

PREVISÃO DE RETORNOS DE ATIVOS COM ALGORITMOS DE DEEP LEARNING

Cristiane Coutinho Inteli AUTORA cristiane.coutinho@sou.inteli.edu.br	Adriana Vieira Universidade Beira do Interior COORIENTADOR adriana.coelhovieira@gmail.com	Geraldo Vasconcelos Inteli ORIENTADOR geraldo.vasconcelos@prof.inteli.edu.br
--	--	---

RESUMO

Este estudo teve como objetivo analisar o desempenho de diferentes arquiteturas de deep learning aplicadas à previsão de retornos de ativos financeiros utilizando séries temporais históricas da bolsa de valores brasileira (B3). Foram testados quatro modelos como LSTM, CNN, Transformer e MLP, com o intuito de avaliar a capacidade de generalização e acurácia preditiva em um contexto caracterizado por alta volatilidade, ruído e não estacionaridade. Os dados utilizados cobriram o período de 2021 a 2024 e incluíram informações de preço de abertura, fechamento, máximos, mínimos e volume de negociação. A variável alvo, retorno, foi calculada como a diferença entre o preço de fechamento e o de abertura, dividida pelo preço de abertura. As bases foram normalizadas e segmentadas em conjuntos de treino (janeiro a agosto) e teste (setembro a dezembro).

Os resultados demonstraram a complexidade do problema de previsão de séries financeiras. O modelo CNN apresentou erros baixos em magnitude ($RMSE = 0,0175$ e $MAE = 0,0079$), mas um R^2 negativo e MAPE elevado, evidenciando falhas de generalização. O Transformer obteve erros absolutos pequenos ($RMSE = 0,0086$ e $MAE = 0,0048$), mas produziu previsões quase constantes, sugerindo incapacidade de aprendizado significativo. A MLP apresentou R^2 elevado (0,8688), porém acompanhada de um MAPE extremamente alto, indicando overfitting. Por fim, a LSTM obteve desempenho insatisfatório ($R^2 = -1145,32$), com erros altos e incapacidade de capturar dependências temporais.

Conclui-se que, embora o deep learning ofereça potencial para análise de dados financeiros, nenhum modelo apresentou desempenho plenamente satisfatório em termos de generalização. A natureza caótica e imprevisível do mercado financeiro impõe desafios significativos à modelagem preditiva, exigindo estratégias complementares, como modelos híbridos, engenharia de atributos aprimorada e inclusão de variáveis exógenas (indicadores macroeconômicos e de sentimento). Assim, este trabalho reforça a importância da experimentação sistemática e da compreensão dos limites das arquiteturas neurais quando aplicadas à previsão de séries temporais financeiras.

Palavras-chave: Dados, Redes Neurais, LSTM, Análise preditiva.

1. INTRODUÇÃO

Ao emprestar recursos a instituições, cria-se um ambiente de negociação com objetivo de recebimento de retorno financeiro baseado em investimentos. Dessa forma, os prêmios que são retornados dessas negociações são denominados “valores mobiliários”. Quando se

adquire um título de dívida, o investidor também assume o risco sobre esse retorno, ou seja, a possibilidade desse investimento ser desvalorizado.

Com o objetivo de maximização de resultados, criou-se toda uma estrutura tecnológica para apoiar e estabelecer a relação de tempo e causalidade nos ganhos de ativos. Para Guedes (2022), o desenvolvimento de algoritmos do setor financeiro foi inserido aos poucos no mercado de negociações e evoluiu à medida que análises mais técnicas de matemática, probabilidade e estatística foram sendo incorporadas nos chamados “robôs investidores”.

Com esse avanço tecnológico, o desenvolvimento de algoritmos com Machine Learning e Deep Learning, tornou-se cada vez mais viável e atrativo, utilizados para seleção de ativos e otimização de carteiras de investimentos. Shimabukuro (2024), atribuiu esses avanços ao crescente aumento na velocidade de processamento computacional e também a diminuição do tamanho físico de memória. Por causa disso se tornou possível, a implementação mais barata de algoritmos de otimização e predição.

Entre técnicas de deep learning, redes neurais recorrentes do tipo LSTM têm se mostrado particularmente eficientes na modelagem de séries temporais financeiras, pois incorporam informações passadas e mecanismos de atualização de memória, permitindo relacionar dados atuais a estados remotos no tempo e evitando problemas de perda ou explosão de gradiente.

Além das LSTMs, arquiteturas baseadas em transformadores de visão, como ViT, DeiT, Swin e ConvMixer, também têm sido aplicadas à previsão de ativos financeiros. Nessas abordagens, séries temporais são convertidas em imagens 2D e enriquecidas com indicadores técnicos, permitindo classificar movimentos do mercado em Comprar, Vender ou Manter.

2. REVISÃO DA LITERATURA

O trabalho de Mesquita 2020, aplica redes neurais LSTM para previsão de séries financeiras da Bovespa, utilizando dez séries temporais como base de análise. O objetivo principal foi verificar a relação entre os resultados de previsão e a validade da Hipótese de Passeio Aleatório (HPA), testada por meio da razão de variância. A escolha da LSTM justifica-se pela sua capacidade de capturar dependências de longo prazo em dados sequenciais, superando limitações das redes recorrentes tradicionais, como explosão e desaparecimento do gradiente. O estudo demonstra que séries que rejeitam a HPA apresentam melhor desempenho preditivo, reforçando a ideia de que a escolha de dados estatisticamente adequados influencia a eficácia do modelo. Além disso, os resultados foram comparados com benchmarks como Buy and Hold e classificadores aleatórios, mostrando que a metodologia proposta pode oferecer vantagens operacionais em estratégias de negociação baseadas em aprendizado profundo.

Gezici 2024 explora o uso de arquiteturas baseadas em transformadores para previsão de preços de ativos e movimentos direcionais. O trabalho avalia diferentes variantes de transformadores de visão (ViT, DeiT e Swin) e uma abordagem baseada em convoluções com incorporação de patches (ConvMixer), além de compará-las com métodos tradicionais como a CNN. Para aplicar essas arquiteturas, os autores transformaram séries temporais financeiras unidimensionais em imagens bidimensionais, enriquecidas por mais de 50 indicadores técnicos (como RSI, EMA, MACD, entre outros), usados como insumos para os

modelos. Os resultados mostram que o ConvMixer apresentou a melhor precisão entre os métodos testados, seguido pelo ViT, enquanto o Swin superou o DeiT. Embora CNNs tradicionais ainda apresentem bom desempenho, os transformadores mostraram maior potencial, indicando uma tendência promissora para o uso dessas arquiteturas no campo financeiro.

Já Shimabukuro 2024, utilizou a previsão de séries temporais financeiras tanto com modelos econométricos clássicos quanto redes neurais recorrentes LSTM. Os modelos econométricos analisados consideram relações lineares entre os valores e seus termos de erro, como MA, AR, ARMA, ARIMA e SARIMA, capturando padrões sazonais e dependências temporais. Em contrapartida, as redes LSTM não pressupõem linearidade, permitindo aprender relações complexas entre dados atuais e passados, além de incorporarem mecanismos que evitam a perda ou explosão do gradiente em sequências longas. O treinamento da rede ocorreu de forma iterativa, utilizando log-retornos e diferenciação fracionária, com ajuste de pesos via retropropagação e gradiente descendente, processamento em lotes e early stopping para prevenir o overfitting. Dessa forma, o estudo evidencia a capacidade das redes LSTM de aprender relações não-lineares em séries financeiras, oferecendo uma alternativa preditiva superior aos métodos tradicionais.

