

# Impacto de técnicas de normalização de conjunto de dados na predição de Ethereum em rede neural convolucional com validação por testes estatísticos

Eduardo José Costa Lopes

<sup>1</sup>Centro Universitário FEI  
CEP 09850-901 – São Bernardo do Campo – SP – Brazil

elopes@fei.edu.br

**Abstract.** *A principal contribuição desta pesquisa é complementar estudos que abordam um problema de previsão de série temporal que compara arquiteturas de redes profundas para a criptomoeda Ethereum (ETH). Este estudo adiciona a comparação de diferentes técnicas de normalização de dados para o modelo de rede neural mais preciso encontrado (rede neural convolucional - TCN). Foi realizado a normalização dos dados de entrada considerando várias técnicas, obtendo-se os valores de acurácia da rede neural citado anteriormente para os diferentes tipos de normalização. Os resultados mostraram que, a normalização MinMax considerada anteriormente nos estudos de comparação de modelos de redes neurais não trouxe melhores resultados validados estatisticamente por teste de hipótese apropriados para a distribuição dos dados utilizados, ainda que a mesma apresente médias de acuracidade mais baixas que as demais técnicas consideradas no trabalho.*

## 1. Introdução

O objetivo deste trabalho é realizar uma comparação entre os vários modelos de normalização de dados aplicados a uma série temporal para avaliar o impacto dos mesmos no resultado de predição de redes neurais convolucionais, sendo as mesmas as mais acuradas em estudos comparativos de arquiteturas de redes neurais para o caso de uso [Lopes 2022]. Tais resultados foram validados estatisticamente através de testes de hipótese com a devida verificação de suas distribuições através de testes de ajuste de bondade de curvas.

### 1.1. Número de Amostras

A objetivo da definição do tamanho da amostra é (i) produzir uma representação justa da população onde as estatísticas amostrais possam ser usadas para inferência populacional e (ii) meios eficientes para estudos populacionais. O primeiro requisito ajuda a minimizar vieses, enquanto o segundo requisito ajuda na conservação de recursos. O tamanho mínimo da amostra responde a pergunta: “quantos dados são suficientes?” [Louangrath 2017]. Dado que validações estatísticas são usadas neste trabalho, devemos garantir que os dados considerados estejam dentro das premissas das técnicas estatísticas utilizadas. De acordo com [Islam 2018], o número mínimo de amostras para que o Teorema do Limite Central [Kwak and Kim 2017] possa ser considerado é de 30 amostras.

## 1.2. Normalização de dados

A normalização é uma etapa de pré-processamento de qualquer tipo de declaração de problema. Especialmente a normalização assume um papel importante no campo da computação para manipulação de dados, como reduzir ou aumentar a gama de dados antes de serem usados em estágios posteriores. Dados não normalizados podem também retardar ou até mesmo impedir a convergência de muitos estimadores baseados em gradientes, como aqueles encontrados em mecanismos de retroprogração de parâmetros de redes neurais. Existem várias técnicas de normalização, dentre as quais foram consideradas:

1. **MinMaxScaler**: redimensiona o conjunto de dados de forma que todos os valores de recursos estejam no intervalo  $[0, 1]$ .
2. **QuantileTransformer**: fornece transformações não lineares nas quais as distâncias entre valores discrepantes e inliers marginais são reduzidas.
3. **RobustScaler**: remove a mediana e dimensiona os dados de acordo com o intervalo quantílico (o padrão é IQR: intervalo interquartil). O IQR é o intervalo entre o 1º quartil (25º quantil) e o 3º quartil (75º quantil).
4. **MaxAbsScaler**: Este estimador dimensiona e traduz cada recurso individualmente, de modo que o valor absoluto máximo de cada recurso no conjunto de treinamento seja 1,0. Ele não desloca/centraliza os dados e, portanto, não destrói qualquer dispersão.
5. **BoxCox**: Transformação paramétrica e monotônica para tornar os dados mais gaussianos. Esta normalização exige que os dados de entrada sejam estritamente positivos, enquanto Yeo-Johnson suporta dados positivos ou negativos. Por padrão, a normalização de média zero e variância unitária é aplicada aos dados transformados.

