

# Robô de Alta Frequência Usando Inteligência Artificial para Predição de Tendências no Mercado de Ações

Henrique H. Cruvinel<sup>1</sup>, Maria J. P. Dantas<sup>2</sup>  
PUC Goiás, Goiânia, GO

O interesse em investimentos na bolsa de valores aumentou significativamente no Brasil após o mercado financeiro ter se tornado mais sólido e seguro. Em 2022, a B3 (Brasil Bolsa Balcão) atingiu cinco milhões de contas de investidores em renda variável somente no mês de janeiro (B3, 2022)[1].

A busca por métodos computacionais para prever movimentos no mercado financeiro tem sido afetada pelo chamado "dilema do caminho aleatório" (random walk dilemma, RWD). Conforme observado por Fama (1970; 1991)[2], a maioria das revisões não conseguiu rejeitar a hipótese de que as flutuações de preços seguem um padrão aleatório. Isso significa que nem todas as ações do mercado podem ser previstas ou determinadas com precisão. Esse é um desafio importante para os pesquisadores que buscam desenvolver métodos mais eficientes de previsão no mercado financeiro.

De acordo com Vasco (2020)[3], uma série temporal é uma sequência de valores ordenados no tempo, que pode ser utilizada para fazer previsões futuras. Cada valor na série é influenciado pelos valores anteriores, de forma que a ordem dos dados é fundamental.

O aprendizado de máquina (machine learning) consiste em utilizar vários neurônios artificiais que possuem entradas, camadas ocultas e saída. Os neurônios compartilham informações entre si para determinar um resultado (LECUN et al., 2015)[4]. Nesse contexto, a rede neural LSTM (Long Short-Term Memory) foi o modelo escolhido para ser utilizado no treinamento e previsão das ações do mercado financeiro.

Uma rede neural LSTM é uma variação de rede neural recorrente que foi projetada para lidar com problemas de sequência e memória de longo prazo. As redes LSTM são gerenciadas por "portões", sendo um deles a "porta de esquecimento" que é capaz de aprender e lembrar informações de longo prazo, tornando-a útil em tarefas como a previsão de séries temporais (HOCHREITER; SCHMIDHUBER, 1997)[5].

O algoritmo genético é uma das meta-heurísticas comumente utilizadas para otimização em problemas complexos, incluindo o aprendizado de máquinas (KURI e GALAVIZ, 2002)[6]. Neste projeto científico, pretende-se utilizar o algoritmo genético como otimizador para ajustar os pesos das camadas da rede neural. O desempenho de cada indivíduo será avaliado por meio do Erro Quadrático Médio (Mean Squared Error, MSE), uma função de avaliação amplamente utilizada em tarefas de regressão. Abaixo encontra-se a formulação matemática do MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

$Y_i$ : i-ésimo valor observado.

$\hat{Y}_i$ : valor predito correspondente.

$n$ : o número de observações.

---

<sup>1</sup>henriquehebert008@gmail.com

<sup>2</sup>mjpdantas@gmail.com

O tratamento de dados é um processo importante para fazer a análise antes de usá-la para o treinamento de dados de uma rede neural. Valores nulos, outliers e sazonalidades são alguns exemplos de processos prejudiciais para a previsão de dados e devem, portanto, ser tratados.

Quanto à preparação dos dados, foram selecionados no Yahoo Finance os ativos. Para exemplificar um dos ativos, utilizamos os dados das ações da Petróleo Brasileiro S.A. Petrobras Preference Shares (PETR4.SA) em frequência diária, do início de 2010 até o final de 2022. Posteriormente, os dados foram separados por data e valor de fechamento. A priori foram removidos os valores nulos e outliers da série temporal. Em seguida, verificou-se a normalidade da série pelo teste de Shapiro-Wilk e concluiu-se que a série não apresenta distribuição normal. Mesmo após tentativas de normalizá-la pelos métodos de ZScore, função logarítmica e Min-Max Scaler, a situação não melhorou. Ademais, verificou-se que a série temporal não passava no teste de estacionaridade e, portanto, foi feita uma diferenciação para melhorar os resultados. Após a diferenciação, verificou-se que a série temporal passa no teste de estacionaridade.

O Objetivo deste projeto é de desenvolver um algoritmo para prever os preços de ativos que compõem a bolsa de valores com métodos de redes neurais artificiais juntamente com algoritmo genético.

## Resultados Parciais

Para o algoritmo genético, foram aplicados os seguintes parâmetros: tamanho da população de 20, taxa de mutação de 5%, número de gerações de 20 e seleção por torneio.

A rede neural mostrou um avanço significativo nos resultados das previsões, alcançando o melhor desempenho com um MSE de 0,5. No entanto, apresenta demora no treinamento da rede neural.

## Agradecimentos

Agradecimento ao CNPQ pela bolsa PIBIC.

## Referências

- [1] B3. **B3 atinge 5 milhões de contas de investidores em renda variável em janeiro.** [https://www.b3.com.br/pt\\_br/noticias/5-milhoes-de-contas-de-investidores.htm](https://www.b3.com.br/pt_br/noticias/5-milhoes-de-contas-de-investidores.htm). Acesso em: 25 mar. 2022. 2022.
- [2] E. F. Fama e K. R. French. “Permanent and temporary components of stock prices”. Em: **Journal of Political Economics** 96.2 (1988), pp. 246–273. DOI: 10.1086/261535.
- [3] Lucas Pimenta Vasco. **Um estudo de redes neurais recorrentes no contexto de previsões no mercado financeiro.** <https://repositorio.ufscar.br/handle/ufscar/13730>. Acesso em: 31 mar. 2023. 2020.
- [4] Y. LeCun, Y. Bengio e G. Hinton. “Deep learning”. Em: **Nature** 521 (2015), pp. 436–444. DOI: 10.1038/nature14539.
- [5] S. Hochreiter e J. Schmidhuber. “Long Short-Term Memory”. Em: **Neural Comput** 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
- [6] M. Á. Kuri e C. J. Galaviz. **Algoritmos Genéticos.** México: Instituto Politécnico Nacional, Universidad Nacional Autónoma de México, Fondo de Cultura Económica, 2002, p. 60. ISBN: 9681663837.