TÍTULO	ARQUITETURA	MÉTRICAS
Utilização de uma rede neural LSTM e teste da razão de variância para previsões em séries de ativos da Bovespa	LSTM	Comparação com Buy and Hold, Classificador Aleatório, Razão de Variância
Deep transformer-based asset price and direction prediction	ViT, DeiT, Swin, ConvMixer, CNN	Acurácia de teste, comparação entre arquiteturas, distribuição de rótulos (Comprar, Manter, Vender)
Deep Learning aplicado à predição de retornos do Ibovespa: uma análise do desempenho da rede neural LSTM utilizando log-retornos e diferenciação fracionária	LSTM com Log-retornos e Diferenciação Fracionária	RMSE, MAE, MAPE, MASE, R^2 , comparação com baseline

3. JUSTIFICATIVA

Soares (2023), destaca que algoritmos de machine learning são amplamente explorados em problemas de regressão e classificação para predição de precificação de retornos e classificação de um determinado ativo. Em contrapartida, Shimabukuro (2024), aponta fatores limitantes relacionados à aplicação de modelos de machine learning para predição de ativos financeiros, pela grande necessidade de volume de dados, baixa sensibilidade a ruído de mercado e risco de overfitting.

Diante desses desafios, investigar estratégias de algoritmos avançados de Deep Learning, como LSTMs e transformadores, torna-se relevante, pois permite explorar métodos preditivos mais robustos, capazes de capturar relações complexas e não-lineares em séries temporais financeiras, contribuindo para decisões de investimento mais informadas e precisas.

4. OBJETIVOS

4.1. OBJETIVOS GERAIS

Análise preditiva de retorno financeiro a partir de diferentes arquiteturas de redes neurais usando técnicas de deep learning sobre uma série temporal.

4.2. OBJETIVOS ESPECÍFICOS

- Realizar análise exploratória das séries temporais financeiras, Identificando padrões, tendências, sazonalidades, outliers e correlações entre variáveis relevantes.
- Implementar diferentes arquiteturas de redes neurais MLP, LSTM, CNN1D e Transformers para a previsão de retornos de ativos financeiros.
- Comparar o desempenho preditivo das arquiteturas utilizando métricas quantitativas (RMSE, MAE, MAPE, R^2).
- Mensurar a capacidade de generalização dos modelos testando em ativos financeiros distintos e em diferentes janelas temporais.
- Propor recomendações de uso das arquiteturas mais eficazes Discutindo suas vantagens e limitações para aplicações reais no mercado financeiro.
- Validando o modelo utilizando precisão, revocação, medida F1 e acurácia.

5. HIPÓTESES

- H1: Arquiteturas de deep learning baseadas em redes recorrentes (LSTM) apresentam maior acurácia preditiva em séries temporais financeiras do que arquiteturas feedforward tradicionais.

- H2: Modelos baseados em mecanismos de atenção ou transformadores (Transformers) são capazes de capturar melhor os ruídos de mercado e generalizar os resultados para diferentes ativos financeiros.
- H3: Modelos de deep learning apresentam menor overfitting em comparação a modelos tradicionais de machine learning (como regressão linear ou random forest) quando aplicados a grandes volumes de dados financeiros.
- H4: Redes MLP podem ser menos eficientes em capturar dependências temporais de séries financeiras.
- H5: O aumento da complexidade do modelo (número de camadas e unidades) só melhora o desempenho preditivo até um ponto, após o qual o risco de overfitting se torna significativo, mesmo com técnicas como early stopping.

6. PROBLEMA DE PESQUISA

Quais arquiteturas de deep learning demonstram maior capacidade de generalização na previsão de retornos financeiros e quais se destacam em termos de métricas de avaliação?

7. METODOLOGIA

7.1. DESCRIÇÃO DOS DADOS UTILIZADOS

Para este estudo foram coletados dados de séries temporais reais públicos da bolsa de valores do Brasil B3 (Brasil, Bolsa, Balcão). Para a coleta das séries históricas de ativos da B3, foram utilizados os arquivos COTAHIST, disponibilizados oficialmente pela bolsa de valores do Brasil (B3). Esses arquivos estão organizados anualmente no formato .ZIP, contendo arquivos de texto (.txt) estruturados segundo o layout definido pela própria B3. Foram extraídos dados dos últimos 4 anos (2021, 2022, 2023), os Datasets contém variáveis sobre data da cotação, ticker, preço máximo, preço mínimo, preço de fechamento. As informações foram organizadas em planilhas csv para cada ano e contém as seguintes variáveis:

NOME DA VARIÁVEL	TIPO DA VARIÁVEL	DESCRIÇÃO DA VARIÁVEL
data	Data	Data de negociação do ativo
ticker	Nominal	Nome de identificação do ativo
preco_abertura	Numérica	Valor da primeira negociação do ativo
preco_max	Numérica	Maior preço pelo qual o ativo foi negociado ao longo do dia.

preco_min	Numérica	Menor preço pelo qual o ativo foi negociado ao longo do dia.
preco_fechamento	Numérica	Menor preço pelo qual o ativo foi negociado ao longo do dia.
volume	Numérica	Número total de negociações multiplicado pelos respectivos preços.

Além disso, foram observadas também o número de linhas contidas em cada Dataset e a quantidade de tickers únicos.

DATASET	NÚMERO DE LINHAS	TICKERS ÚNICOS
Dataset 2021	1831862	101995
Dataset 2022	2117440	121201
Dataset 2023	2257424	133292
Dataset 2024	2635561	167131

7.2. ARQUITETURAS UTILIZADAS

A escolha das arquiteturas LSTM, CNN 1D, Transformers e MLP fundamenta-se nas características próprias de cada modelo e em sua aplicabilidade à previsão de retornos de ativos.

- **LSTM:** São redes recorrentes capazes de processar dados sequenciais e superar problemas como explosão ou desaparecimento de gradiente, utilizando portas que controlam o fluxo de informação no tempo. Serão utilizadas por sua capacidade de capturar dependências de longo prazo.
- **CNN 1D:** Embora originalmente aplicadas em visão computacional, CNNs têm se mostrado eficazes na predição de preços financeiros ao identificar padrões locais em séries temporais. Serão usadas para extrair características de curto prazo e padrões repetitivos presentes nos retornos dos ativos.
- **Transformers:** Baseiam-se em mecanismos de atenção e autoatenção, permitindo relacionar eventos distantes em uma série de forma eficiente. Serão empregados pela

flexibilidade em capturar relações de longo alcance entre variáveis e pela tendência de superarem redes recorrentes e convolucionais em tarefas complexas.

- **MLP:** Redes neurais multicamadas que processam entradas de forma densa e não sequencial. Serão utilizadas como baseline, possibilitando comparar o ganho de desempenho dos modelos especializados (LSTM, CNN e Transformers) frente a uma arquitetura mais simples.

7.3. LIMPEZA DE DADOS E ANÁLISE

- VERIFICAÇÃO DE DADOS NULOS OU VAZIOS E ZERADOS

Inicialmente, foi realizada uma análise para identificar linhas que apresentassem ao menos uma coluna com valor nulo, vazio ou igual a zero. As linhas que atendiam a esses critérios foram removidas do dataset, garantindo maior consistência e confiabilidade dos dados utilizados nos modelos.

- CRIAÇÃO DA COLUNA DE RETORNO

Considerando que o objetivo do estudo é prever o retorno financeiro, foi criada uma coluna específica chamada retorno. Essa coluna corresponde ao resultado da diferença entre o preço de fechamento e o preço de abertura, dividido pelo preço de abertura, representando a variação percentual diária do ativo.

- ELIMINAR DUPLICATAS

Foi verificado que o dataset não possui registros duplicados, garantindo que cada linha corresponda a um evento único de negociação.

- TRATAMENTO DE OUTLIERS

Optou-se por não tratar os outliers presentes no dataset, considerando que a distribuição das variáveis de interesse apresenta uma assimetria à direita. A remoção ou alteração desses valores extremos poderia distorcer a variabilidade natural dos dados e impactar negativamente a capacidade preditiva dos modelos, uma vez que os outliers refletem movimentos significativos nos retornos financeiros que o modelo precisa aprender a capturar. Dessa forma, todos os valores foram mantidos, preservando a integridade da distribuição original e permitindo que as arquiteturas de redes neurais pudessem aprender padrões reais, incluindo eventos extremos, que são relevantes para previsões mais precisas em séries temporais financeiras.

