

# Um Modelo de IA para Predição De Vazão De Rede considerando as Características da Janela de Congestionamento do TCP

Ariel Lima De Carvalho Portela<sup>1</sup>, Rafael Lopes Gomes<sup>1</sup>

<sup>1</sup>Universidade Estadual do Ceará (UECE)

ariel.portela@aluno.uece.br, rafa.lopes@uece.br

**Abstract.** *Performance prediction is essential for proactive network management, particularly for identifying bottlenecks through throughput forecasting. Since congestion control algorithms like TCP Cubic and TCP BBR significantly influence traffic patterns through distinct operational mechanisms, they directly impact the accuracy of predictive models. This dissertation proposes a forecasting framework utilizing LSTM and GRU neural networks to predict throughput performance, explicitly accounting for the applied TCP algorithm and its statistical behavior. Experiments conducted with real-world data from the Brazilian National Research and Educational Network (RNP) evaluate the impact of TCP congestion control on the predictive process. The study results underscore the necessity of aligning predictive models with the specific characteristics of the network environment and the employed congestion control mechanisms to optimize operational management.*

**Resumo.** *A predição de desempenho é essencial para o gerenciamento proativo de redes, sendo a predição de vazão crucial para a antecipação de gargalos. Visto que algoritmos de controle de congestionamento, como TCP Cubic e TCP BBR, impõem dinâmicas distintas ao tráfego, suas características influenciam diretamente a precisão dos modelos preditivos. Este trabalho apresenta a dissertação que propõe uma abordagem de predição de vazão utilizando redes neurais de aprendizado profundo, considerando explicitamente o algoritmo TCP aplicado e seu comportamento estatístico. Experimentos realizados com dados reais da Rede Nacional de Ensino e Pesquisa (RNP) avaliam o impacto do controle de congestionamento no processo preditivo. Os resultados reforçam a necessidade de alinhar modelos preditivos às especificidades do protocolo de transporte e do ambiente de rede para otimizar decisões operacionais.*

## 1. Introdução

A predição de desempenho assume um papel crucial no gerenciamento de redes, permitindo que administradores antecipem mudanças e preparem a infraestrutura através da análise de dados históricos e modelagem preditiva [Silveira et al. 2023]. O objetivo central é a identificação precoce de gargalos, possibilitando ações preventivas que garantam a alocação eficaz de recursos, especialmente durante períodos de tráfego de pico, evitando a degradação do serviço.

Essa postura proativa não apenas otimiza a eficiência de custos ao evitar superdimensionamentos [Miyazawa et al. 2020], mas é fundamental para minimizar pro-

blemas como lentidão e desconexões [Gomes et al. 2016, Gomes et al. 2013]. Ao garantir a resiliência da rede e apoiar o planejamento de infraestrutura a longo prazo, a predição assegura que a capacidade evolua em alinhamento com os requisitos dos usuários [Kanaya et al. 2020].

No entanto, a precisão da predição de Vazão depende intrinsecamente do comportamento dos algoritmos de Controle de Congestionamento TCP. Algoritmos distintos, como o TCP Cubic e o TCP BBR, gerenciam o envio de dados e respondem ao congestionamento de maneiras significativamente diferentes, introduzindo variáveis que devem ser isoladas e analisadas para um prognóstico correto.

Nesse contexto, este trabalho descreve a dissertação que apresenta um modelo preditivo de Vazão baseado no tipo de algoritmo TCP utilizado, empregando redes LSTM e GRU. Experimentos realizados com dados reais da Rede Nacional de Ensino e Pesquisa (RNP) demonstram que a incorporação das características específicas do TCP permite alcançar baixos níveis de erro, validando o impacto positivo dessa abordagem no gerenciamento eficaz da rede.

De maneira geral a dissertação possui as seguintes contribuições: (1) Análise do impacto dos algoritmos de controle de congestionamento TCP (Cubic e BBR) no comportamento estatístico da vazão de rede; (2) Proposição de um modelo de predição de vazão sensível ao algoritmo TCP, superando abordagens genéricas que ignoram esse fator; (3) Utilização de dados reais, aumentando a validade prática dos resultados; (4) Definição de uma metodologia completa de coleta, pré-processamento e análise de séries temporais de rede, incluindo interpolação de dados ausentes; (5) Avaliação dos modelos por meio de métricas quantitativas (RMSE, acurácia) e métricas de similaridade temporal (DTW e correlação de Pearson); e, (6) Contribuição prática para planejamento de capacidade, alocação de recursos e gerenciamento proativo de redes.

## **2. Trabalhos relacionados**

Diversos estudos abordam a predição de tráfego de rede combinando métodos estatísticos e de aprendizado de máquina para capturar padrões complexos. Yang et al. [Yang et al. 2021] propõem um modelo híbrido ARIMA-BPNN otimizado por SA para identificar características lineares e não lineares, embora a abordagem não contemple a predição de desempenho da rede. Em uma análise complementar, Sone et al. [Sone et al. 2020] investigam a sazonalidade e a relação não linear entre número de usuários e volume de dados em redes corporativas, aplicando modelos como SARIMA, LSTM e GRU para compreender dinâmicas temporais e espaciais.

