



# Avaliação de Modelos de Aprendizado de Máquina na Detecção de Volatilidade Atípica em Séries Financeiras

Renata Cecconi de Souza<sup>1</sup>, Denis Fernando Wolf<sup>2</sup>  
ICMC-USP

## 1 Introdução

*Outliers* são observações que se desviam significativamente do padrão dos dados [4], podendo indicar erros ou eventos raros de grande impacto, como crises econômicas ou oscilações abruptas de preços. Identificá-los é crucial no contexto financeiro para ajustar estratégias de negociação e gestão de risco.

A detecção de *outliers* em séries temporais financeiras é desafiadora devido ao ruído nos dados e às dependências temporais complexas. Além disso, é necessário diferenciar entre *outliers* locais e globais. A volatilidade atípica ocorre quando há variações abruptas e inesperadas nos preços de ativos, muitas vezes causadas por eventos externos significativos [2]. Detectar essas anomalias em tempo hábil é essencial para investidores ajustarem suas estratégias.

Este estudo contempla cinco modelos de aprendizado de máquina, avaliados pelo *F1-Score*, métrica adequada para problemas desbalanceados, para a captura de volatilidade atípica.

## 2 Objetivos

Este estudo visa:

- Implementar e comparar cinco modelos de aprendizado de máquina na detecção de *outliers*: *Isolation Forest*, *LOF*, *One-Class SVM*, *LSTM Autoencoder* e *Anomaly Transformer*.
- Avaliar o desempenho dos modelos com base no *F1-Score*.
- Identificar as datas em que todos os modelos detectaram os mesmos *outliers*, garantindo consistência.

---

<sup>1</sup>renata.cecconi.souza@usp.br

<sup>2</sup>denis@icmc.usp.br

### 3 Metodologia

Foram utilizados dados históricos de dez ações líquidas (VALE3, PETR4, ITUB4, BBDC4, B3SA3, ABEV3, BBAS3, PETR3, PRIO3, RENT3) negociadas na B3, a Bolsa de Valores Brasileira, e o Dólar americano, no período de 1 de janeiro de 2014 a 1 de janeiro de 2024. As séries temporais foram processadas para obter os retornos logarítmicos e aplicar os modelos de aprendizado de máquina. A seguir, são apresentados os gráficos de identificação de volatilidade atípica para cada janela de 20, 30 e 60 dias, com e sem *outliers* sintéticos. Os modelos não assumem normalidade dos dados, utilizando abordagens robustas para identificar *outliers*. O critério  $3\sigma$  [3] foi aplicado por ser amplamente utilizado na literatura na detecção de anomalias .

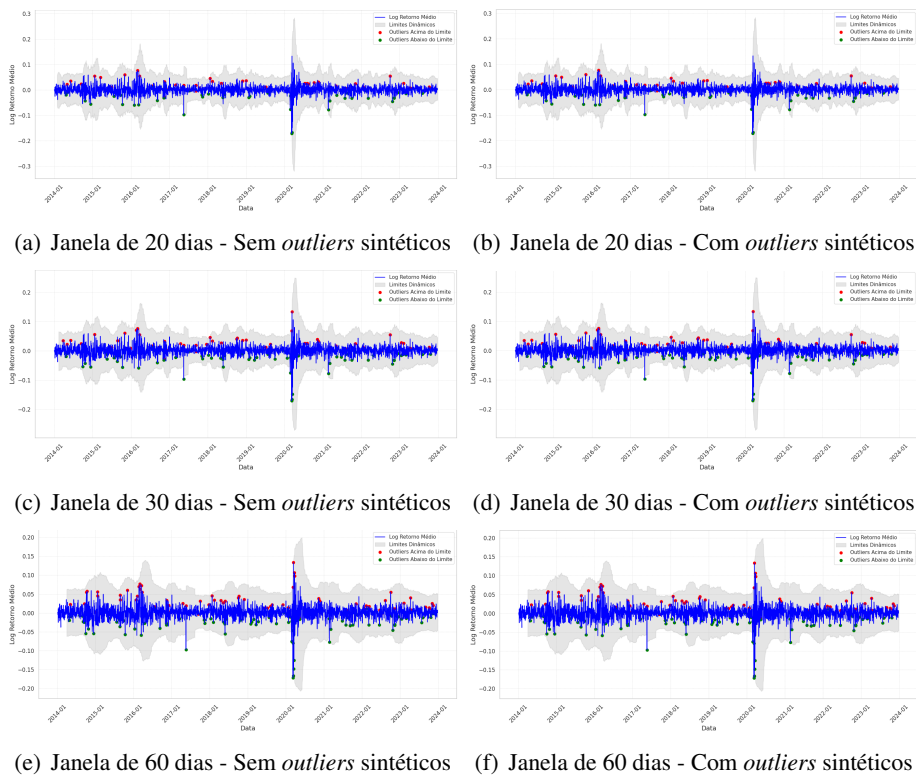


Figura 1: Detecção de *outliers* em diferentes janelas de análise, com e sem *outliers* sintéticos.

### 4 Resultados

#### 4.1 Modelos de Aprendizado de Máquina

Estando definidos pela metodologia os pontos identificados como *outliers* nas janelas móveis, aplicou-se os modelos de aprendizado de máquina para avaliar seus desempenhos.

- **Isolation Forest:** isola os *outliers* particionando os dados, sendo eficiente para detectar *outliers* globais [5].