- NORMALIZAÇÃO / PADRONIZAÇÃO

A etapa de normalização e padronização das variáveis numéricas, foi aplicado a normalização minMax em todas as features numéricas do dataset, com o objetivo de deixar os valores entre 1 e 0. Essa normalização foi feita para garantir que o valor alto do volume comparado a outras variáveis, não gerasse de alguma forma, overfitting.

- ENCODING DE VARIÁVEIS CATEGÓRICAS

As variáveis categóricas não serão utilizadas no modelo.

- ELIMINAÇÃO DE EXEMPLOS COM MENOS DE 250 EXEMPLOS

Para o desenvolvimento do modelo de rede neural, é necessário que o input relacionado ao ticker, tenha ao menos 250 exemplos, seria o número de dias úteis em que os ativos são negociados na bolsa. Dessa forma, para uma análise temporal, ter os dados de forma contínua contribui para um melhor resultado.

- **QUANTIDADE DE EXEMPLOS DE CADA VARIÁVEL**

Para o estudo de análise exploratória, os datasets foram divididos em anos, sendo eles 2021, 2022, 2023 e 2024. Para cada ano foi verificado a quantidade de exemplos (linhas).

ANO	QUANTIDADE DE EXEMPLOS
2021	1628332
2022	1887323
2023	1990322
2024	2265471

	preco_abertura	preco_max	preco_min	preco_fechamento	volume	retorno
count	1.628332e+06	1.628332e+06	1.628332e+06	1.628332e+06	1.628332e+06	1.628332e+06
mean	5.742379e+01	5.912532e+01	5.582567e+01	5.744157e+01	4.850943e+08	1.299595e-02
std	5.633530e+02	5.727851e+02	5.558302e+02	5.645718e+02	7.160942e+09	3.808237e-01
min	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e+00	-9.956522e-01
25%	3.600000e-01	4.100000e-01	3.200000e-01	3.600000e-01	1.500000e+05	-2.803738e-02
50%	1.600000e+00	1.740000e+00	1.500000e+00	1.600000e+00	1.360900e+06	0.000000e+00
75%	1.265000e+01	1.287000e+01	1.243000e+01	1.264000e+01	1.144000e+07	1.486520e-02
max	1.295160e+05	1.295160e+05	1.295160e+05	1.295160e+05	2.471088e+12	1.940000e+02

Figura 1: Análise Descritiva 2021

	preco_abertura	preco_max	preco_min	preco_fechamento	volume	retorno
count	1.887323e+06	1.887323e+06	1.887323e+06	1.887323e+06	1.887323e+06	1.887323e+06
mean	5.919021e+01	6.130324e+01	5.712483e+01	5.921150e+01	3.743698e+08	1.562686e-02
std	5.541037e+02	5.668977e+02	5.423721e+02	5.551114e+02	6.582166e+09	5.902354e-01
min	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e+00	-9.950495e-01
25%	3.000000e-01	3.500000e-01	2.700000e-01	3.000000e-01	1.078075e+05	-2.846488e-02
50%	1.320000e+00	1.430000e+00	1.230000e+00	1.330000e+00	9.600000e+05	0.000000e+00
75%	7.800000e+00	7.990000e+00	7.620000e+00	7.800000e+00	8.042438e+06	1.758191e-02
max	1.168860e+05	1.168860e+05	1.168860e+05	1.168860e+05	2.487180e+12	6.490000e+02

Figura 2: Análise Descritiva 2022

	preco_abertura	preco_max	preco_min	preco_fechamento	volume	retorno
count	1.990322e+06	1.990322e+06	1.990322e+06	1.990322e+06	1.990322e+06	1.990322e+06
mean	4.338949e+01	4.477701e+01	4.207376e+01	4.341316e+01	2.934385e+08	1.686440e-02
std	5.231317e+02	5.309452e+02	5.163104e+02	5.238585e+02	4.997054e+09	4.856598e-01
min	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e+00	-9.950000e-01
25%	2.500000e-01	2.900000e-01	2.200000e-01	2.500000e-01	6.760000e+04	-2.206736e-02
50%	1.100000e+00	1.180000e+00	1.010000e+00	1.100000e+00	5.599440e+05	0.000000e+00
75%	6.440000e+00	6.600000e+00	6.290000e+00	6.450000e+00	4.990200e+06	1.438849e-02
max	1.280060e+05	1.280060e+05	1.280060e+05	1.280060e+05	1.774876e+12	3.290000e+02

Figura 3: Análise Descritiva 2023

	preco_abertura	preco_max	preco_min	preco_fechamento	volume	retorno
count	1.628332e+06	1.628332e+06	1.628332e+06	1.628332e+06	1.628332e+06	1.628332e+06
mean	5.742379e+01	5.912532e+01	5.582567e+01	5.744157e+01	4.850943e+08	1.299595e-02
std	5.633530e+02	5.727851e+02	5.558302e+02	5.645718e+02	7.160942e+09	3.808237e-01
min	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e-02	1.000000e+00	-9.956522e-01
25%	3.600000e-01	4.100000e-01	3.200000e-01	3.600000e-01	1.500000e+05	-2.803738e-02
50%	1.600000e+00	1.740000e+00	1.500000e+00	1.600000e+00	1.360900e+06	0.000000e+00
75%	1.265000e+01	1.287000e+01	1.243000e+01	1.264000e+01	1.144000e+07	1.486520e-02
max	1.295160e+05	1.295160e+05	1.295160e+05	1.295160e+05	2.471088e+12	1.940000e+02

Figura 4: Análise Descritiva 2024

Distribuição dos preços: Os preços médios de abertura e fechamento variam entre 41 e 57 unidades monetárias, com desvio padrão alto, refletindo a presença de valores extremos

(outliers). Os valores mínimos frequentemente são muito próximos de zero, enquanto os máximos chegam a mais de 100 mil unidades.

Volume de negociação: O volume médio diário varia de aproximadamente $2,46 \times 10^8$ a $4,85 \times 10^8$, mas com valores máximos chegando a $2,48 \times 10^{12}$, indicando dias de negociação excepcionalmente intensos.

Retorno financeiro: O retorno médio diário é pequeno, em torno de 1,3% a 1,6%, enquanto o desvio padrão é relativamente alto devido a oscilações extremas ocasionais. A mediana do retorno é zero em todos os anos, mostrando que a maioria dos dias apresenta variações pequenas.