Outras abordagens focam em cenários específicos ou em técnicas avançadas de otimização e incerteza. Sepasgozar e Pierre [Sepasgozar and Pierre 2022] desenvolvem um modelo restrito a redes veiculares (VANETs), correlacionando tráfego de rede e rodoviário, enquanto Redhu et al. [Redhu et al. 2023] aplicam otimização por enxame de partículas (PSO) em redes Bi-LSTM para ajuste global de parâmetros. Adicionalmente, Wang et al. [Wang et al. 2024] introduzem uma estrutura híbrida com redes bayesianas (BNN) e LSTM, visando não apenas a predição em tempo real, mas também a quantificação de incertezas em dados sequenciais.

Apesar desses avanços significativos, a literatura ainda carece de uma abordagem abrangente que detalhe o impacto específico dos algoritmos de Controle de Congestio-

namento TCP sobre o comportamento estatístico da vazão. Essa lacuna é crítica, especialmente em cenários que simulam condições reais e dados heterogêneos, pois limita a compreensão de como esses protocolos interagem com as dinâmicas de redes contemporâneas, comprometendo a eficácia prática das soluções de predição existentes.

### 3. Proposta

Redes de alto desempenho, como as utilizadas em grandes infraestruturas acadêmicas e de pesquisa, são compostas por rotas e conexões complexas que apresentam comportamentos variáveis ao longo do tempo. Esses comportamentos são influenciados diretamente pelos algoritmos de controle de congestionamento TCP, como TCP Cubic e TCP BBR, que adotam estratégias distintas para gerenciar o tráfego de rede e lidar com situações adversas. Por exemplo, enquanto o TCP Cubic reage a perdas de pacotes de forma agressiva, reduzindo a taxa de envio, o TCP BBR utiliza estimativas de largura de banda e tempo de ida e volta para otimizar o desempenho [Ferreira et al. 2024].

Essas características aumentam a complexidade da predição, uma vez que a variabilidade dos dados torna os modelos de predição mais suscetíveis a erros. Com base nisso, este trabalho propõe um sistema de predição de Vazão utilizando modelos de redes neurais, como LSTM e GRU, com o suporte de técnicas de pré-processamento e avaliação estatística.

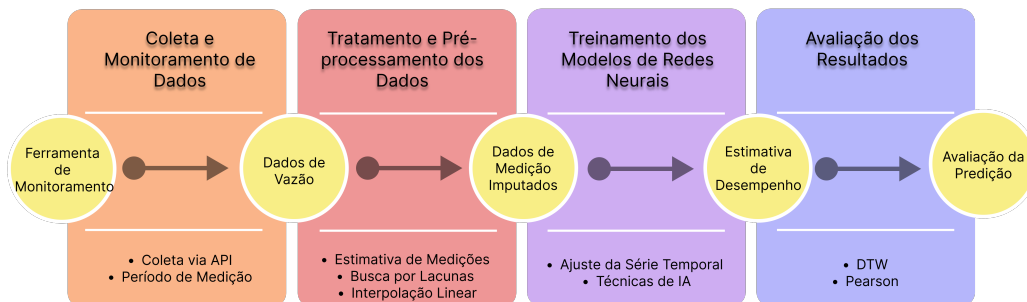


Figura 1. Sistema Proposto

Como proposta, foi desenvolvido um sistema para predição de desempenho baseado em modelos preditivos de redes neurais e análise estatística das séries temporais. O sistema é projetado para funcionar em duas etapas principais: a primeira, focada no tratamento e análise dos dados de Vazão obtidos da RNP, e a segunda, voltada para o treinamento e avaliação dos modelos preditivos.

O sistema proposto realiza as seguintes etapas: (I) Coleta e Monitoramento de Dados; (II) Tratamento e Pré-processamento dos Dados; (III) Treinamento dos Modelos de Redes Neurais; e (IV) Avaliação dos Resultados.

#### 3.1. Coleta e Monitoramento

A Coleta e Monitoramento de Dados é a etapa inicial do sistema de predição, sendo fundamental para gerar uma base de informações confiável e adequada para alimentar os modelos preditivos. Dessa forma, foram utilizados dados da RNP por meio do Serviço de Monitoramento da Rede Ipê (MonIPÊ) para realizar experimentos baseados em informações

do mundo real. O MonIPÊ adota o padrão internacional de monitoramento perfSONAR, no qual as medições de Vazão são realizadas a cada 4 horas, enquanto as medições de perda e atraso ocorrem a cada 5 minutos [Ribeiro et al. 2023].

A rede RNP interliga todos os estados brasileiros por meio de links que podem estabelecer conexões diretas, quase diretas ou atravessar múltiplos intermediários. Para melhor compreensão, essas rotas são classificadas como Recorrentes (R) quando percorrem poucos ou nenhum link intermediário e Não Recorrentes (NR) quando exigem a passagem por diversos links antes de alcançar o destino final.

