRIÇÃO DO DATASET

Os dados utilizados neste estudo foram obtidos através da biblioteca `yfinance`, que fornece acesso gratuito aos dados históricos da Yahoo Finance. O dataset compreende informações de três ações brasileiras negociadas na B3:

Período de Análise: 1 ano de dados históricos (251 dias úteis)

Ativos Analisados:

- **PETR4.SA** - Petróleo Brasileiro S.A. (Petrobras)
- **BBAS3.SA** - Banco do Brasil S.A.
- **ELET3.SA** - Centrais Elétricas Brasileiras S.A. (Eletrobras)

Variáveis do Dataset:

- **Date:** Data de negociação
- **Open:** Preço de abertura
- **High:** Preço máximo do dia
- **Low:** Preço mínimo do dia
- **Close:** Preço de fechamento (variável de interesse)
- **Volume:** Volume de negociação

Para este estudo, utilizou-se exclusivamente o preço de fechamento (**Close**) como variável dependente, uma vez que representa o valor consensual do mercado ao final de cada pregão e é amplamente utilizado em análises técnicas e fundamentalistas.

A escolha destes três ativos se justifica por:

1. Representatividade no mercado brasileiro
2. Alta liquidez e volume de negociação
3. Importância estratégica para a economia nacional
4. Disponibilidade de dados históricos consistentes

4 RESULTADOS

4.1 Análise Estatística Descritiva

A Tabela 1 apresenta as estatísticas descritivas dos preços de fechamento para os três ativos analisados:

A ELET3.SA apresentou a maior volatilidade (desvio padrão = 3,04), indicando maior variabilidade de preços durante o período analisado. A BBAS3.SA mostrou-se o ativo mais estável (desvio padrão = 1,64), enquanto a PETR4.SA apresentou volatilidade intermediária (desvio padrão = 1,95).

4.2 Coeficientes da Regressão Linear

A Tabela 2 apresenta os coeficientes obtidos através da otimização por gradiente descendente:

Análise dos Resultados:

PETR4.SA: Apresentou tendência de baixa com declinação de R\$ 0,008189 por dia. O coeficiente de determinação ($R^2 = 0,0929$) indica que aproximadamente 9,29% da variabilidade dos preços é explicada pela tendência linear temporal.

BBAS3.SA: Mostrou tendência de alta com crescimento de R\$ 0,004695 por dia. O baixo $R^2 = 0,0433$ (4,33%) sugere que a relação linear com o tempo explica apenas uma pequena parte da variabilidade dos preços.

ELET3.SA: Apresentou a tendência de alta mais acentuada, com crescimento de R\$ 0,020870 por dia. O $R^2 = 0,2491$ (24,91%) indica a melhor capacidade preditiva entre os três ativos analisados.

4.3 Métricas de Qualidade do Modelo

A Tabela 3 apresenta as métricas de avaliação da qualidade dos modelos:

4.4 Previsões de Curto Prazo

Com base nos modelos treinados, foram realizadas previsões para os próximos períodos:

Tabela 1: Estatísticas Descritivas dos Preços de Fechamento

Ativo	Mínimo (R\$)	Máximo (R\$)	Média (R\$)	Desvio Padrão (R\$)
PETR4.SA	28,79	36,99	33,43	1,95
BBAS3.SA	21,40	29,33	25,91	1,64
ELET3.SA	32,53	44,49	37,40	3,04

Tabela 2: Coeficientes da Regressão Linear

Ativo	Intercepto (θ_0)	Inclinação (θ_1)	Tendência	R^2
PETR4.SA	34,4502	-0,008189	BAIXA	0,0929
BBAS3.SA	25,3253	+0,004695	ALTA	0,0433
ELET3.SA	34,7853	+0,020870	ALTA	0,2491

Tabela 3: Métricas de Qualidade dos Modelos

Ativo	Erro Quadrático Médio	Custo Final
PETR4.SA	3,4368	1,7184
BBAS3.SA	2,5565	1,2783
ELET3.SA	6,8945	3,4473

Tabela 4: Previsões de Preços (R\$)