- DISTRIBUIÇÃO DAS VARIÁVEIS

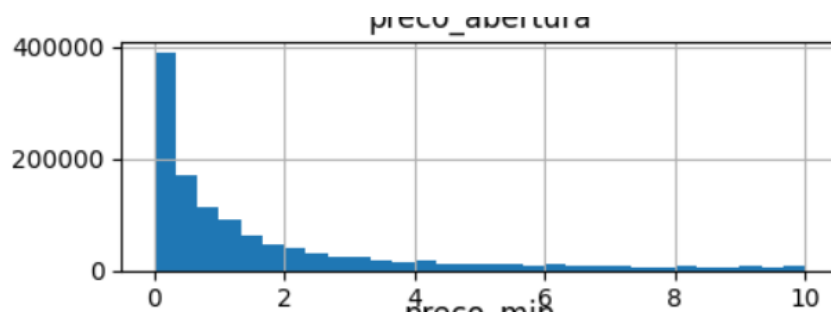


Figura 4 :Distribuição de preço de abertura

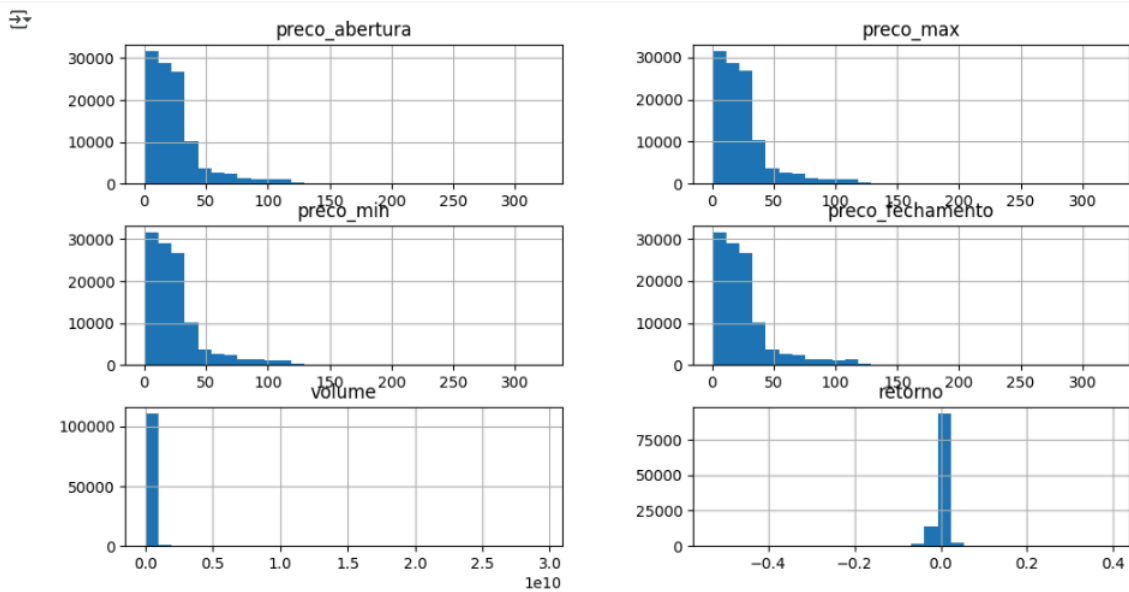
A seguinte distribuição descreve o comportamento de uma variável do dataset, em um intervalo de 0 a 10. O dataset original apresentou um volume muito grande de exemplos concentrados entre 0 e 1. Mesmo com esse concentramento o dataset ainda apresentou “ruído” relacionado a valores muito altos.

- DISTRIBUIÇÃO APÓS SEPARAÇÃO DE DIAS ÚTEIS

Após a remoção de tickers que continham menos de 230 exemplos, foi feita uma nova avaliação da distribuição, que demonstrou estar menos concentrada e com valores mais equilibrados:

- 2021

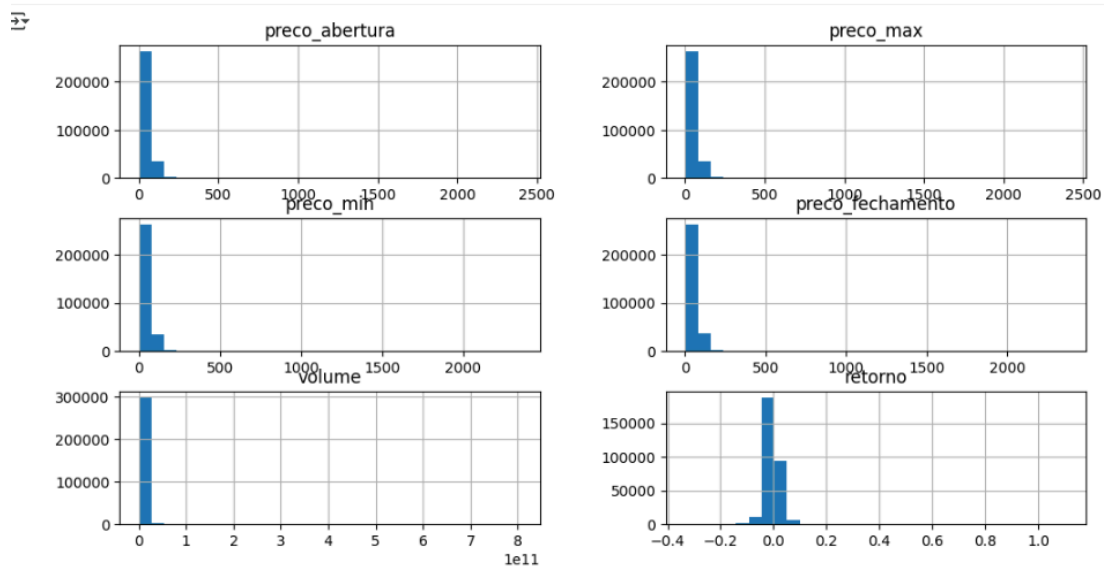
```
df2021.hist(bins=30, figsize=(12,6))
plt.show()
```



- 2022

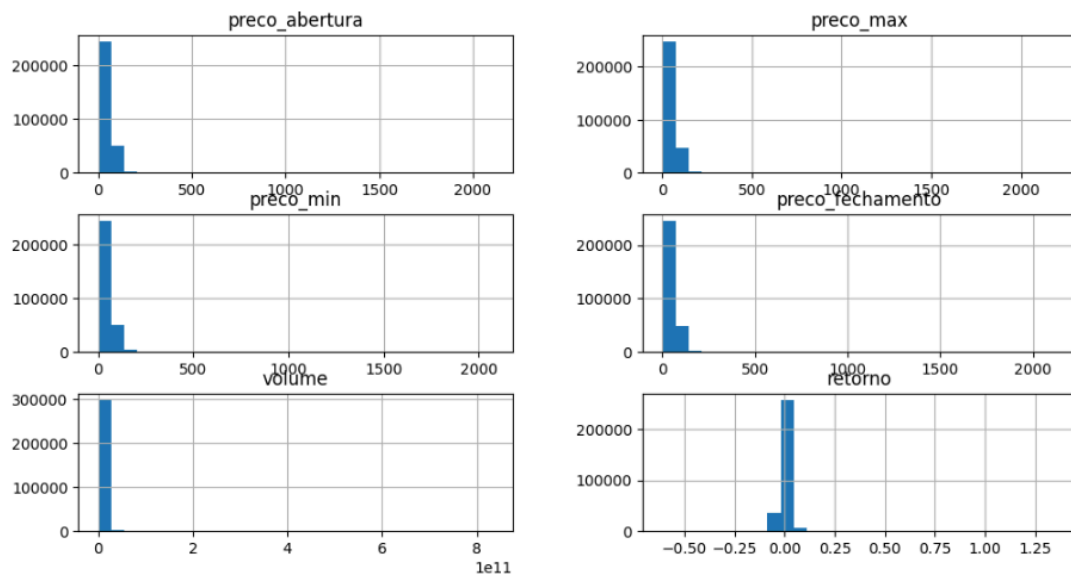
Dados relacionados ao ano de 2022 ainda apresentam ruídos, a respeito de poucos dados com alto valor.

```
df2022.hist(bins=30, figsize=(12,6))
plt.show()
```