- **Local Outlier Factor (LOF)**: mede a densidade local, identificando *outliers* em regiões menos densas [1].
- **One-Class SVM**: constrói um hiperplano para separar dados normais de anomalias [7].
- **LSTM Autoencoder**: rede neural que tenta reconstruir os dados; falhas significativas indicam *outliers* [6].
- **Anomaly Transformer**: baseado em *Transformer* [8], captura padrões de longo prazo [9].

O estudo avalia os modelos conforme a média do desempenho obtido nas seis janelas. O *Isolation Forest* apresentou o melhor desempenho geral em termos de *F1-Score* na tarefa de identificar volatilidade atípica, como exposto na Tabela 1. A figura 2 mostra o desempenho de cada modelo em cada janela de análise.

Tabela 1: Desempenho Geral dos Modelos (média das seis janelas móveis)

Modelo	Acurácia	Precisão	Revocação	<i>F1-Score</i>
<i>Isolation Forest</i>	0,973	0,986	0,978	0,983
<i>Local Outlier Factor</i>	0,954	0,970	0,960	0,976
<i>Anomaly Transformer</i>	0,934	0,951	0,945	0,966
<i>LSTM Autoencoder</i>	0,932	0,950	0,940	0,965
<i>One-Class SVM</i>	0,919	0,931	0,926	0,956

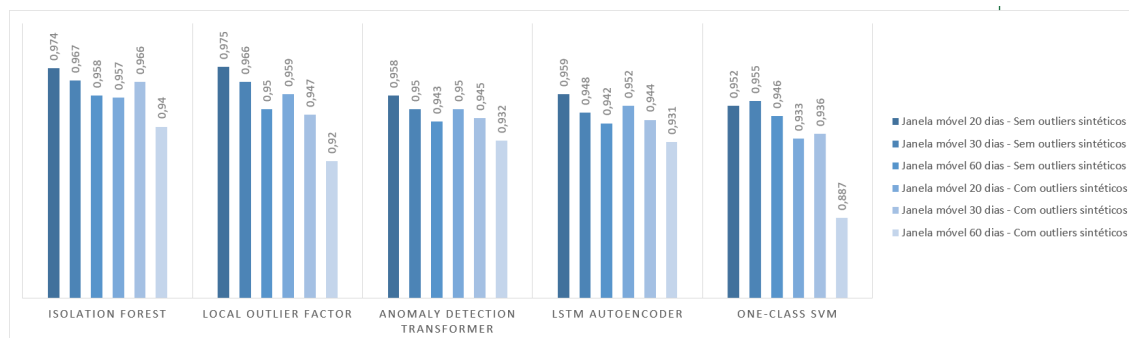


Figura 2: Desempenho de cada modelo por janela de análise

A Tabela 2 apresenta as datas em que todos os modelos identificaram os mesmos eventos como *outliers* considerando todas as janelas.

Tabela 2: Datas identificadas como *outliers* por todos os modelos

2018	2020	2021
31-jul	16-mar	09-mar
26-set	17-mar	10-mar
28-set	18-mar	11-mar
01-out	19-mar	12-mar
02-out	20-mar	15-mar

Essas datas coincidem com eventos significativos que afetaram os mercados financeiros, como o início da pandemia de COVID-19 e tensões políticas, resultando em aumento da volatilidade.

## 5 Conclusão

O *Isolation Forest* destacou-se como o melhor modelo, apresentando o maior *F1-Score* devido à sua capacidade de isolar rapidamente *outliers* ao particionar os dados em subespaços, o que o torna eficiente para detectar anomalias globais em grandes conjuntos de dados. Como o método não pressupõe distribuição normal dos dados, ele é mais robusto em cenários de alta volatilidade e ruído, comuns em séries financeiras. O LOF também apresentou bom desempenho, principalmente em identificar *outliers* locais, devido à sua abordagem baseada em densidade. Já o *Anomaly Transformer* e o *LSTM Autoencoder*, por utilizarem arquiteturas mais complexas, são melhores na detecção de padrões de longo prazo, o que justifica seu desempenho ligeiramente inferior em relação aos métodos mais simples em um conjunto de dados onde a detecção de *outliers* pontuais é mais relevante do que a análise de tendências de longo prazo. Esses resultados indicam que, em ambientes de alta volatilidade, como o mercado financeiro, modelos mais simples e diretos podem oferecer uma performance superior na detecção de *outliers*, enquanto modelos mais complexos podem ser mais adequados para outras aplicações, como previsão de séries temporais.

## Referências

- [1] M. Breunig, H. Kriegel, R. Ng, J. Sander, *LOF: Identifying Density-Based Local Outliers*, ACM SIGMOD, 2000.
- [2] R. F. Engle, *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation*, *Econometrica*, 1982.
- [3] FREEDMAN, D.; PISANI, R.; PURVES, R. *Statistics*. 1st. ed. New York, NY: W. W. Norton & Company, 1978. ISBN 978-0-393-01352-8.
- [4] D. Hawkins, *Identification of Outliers*, Chapman and Hall, 1980.
- [5] F. T. Liu, K. M. Ting, Z. Zhou, *Isolation Forest*, IEEE ICDM, 2008.
- [6] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, *LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection*, arXiv, 2016.
- [7] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, *Estimating the Support of a High-Dimensional Distribution*, *Neural Computation*, 2001.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, I. Polosukhin, *Attention is All You Need*, *NeurIPS*, 2017.
- [9] Jiehui Xu, Haixu Wu, Jianmin Wang, Mingsheng Long, *Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy*, International Conference on Learning Representations, 2022. Disponível em: [https://openreview.net/forum?id=LzQQ89U1qm\\_](https://openreview.net/forum?id=LzQQ89U1qm_).