A normalização **MinMaxScaler** é muito sensível a presença de valores discrepantes. Em dados apenas positivos, tanto **MinMaxScaler** quanto **MaxAbsScaler** se comportam de maneira semelhante.

## 1.3. Testes Estatísticos

Mais do que resultados de métricas de precisão de valor único são necessários para considerar se um modelo de aprendizado profundo tem melhor desempenho. Os valores de inicialização dos hiperparâmetros e do modelo podem afetar significativamente os resultados obtidos por esses modelos, introduzindo assim a variabilidade dos resultados por modelo. Normalmente, o teste de T-student pareado [Student 1908] que é usado para comparar amostras médias é baseado em suposições ou parâmetros particulares: as amostras de dados que medem esses parâmetros são sorteadas aleatoriamente de uma população média, com base em observações independentes, medidas com uma escala de intervalo ou razão, possuem um tamanho adequado e se assemelham aproximadamente a uma distribuição normal [Corder and Foreman 2014]. Como essas condições são desconhecidas, o teste de bondade de ajuste de Kolmogorov-Smirnov [Massey Jr 1951] será usado para determinar se as amostras de dados obtidas atendem a níveis aceitáveis de normalidade, exigindo testes não paramétricos adicionais.

### 1.3.1. Teste de Kolmogorov-Smirnov

É um procedimento para examinar a concordância entre dois conjuntos de valores comparando duas distribuições de frequência cumulativas. Uma distribuição de frequência cumulativa ajuda a encontrar o número de observações acima ou abaixo de um valor específico em uma amostra de dado [Corder and Foreman 2014]. Os testes bilaterais consideram a hipótese nula como se duas distribuições fossem idênticas,  $F(x)=G(x)$  para todos os  $x$ , e a alternativa é que elas não são idênticas.

### 1.3.2. Teste de Wilcoxon

O teste não paramétrico de Wilcoxon é um teste de hipótese usado para comparar amostras agrupadas para determinar se as classificações médias da população diferem [Derrick and White 2017]. O teste unilateral considerado neste trabalho tem a hipótese nula de que a média é positiva contra a alternativa de que é negativa.

## 2. Trabalhos relacionados

Base deste trabalho, uma rede neural convolucional (CNN) é uma classe de redes neurais profundas mais comumente aplicada a tarefas de visão computacional. As séries temporais também podem ser tratadas como uma imagem unidimensional que um modelo CNN pode analisar [Goodfellow et al. 2016]. Novas abordagens que aplicam conceitos de CNN podem ser encontradas, como Redes Convolucionais Temporais (TCN) [Chen et al. 2020]. Os resultados de conjuntos de dados industriais e públicos mostram que a estrutura produz um desempenho superior em comparação com outros métodos de última geração, tanto em previsão pontual como probabilística. O trabalho realizado neste artigo amplia os trabalhos de [Lopes 2022].

## 3. Métodos e Materiais

Para alcançar os resultados deste trabalho foram utilizados cinco tipos de normalização diferentes. Em seguida, foi realizada uma avaliação de acurácia usando uma rede neural convolucional (TCN). Em resumo, o processo consiste em cinco etapas: (1) coleta de dados históricos de criptomoedas; (2) exploração e visualização de dados; (3) normalização; (4) execução de rede neural; (5) extração e comparação dos resultados das métricas de acurácia do modelo com validação estatística.

### 3.1. Critérios de Avaliação

Foi considerada uma combinação de testes estatísticos e métricas de precisão do modelo.

De acordo com [Pintelas et al. 2020], o Erro Médio Médio (MAE) e o Erro Quadrático Médio (RMSE) são as métricas de desempenho mais utilizadas para avaliar o desempenho de regressão dos modelos de previsão. Dado que:  $n$  = número de amostras,  $x_i$  = valor real da série no tempo ( $i$ ) e  $y_i$  = valor previsto da série no tempo ( $i$ ):

$$MAE = \left(\frac{1}{n}\right) \sum_{i=1}^n |y_i - x_i| \quad (1)$$

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (x_i - y_i)^2} \tag{2}$$

### 3.2. Bibliotecas e Técnicas

O conjunto de dados de Ethereum foi coletado do sitio web [Glassnode 2022]. A normalização da base de dados visa alterar os valores das colunas numéricas nos dados de entrada do conjunto para uma escala padrão sem distorcer as diferenças nos intervalos de dados. Essas normalizações consideraram os conjuntos de treinamento, validação e teste de forma separada.