- 2023

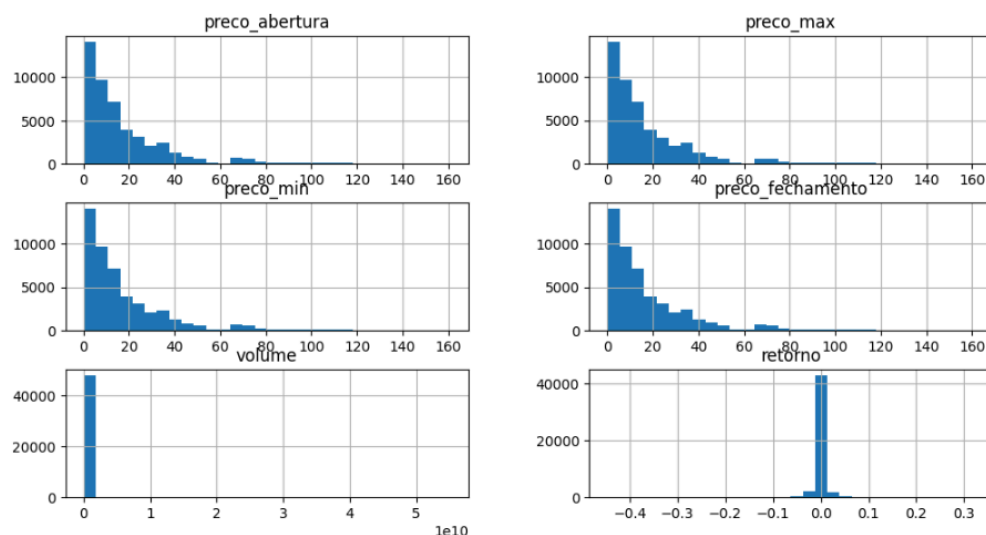
12



- 2024

```
df2024.hist(bins=30, figsize=(12,6))
plt.show()
```

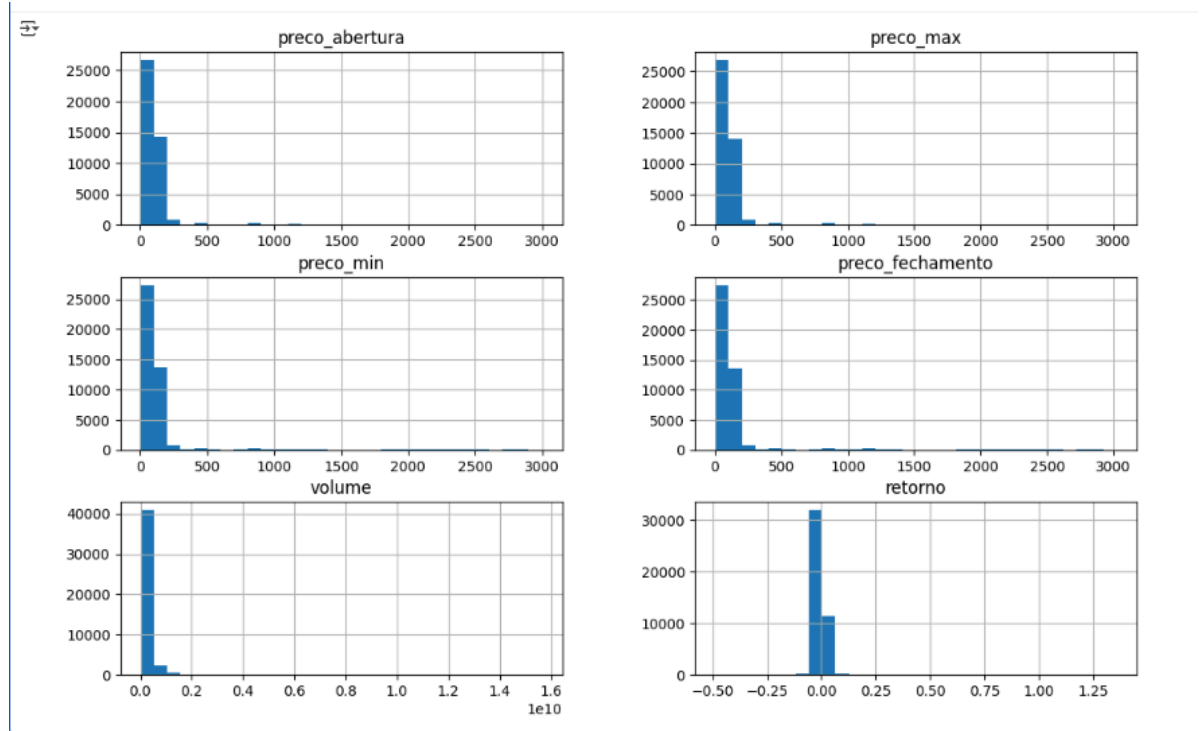
13



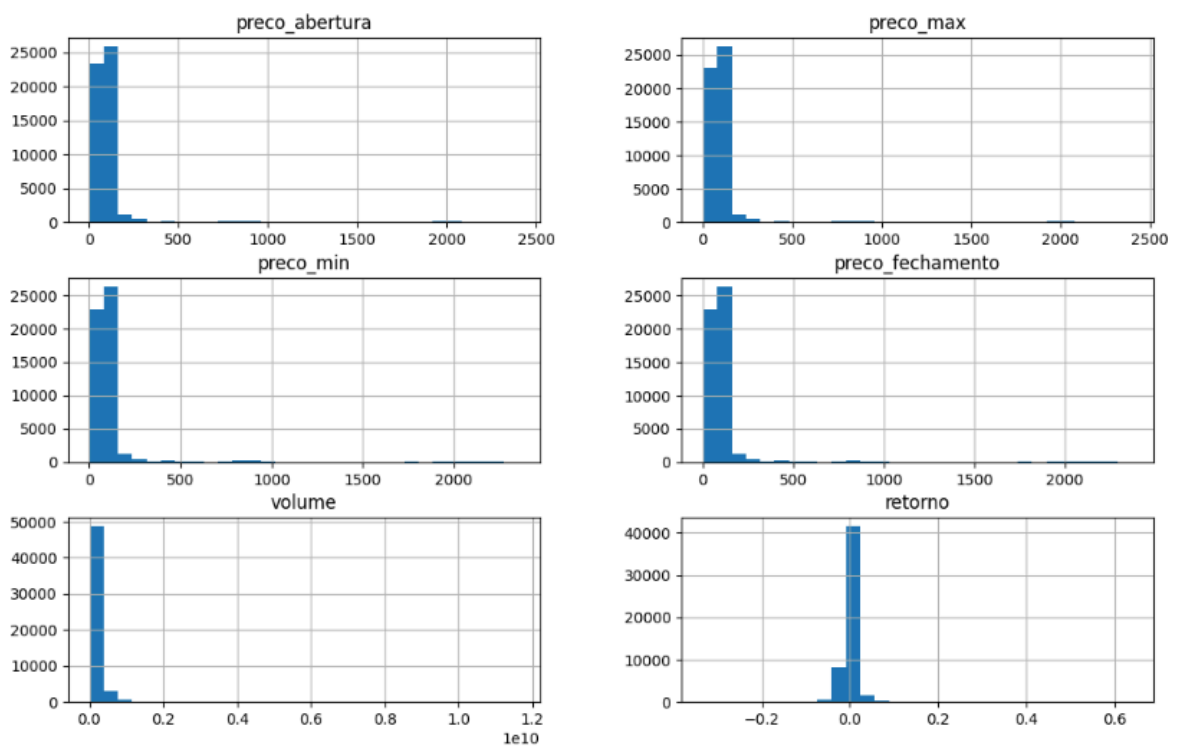
- DISTRIBUIÇÃO APÓS A SEGREGAÇÃO EM TIPO DE MERCADO E CÓDIGO BDI

Para o pré-processamento dos dados nas redes neurais, os dados foram segregados, para serem analisados tickers relacionados a mercado a vista e que fossem fundos de investimento imobiliário. Dessa forma, obtemos uma distribuição mais próxima da normal.