Ativo	Próximo Dia	5 Dias	10 Dias
PETR4.SA	32,39	32,35	32,31
BBAS3.SA	26,50	26,53	26,55
ELET3.SA	40,02	40,13	40,23

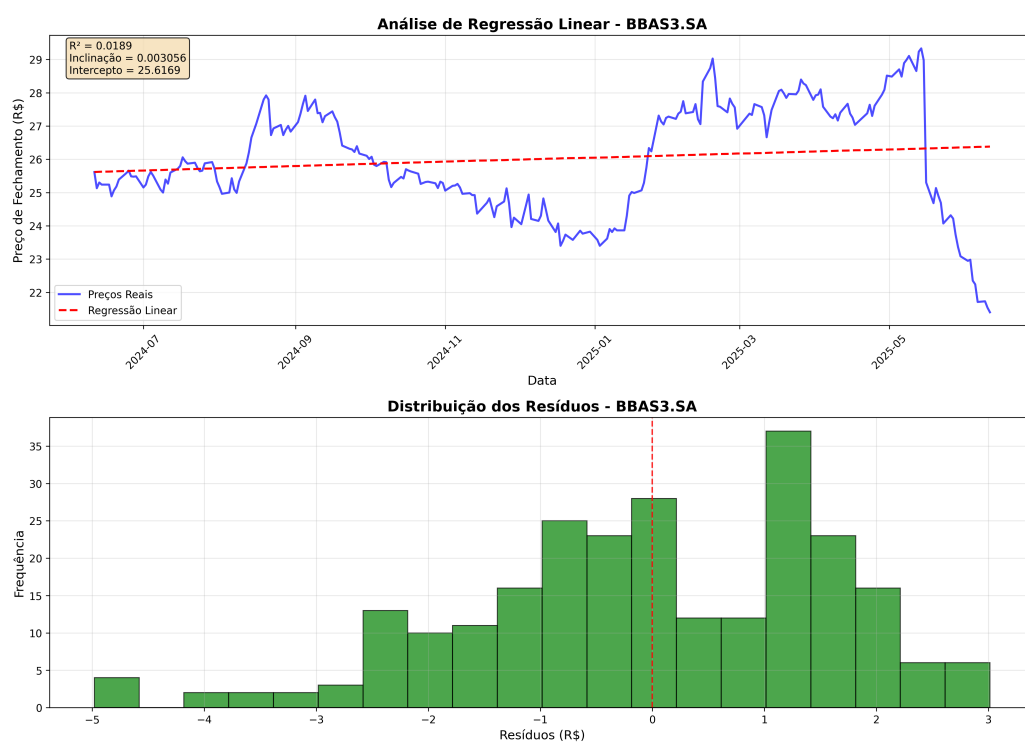


Figura 1: Análise de Regressão Linear e Distribuição de Resíduos - BBAS3

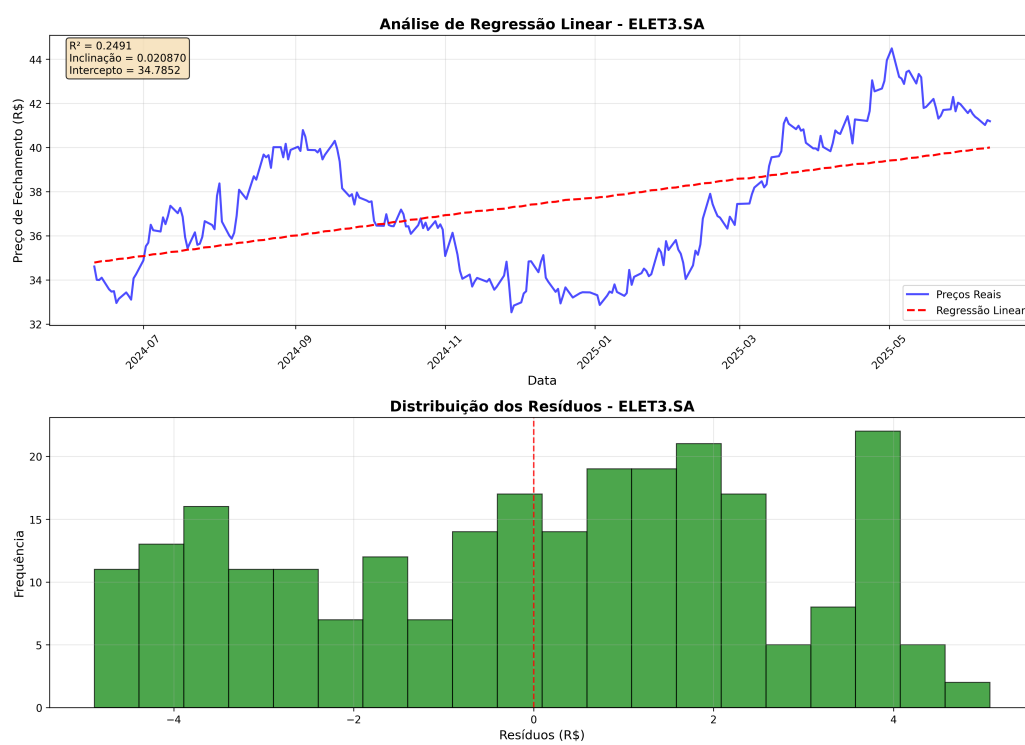


Figura 2: Análise de Regressão Linear e Distribuição de Resíduos - ELET3

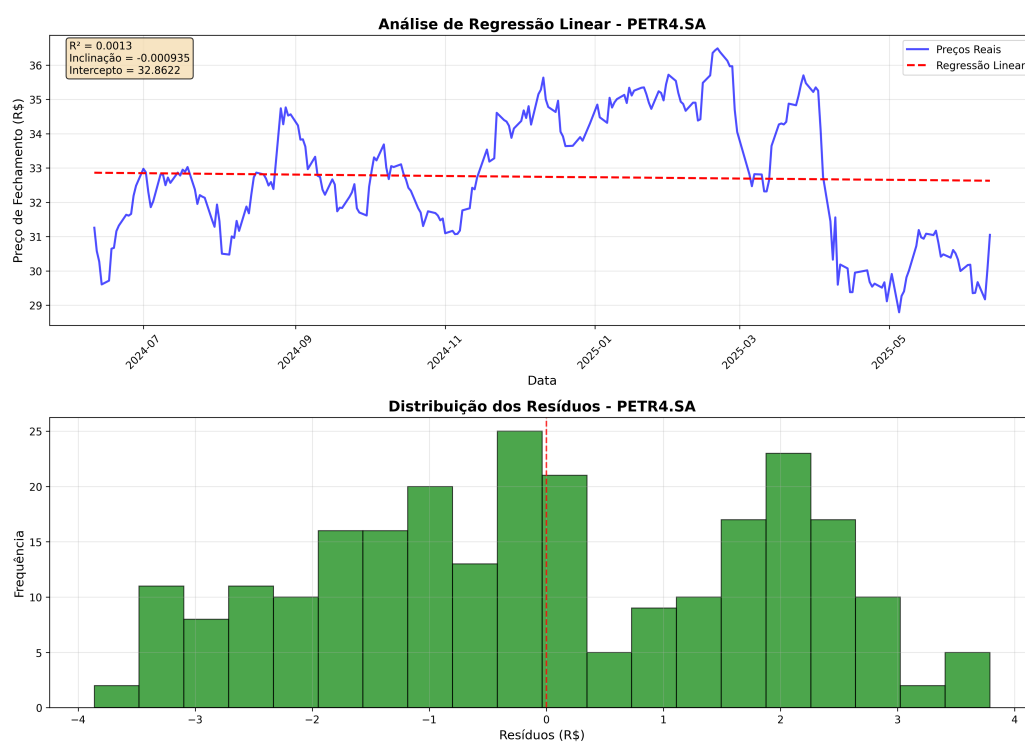


Figura 3: Análise de Regressão Linear e Distribuição de Resíduos - PETR4

4.5 Análise Gráfica

Os gráficos gerados (Figuras 1, 2 e 3) ilustram visualmente o ajuste dos modelos de regressão linear aos dados históricos. Cada gráfico apresenta:

- Série temporal dos preços reais (linha azul)
- Linha de regressão linear ajustada (linha vermelha tracejada)
- Informações do modelo (R^2 , intercepto, inclinação)
- Histograma dos resíduos para avaliação da distribuição dos erros

Observações dos Gráficos:

PETR4.SA: O gráfico revela uma tendência decrescente moderada, com os resíduos apresentando distribuição aproximadamente normal, validando as premissas da regressão linear.