### 4. Resultados e Discussão

Os códigos utilizados em python, configuração de rede neural e seus hiperparâmetros bem como seus resultados de acurácia podem ser encontrados no repositório github [Lopes 2024].

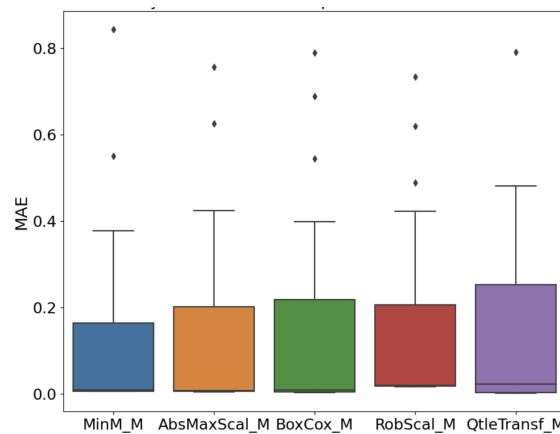
Com base nos testes de Kolmogorov-Smirnov realizados no conjunto completo de resultados composto por 30 amostras calculadas com trinta valores iniciais diferentes (0–29) de semente aleatória de inicialização da rede neural “seed”, a Tabela 1 mostra que nenhuma normalização exibe um comportamento de distribuição normal com 1 por cento de significância ao rejeitar a hipótese nula do teste (os valores p são menores que o nível de significância  $\alpha = 0,01$ ), justificando assim a necessidade de testes não paramétricos.

Tabela 1. Resultados do Teste de Kolmogorov-Smirnov.

Normalização	p-value	
	MAE	RMSE
MinMax	$4.25 \times 10^{-12}$	$1.89 \times 10^{-11}$
AbsMaxScaler	$1.70 \times 10^{-10}$	$9.77 \times 10^{-9}$
BoxCox	$5.66 \times 10^{-10}$	$6.22 \times 10^{-9}$
RobustScaler	$9.26 \times 10^{-11}$	$1.15 \times 10^{-8}$
QuantileTransformer	$2.72 \times 10^{-9}$	$5.10 \times 10^{-8}$

Tabela 2. Resultados do Teste de Wilcoxon - Acurácia MAE.

Técnica A	Técnica B	p-value	Menor Lista - média
MinMax	AbsMaxScaler	0.640	MinMax
MinMax	BoxCox	0.439	MinMax
MinMax	RobustScaler	<b>0.015</b>	MinMax
MinMax	QuantileTransformer	0.839	MinMax
AbsMaxScaler	BoxCox	0.370	AbsMaxScaler
AbsMaxScaler	RobustScaler	<b>0.001</b>	AbsMaxScaler
AbsMaxScaler	QuantileTransformer	0.855	AbsMaxScaler
BoxCox	RobustScaler	<b>0.047</b>	BoxCox
BoxCox	QuantileTransformer	0.903	QuantileTransformer
RobustScaler	QuantileTransformer	0.349	QuantileTransformer



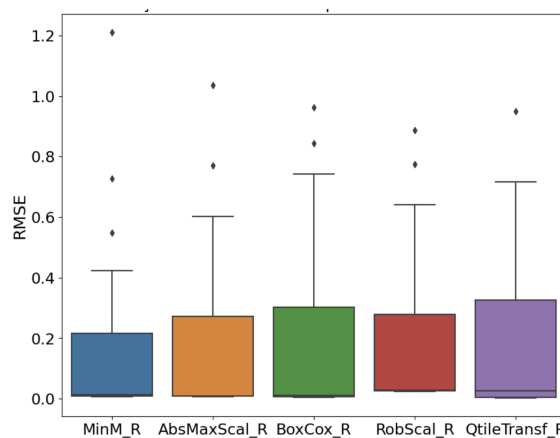
**Figura 1. BoxPlot - Métrica MAE para as diferentes técnicas de normalização.**