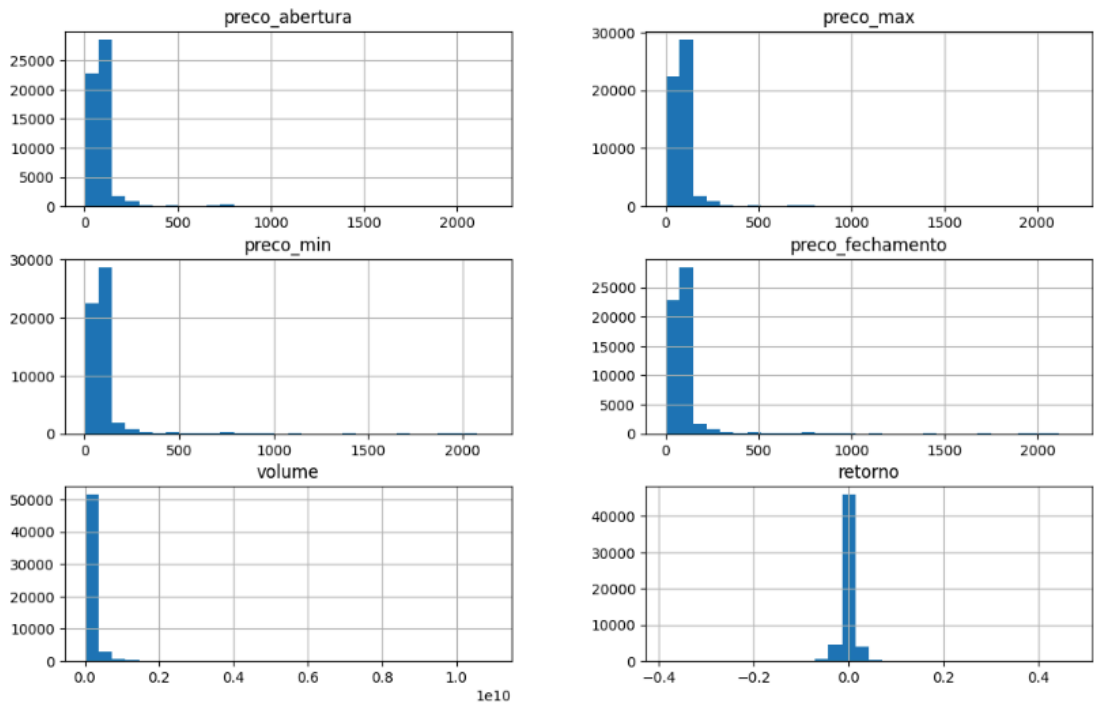
- 2021



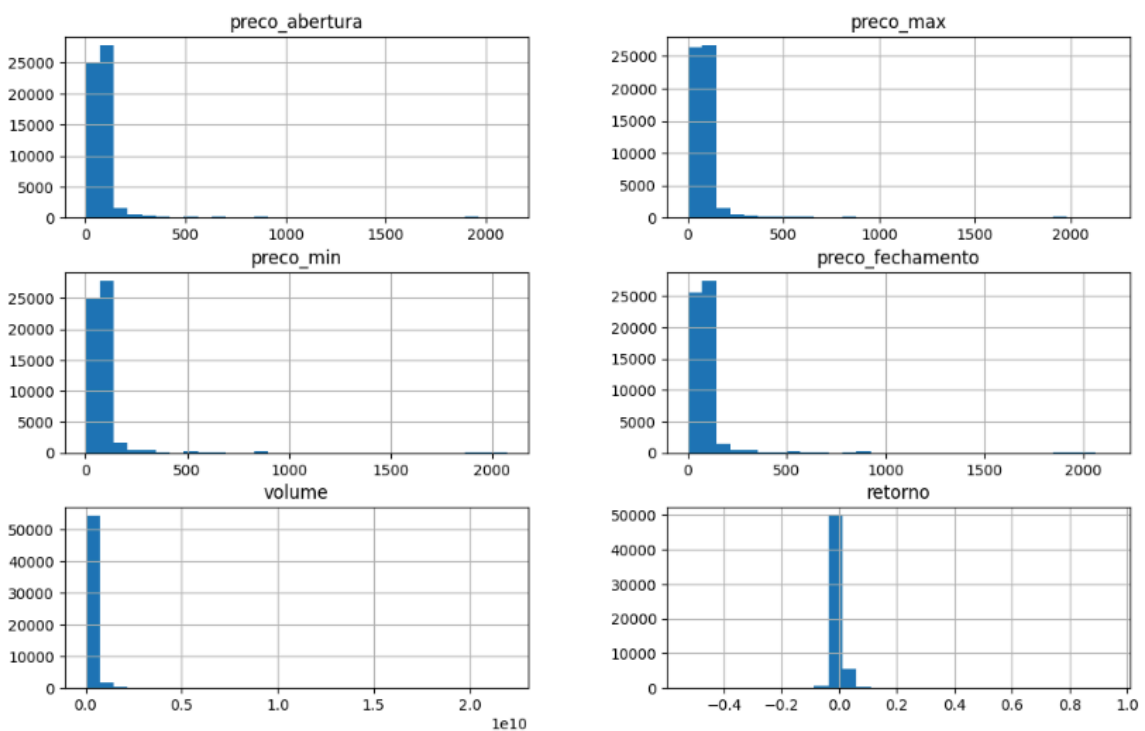
- 2022



- 2023



- 2024



- MATRIZ DE CORRELAÇÃO

Pela baixa quantidade de variáveis, não será incluído no estudo a matriz de correlação.

7.4. PRÉ PROCESSAMENTO DE DADOS

Para utilização dos dados nas redes neurais testadas, foram necessários alguns passos de pré processamento, para garantir o aprendizado integral das redes sobre a base de dados.

A primeira etapa foi organizar o dataset de forma temporal, ou seja, os dados mais antigos primeiro, seguindo uma ordem cronológica. Essa organização foi feita junto com o agrupamento de tickers, de forma que o mesmo ticker ficasse agrupado de ordem cronológica. Isso é importante principalmente para redes LSTM.

Em seguida, o dataset foi separado em um ticker específico de ativos de fundos de investimento imobiliários no período de 4 anos (de 2021 a 2024).

Após essa organização, o dataset foi dividido entre treino e teste. Esta etapa é essencial para modelos mais voltados para séries temporais, é importante que o dataset esteja dividido corretamente para que o modelo consiga aprender os padrões cronologicamente. Dessa forma, o dataset foi dividido em, dados de treinos com valores relacionados aos meses de janeiro a agosto, para que o modelo aprendesse com base na sazonalidade também e os dados de teste ficaram com os meses seguintes.

7.5. ARQUITETURAS APLICADAS

MLP: Essa arquitetura é um perceptron multicamada simples voltado para regressão. O modelo começa com uma camada densa de 64 neurônios com ativação ReLU, seguida por uma camada densa intermediária de 32 neurônios também com ReLU, permitindo a captura de padrões não lineares dos dados. A camada de saída contém um único neurônio linear para prever valores contínuos. O treinamento utiliza o otimizador Adam, função de perda MSE (erro quadrático médio) e métrica MAE (erro absoluto médio), com batch size de 32 e número de épocas configurável.

Transformers: O modelo foi estruturado com duas camadas Transformer Encoder, cada uma contendo Multi-Head Attention com 4 cabeças, normalização em camadas e rede feed-forward interna de 64 neurônios. Após o empilhamento dos encoders, os vetores foram achatados, passados por uma camada densa de 64 neurônios com ReLU e Dropout, e finalizados em uma camada de saída densa com um neurônio para regressão. O otimizador foi Adam com perda MSE.