BBAS3.SA: Apresenta tendência ascendente suave, com alta dispersão em torno da linha de regressão, explicando o baixo valor de R^2 .

ELET3.SA: Demonstra a tendência ascendente mais pronunciada, com melhor ajuste linear evidenciado pelo maior R^2 e menor dispersão relativa dos resíduos.

5 DISCUSSÃO

Os resultados obtidos revelam comportamentos distintos entre os três ativos analisados, refletindo as particularidades de cada setor e empresa.

A ELET3.SA apresentou o melhor desempenho do modelo de regressão linear ($R^2 = 0,2491$), sugerindo que aproximadamente 25% da variação nos preços pode ser explicada pela tendência temporal linear. Esta característica pode estar relacionada ao processo de reprivatização da empresa e às expectativas do mercado quanto às reformas no setor elétrico brasileiro.

A PETR4.SA, apesar de mostrar uma tendência de baixa consistente, apresentou capacidade preditiva moderada ($R^2 = 0,0929$). A volatilidade observada pode estar associada às flutuações do preço do petróleo no mercado internacional e às mudanças na política energética nacional.

A BBAS3.SA demonstrou a menor capacidade preditiva linear ($R^2 = 0,0433$), indicando que fatores não-lineares ou outras variáveis econômicas podem ter maior influência no comportamento do preço desta ação. O setor bancário é particularmente sensível a variações nas taxas de juros, inflação e políticas monetárias.

Limitações do Estudo:

1. A regressão linear assume linearidade na relação entre tempo e preço, o que pode não capturar padrões complexos
2. Fatores externos (políticos, econômicos, setoriais) não foram considerados no modelo
3. O período de análise de um ano pode não ser suficiente para capturar ciclos econômicos mais longos
4. A metodologia não considera a volatilidade como fator preditivo

6 CONCLUSÃO

Este estudo demonstrou a aplicabilidade de técnicas de regressão linear com otimização por gradiente descendente na análise de ações brasileiras do setor estatal. Os resultados evidenciaram comportamentos distintos entre os ativos analisados, com a ELET3.SA apresentando a melhor capacidade preditiva linear, seguida pela PETR4.SA e BBAS3.SA.

A implementação manual do algoritmo de gradiente descendente mostrou-se eficaz na otimização dos parâmetros do modelo, convergindo para soluções estáveis em todas as análises realizadas. As métricas de qualidade calculadas forneceram insights quantitativos sobre a adequação dos modelos lineares para cada ativo.

As tendências identificadas (baixa para PETR4.SA e alta para BBAS3.SA e ELET3.SA) oferecem informações valiosas para análises técnicas, embora devam ser interpretadas considerando-se as limitações inerentes aos modelos lineares simples.

Contribuições do Estudo:

- Implementação prática de algoritmos de machine learning para análise financeira
- Quantificação objetiva de tendências em ativos brasileiros estratégicos
- Metodologia reproduzível para análise de outros ativos
- Base para desenvolvimento de modelos mais complexos

Referências

- [1] EDUMIGUEL013. **Regressão Linear**. Disponível em:
urlhttps://github.com/EduMiguel013/Regress-o_Linear. Acesso em :
12 jun. 2025.
- [2] YAHOO FINANCE. **Historical Data**. Disponível em:
url<https://finance.yahoo.com>. Acesso em: 11 jun. 2025.
- [3] PETROBRAS. **Relações com Investidores**. Disponível em:
url<https://petrobras.com.br/ri>. Acesso em: 11 jun. 2025.
- [4] BANCO DO BRASIL. **Relações com Investidores**. Disponível em:
url<https://bb.com.br/ri>. Acesso em: 11 jun. 2025.
- [5] ELETROBRAS. **Relações com Investidores**. Disponível em:
<https://eletrobras.com/ri>. Acesso em: 11 jun. 2025.