Podemos verificar para a métrica de acurácia MAE que:

1. A técnica RobustScaler frequentemente apresenta maiores valores médios de MAE comparado a outras técnicas.

**Tabela 3. Resultados do Teste de Wilcoxon - Acurácia RMSE.**

<i>Técnica A</i>	<i>Técnica B</i>	<i>p-value</i>	<i>Menor Lista - média</i>
MinMax	AbsMaxScaler	0.655	MinMax
MinMax	BoxCox	0.393	MinMax
<b>MinMax</b>	<b>RobustScaler</b>	<b>0.017</b>	<b>MinMax</b>
MinMax	QuantileTransformer	0.807	MinMax
AbsMaxScaler	BoxCox	0.183	AbsMaxScaler
<b>AbsMaxScaler</b>	<b>RobustScaler</b>	<b>0.002</b>	<b>AbsMaxScaler</b>
AbsMaxScaler	QuantileTransformer	0.583	AbsMaxScaler
BoxCox	RobustScaler	0.066	BoxCox
BoxCox	QuantileTransformer	0.670	QuantileTransformer
RobustScaler	QuantileTransformer	0.114	QuantileTransformer



**Figura 2. BoxPlot - Métrica RMSE para as diferentes técnicas de normalização.**

Para a métrica de acurácia RMSE, pode-se concluir que:

1. A técnica QuantileTransformer é frequentemente listada como a "Menor Lista" em várias comparações. A quantidade de p-values significativos (menores que 0,05) é baixa, indicando que muitas diferenças não são estatisticamente significativas.
2. O gráfico de boxplot da figura 2 mostra que a técnica QuantileTransformer tem uma menor mediana e menor dispersão de valores RMSE.

## 5. Conclusão e Trabalhos futuros

Este trabalho apresentou a comparação de técnicas diferentes de normalização aplicadas a uma série temporal e verificou a acurácia dos resultados produzidos por um modelo de rede neural específico. Através de testes de hipótese, não se pode verificar que uma técnica específica tenha resultados estatisticamente melhores que as demais. Trabalhos futuros na área de aprendizado profundo e redes neurais deverão avaliar as técnicas apropriadas de normalização que garantam a melhor acurácia.

## Referências

- Chen, Y., Kang, Y., Chen, Y., and Wang, Z. (2020). Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399:491–501. page.33
- Corder, G. W. and Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons. page.22, page.33
- Derrick, B. and White, P. (2017). Comparing two samples from an individual likert question. *International Journal of Mathematics and Statistics*, 18(3). page.33
- Glassnode, W. (2022). On-chain data. <https://docs.glassnode.com/general-info/on-chain-data>. Accessed: 2022-06-09. page.44
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press. page.33
- Islam, M. R. (2018). Sample size and its role in central limit theorem (clt). *Computational and Applied Mathematics Journal*, 4(1):1–7. page.11
- Kwak, S. G. and Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, 70(2):144. page.11
- Lopes, E. (2024). Impacto de técnicas de normalização de conjunto de dados na predição de ethereum em rede neural convolucional com validação por testes estatísticos - github site. page.44
- Lopes, E. J. C. (2022). Uso de redes neurais profundas para previsão de curto prazo do preço da criptomoeda ethereum. page.11, page.33
- Louangrath, P. (2017). Minimum sample size method based on survey scales. *Int. J. Res. Methodol. Soc. Sci*, 3(3):44–52. page.11
- Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78. page.22
- Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., and Pintelas, P. (2020). Investigating the problem of cryptocurrency price prediction: a deep learning approach. In *IFIP International conference on artificial intelligence applications and innovations*, pages 99–110. Springer, Springer. page.33
- Student (1908). The probable error of a mean. *Biometrika*, pages 1–25. page.22