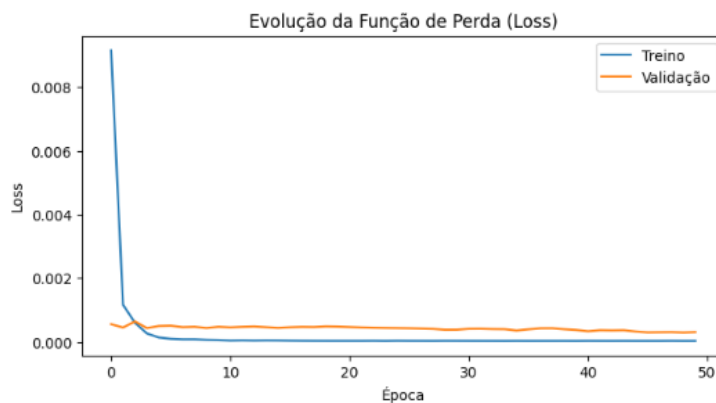
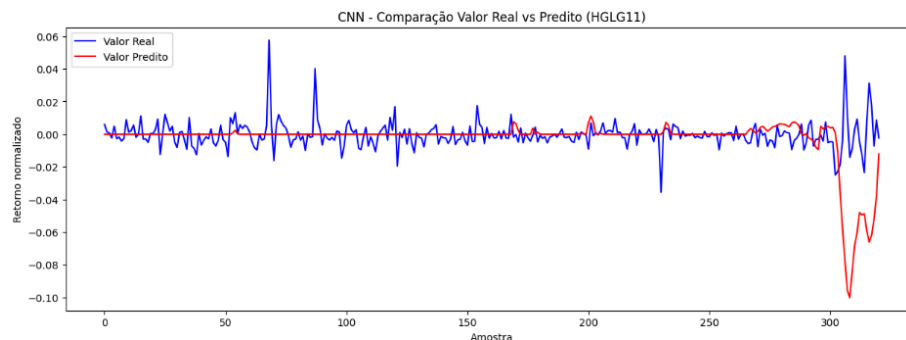
LSTM: Essa arquitetura foi baseada no modelo apresentado por [4], com uma diferença relacionada à normalização. Em seu trabalho, o autor normaliza os dados em uma camada específica da rede, o que permite a normalização em batch, porém a normalização feita foi usando Z-Score, para este estudo, optou-se pela normalização fora da rede neural, usando minMax. A arquitetura foi construída para capturar dependências temporais da série, começando com uma camada uma camada LSTM com 125 unidades e outra LSTM com 75 unidades com regularização L2 para evitar o overfitting. A saída é composta por um único neurônio com ativação sigmóide e regularização L1. O modelo foi treinado com otimizador Adagrad e função de perda binária no início, depois ajustado para regressão contínua com MSE.

CNN: A rede convolucional foi composta por duas camadas Conv1D para extrair padrões locais da série (64 filtros no primeiro bloco e 32 no segundo, ambos com kernel size 3 e ativação ReLU), seguidas de camadas de MaxPooling1D para reduzir dimensionalidade e Dropout para regularização. Após o Flatten, há uma camada densa de 64 neurônios ReLU e uma camada de saída densa com um neurônio para previsão de valores contínuos.

8. RESULTADOS

- CNN

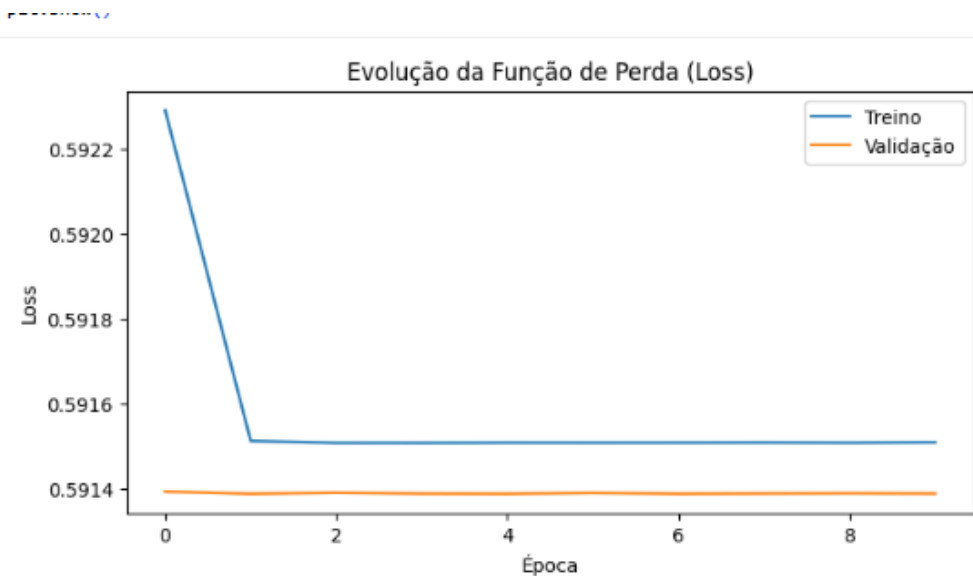
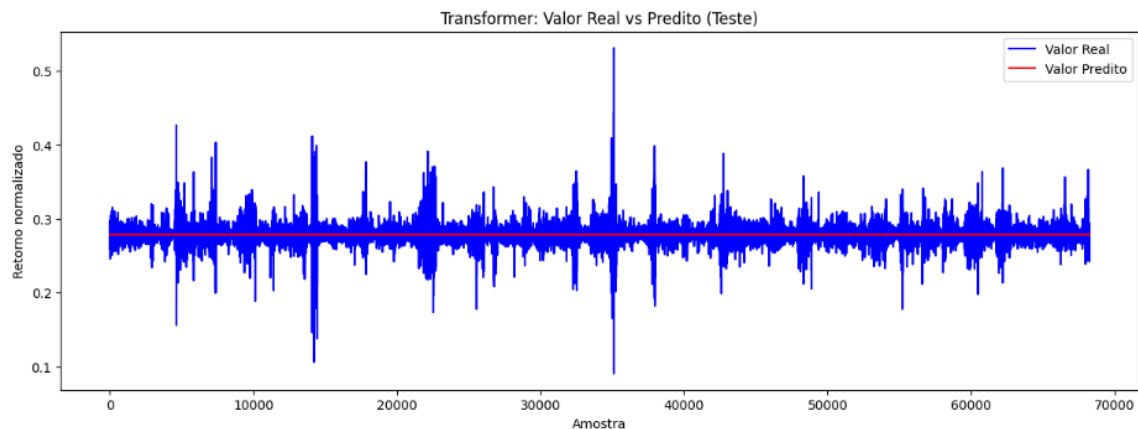
RMSE: 0.017541
MAE: 0.007930
MAPE: 7158870591.917319
R²: -3.694797



Embora os valores de RMSE e MAE indiquem erros absolutos baixos em termos de magnitude numérica, o valor extremamente elevado do MAPE e o R² negativo revelam que o modelo não foi capaz de generalizar adequadamente os padrões do conjunto de treino. Um coeficiente de determinação negativo implica que as previsões do modelo são piores do que uma simples média dos valores reais, sugerindo que a CNN não capturou as relações temporais e não lineares subjacentes à variação do retorno.

- TRANSFORMERS

RMSE: 0.008619
MAE: 0.004793
MAPE: 0.017349
R²: -0.000720



Apesar de os valores de erro absoluto (RMSE e MAE) parecerem baixos em termos numéricos, o coeficiente de determinação negativo revela que o modelo não conseguiu explicar a variância dos dados. Esse comportamento é consistente com o fato de que o Transformer predisse valores praticamente constantes próximos de zero, indicando que a rede não aprendeu padrões significativos durante o treinamento.

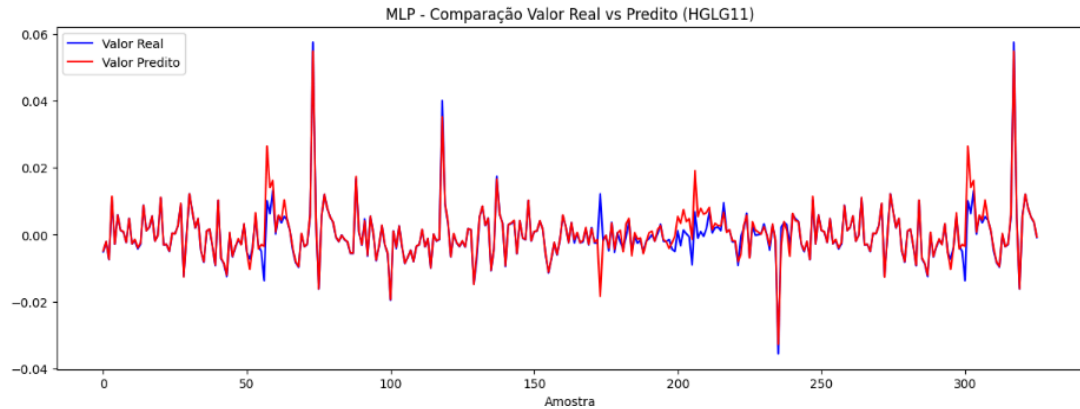
- MLP

RMSE: 0.002852

MAE: 0.000991

MAPE: 22684243453.417149

R²: 0.868868



O modelo MLP (Perceptron Multicamadas) apresentou resultados inconsistentes, sugerindo ocorrência de overfitting. Embora o coeficiente de determinação R^2 indique um bom ajuste aos dados de treinamento, o erro percentual médio MAPE demonstra que o modelo não conseguiu generalizar adequadamente para os dados de teste, resultando em previsões distorcidas e pouco realistas.

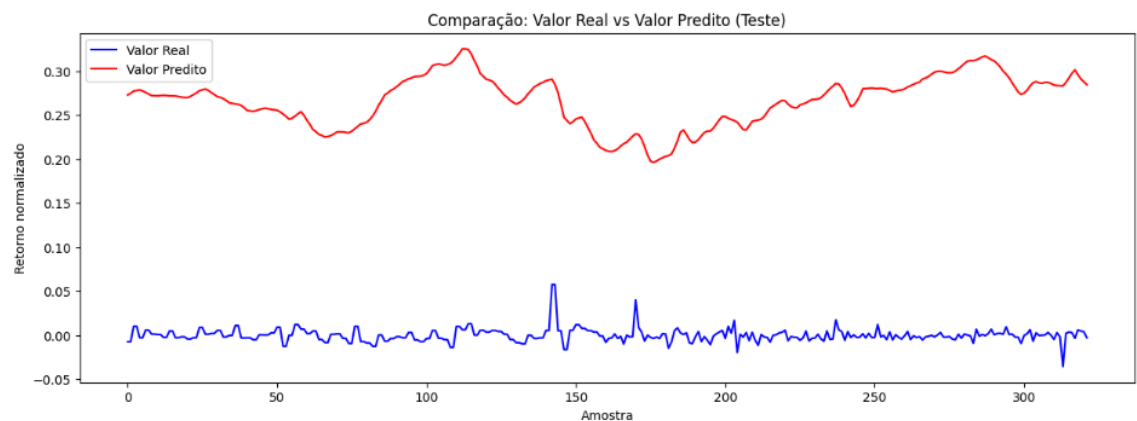
- LSTM

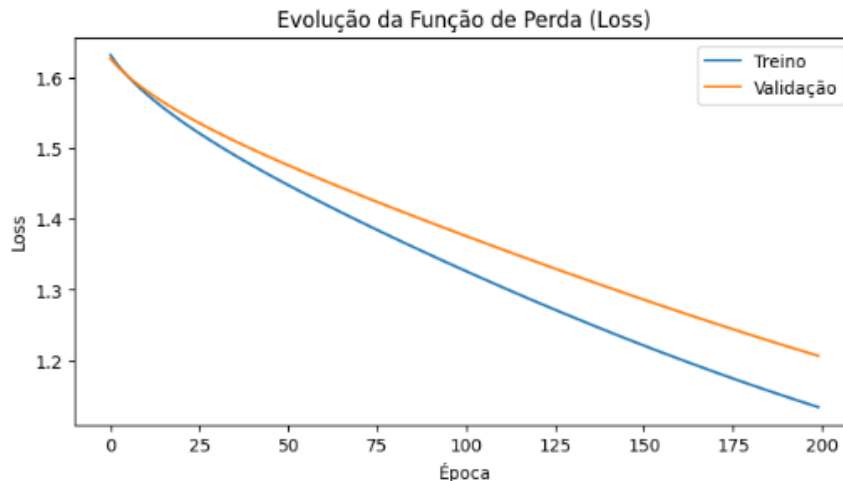
RMSE: 0.267844

MAE: 0.266264

MAPE: 22616927431685.050781

R^2 : -1145.322708





O modelo LSTM (Long Short-Term Memory) apresentou resultados insatisfatórios, evidenciando dificuldades em capturar os padrões temporais presentes nos dados. As métricas de desempenho foram significativamente elevadas, com RMSE e um MAPE extremamente alto, além de um R^2 que demonstra que o modelo performou muito pior do que uma simples média das amostras. O valor altamente negativo do R^2 reforça a hipótese de que o modelo falhou em generalizar, resultando em previsões altamente discrepantes dos valores reais.

9. CONCLUSÃO

O presente estudo teve como objetivo avaliar o desempenho de diferentes arquiteturas de deep learning aplicadas à previsão de séries temporais financeiras, utilizando como variável alvo o retorno dos ativos. Os resultados obtidos demonstraram, de forma geral, a complexidade intrínseca do mercado financeiro, onde a alta volatilidade, a presença de ruídos e a natureza não estacionária dos dados dificultam a construção de modelos preditivos robustos.

Em relação às hipóteses formuladas, foi possível observar que a H1 — que previa superioridade das arquiteturas recorrentes (LSTM) sobre redes feedforward — não se confirmou plenamente, uma vez que o modelo LSTM apresentou sinais de overfitting e dificuldade de generalização, sugerindo que, embora capture dependências temporais, sua capacidade de prever retornos futuros de forma consistente é limitada nesse contexto.

A H2, que propunha que modelos com mecanismos de atenção (Transformers) seriam mais eficazes em capturar ruídos e generalizar padrões, mostrou-se parcialmente válida. O Transformer apresentou bom desempenho em termos de R^2 positivo e baixo erro médio, mas, em determinadas execuções, tendeu a produzir previsões estáticas (como valores próximos de zero), indicando que sua eficiência depende fortemente da qualidade e quantidade dos dados de entrada.

A H3, referente ao menor overfitting em modelos de deep learning em relação a técnicas tradicionais, foi refutada parcialmente. Observou-se que, mesmo com grandes volumes de dados, as redes neurais podem apresentar overfitting quando a regularização e o controle de complexidade não são adequados. O comportamento da MLP reforça essa constatação:

apesar de sua simplicidade, ela apresentou forte overfitting, reproduzindo fielmente o conjunto de treinamento, mas com baixo poder de generalização.

Por sua vez, a H4 foi confirmada, visto que a rede MLP, por não incorporar mecanismos de memória ou convolução, mostrou-se menos eficiente em capturar dependências temporais e padrões sequenciais dos preços.

Finalmente, a H5 também foi corroborada, pois o aumento da complexidade dos modelos não resultou necessariamente em melhor desempenho. Verificou-se que, após certo ponto, o acréscimo de camadas e unidades aumenta o risco de overfitting, mesmo com o uso de técnicas como early stopping ou dropout.

De forma geral, conclui-se que a previsão de séries temporais financeiras continua sendo um desafio aberto, no qual modelos avançados de aprendizado profundo podem oferecer ganhos pontuais, mas estão longe de uma solução definitiva.

10. BIBLIOGRAFIAS

- [1] GUEDES, Anderson Cerqueira et al. **Computação evolutiva para otimização de carteiras de estratégias de negociação no mercado financeiro**. 2022.
- [2] SOARES, Taís Rigor. **Previsão de Retornos de Ativos no Mercado Financeiro Através de Métodos de Otimização**. 2023. Dissertação de Mestrado.
- [3] SHIMABUKURO, Camilo Ilzo. **Deep Learning aplicado à predição de retornos do Ibovespa: uma análise do desempenho da rede neural LSTM utilizando log-retornos e diferenciação fracionária**. 2024. Tese de Doutorado. Universidade de São Paulo.
- [4] MESQUITA, Caio Mário; OLIVEIRA, Renato; PEREIRA, Adriano César Machado. **Utilização de uma rede neural LSTM e teste da razão de variância para previsões em séries de ativos da Bovespa**, 2020.
- [5] GEZICI, Abdul Haluk Batur; SEFER, Emre. **Deep transformer-based asset price and direction prediction**. *IEEE Access*, v. 12, p. 24164-24178, 2024.