Diante desse cenário, a Vazão de rede foi escolhida como foco de análise por sua relevância na avaliação da capacidade de transmissão das aplicações. Essa abordagem também permite investigar o comportamento de diferentes algoritmos de controle de congestionamento TCP, como Cubic e BBR, no contexto de processos de predição.

A extração e a análise dos dados fornecidos pela plataforma MonIPÊ desempenham um papel essencial em diagnósticos da infraestrutura de rede e planejamento de capacidade.

### 3.2. Tratamento e Pré-processamento

A etapa de pré-processamento é fundamental para assegurar a qualidade dos dados de entrada, embora arquiteturas recorrentes como LSTM e GRU apresentem robustez frente a séries temporais não estacionárias [Lara-Benítez et al. 2021]. Para a compreensão da estrutura subjacente da série temporal, composta por Tendência ( $T$ ), Sazonalidade ( $S$ ) e Erro ( $E$ ), empregou-se a decomposição STL (Seasonal-Trend decomposition using LO-ESS), baseada em suavização local [Cleveland et al. 1990].

A decomposição pode ser representada de forma aditiva, quando a amplitude das variações permanece aproximadamente constante ao longo do tempo:

$$Z_t = T_t + S_t + E_t$$

ou de forma multiplicativa, quando a amplitude das flutuações varia proporcionalmente ao nível da série:

$$Z_t = T_t \times S_t \times E_t$$

Para o tratamento de lacunas temporais decorrentes de falhas de medição, adotou-se a técnica de Interpolação Linear, que estima valores ausentes assegurando a continuidade da série temporal. A interpolação é definida por:

$$y = y_1 + \frac{(x - x_1)}{(x_2 - x_1)} \cdot (y_2 - y_1) \quad (1)$$

Esse procedimento reduz os impactos da descontinuidade dos dados sobre o desempenho dos modelos preditivos.

Por fim, com o objetivo de adequar diferentes escalas e unidades à sensibilidade das redes neurais, aplicou-se a normalização Min-Max, que transforma os dados para um intervalo padronizado entre 0 e 1. Como resultado, obtém-se um conjunto de dados contínuo, normalizado e com ruídos atenuados, proporcionando condições adequadas para o treinamento eficiente dos modelos LSTM e GRU.

### 3.3. Treinamento dos Modelos

Nesta etapa, os dados pré-processados são utilizados como entrada para o treinamento de modelos de redes neurais profundas, como LSTM e GRU. Esses modelos são amplamente utilizados para tarefas de predição em séries temporais, devido à sua capacidade de capturar padrões temporais complexos e dependências de longo prazo no comportamento da Vazão [Yamak et al. 2019].

#### 3.3.1. Processo de Treinamento

Os dados pré-processados, organizados em sequências temporais, são utilizados como entrada para os modelos LSTM e GRU. Cada sequência contém um número fixo de observações passadas, utilizadas para prever o valor de Vazão no próximo intervalo. O treinamento é realizado de forma iterativa, ajustando os pesos dos modelos para minimizar o erro entre os valores previstos e os valores reais.

A qualidade das predições geradas pelos modelos LSTM e GRU é avaliada utilizando a métrica RMSE, que é mais apropriado para representar o desempenho de modelos [Chai and Draxler 2014]. Um valor de RMSE mais baixo indica previsões mais precisas e, conseqüentemente, um melhor desempenho do modelo.

Ao final, os modelos são ajustados para capturar padrões temporais complexos, selecionando-se a configuração com o menor RMSE para garantir a precisão em relação aos valores reais. A documentação dos resultados fundamenta futuras comparações e ajustes paramétricos, assegurando que o sistema forneça informações essenciais para a otimização contínua e o gerenciamento proativo da rede.

### 3.4. Avaliação

Após o treinamento, para realizar uma avaliação abrangente, calculamos o RMSE [Brito et al. 2026] para cada caso, comparando os valores estimados com os valores reais. Além disso, analisamos a similaridade entre os dados estatísticos das séries temporais resultantes dos comportamentos do TCP. Para isso, DTW e a Correlação de Pearson como principais técnicas de comparação.

O DTW é uma métrica baseada em distância que alinha duas séries temporais, mesmo que tenham comprimentos ou fases diferentes. Valores elevados de distância DTW indicam diferenças significativas entre as séries em termos de tendência, sazonalidade e resíduos. Já a Correlação de Pearson mede a relação linear entre duas séries temporais, atribuindo um valor que varia de -1 a 1. Enquanto 1 representa uma correlação linear positiva perfeita, -1 indica uma correlação linear negativa perfeita, e 0 aponta para a ausência de relação linear.

Ademais, é comum que empresas e provedores de internet definam faixas de desempenho para classificar medições, permitindo suavizar variações e estabelecer limites com base nos requisitos dos serviços que utilizam a rede [Bayne et al. 2017]. Seguindo a abordagem do modelo MonIPÊ da RNP, foram adotadas cinco categorias de Vazão de rede: *Red*, abaixo de 0,2 Gbit/s; *Orange*, entre 0,2 e 0,5 Gbit/s; *Yellow*, entre 0,5 e 0,8 Gbit/s; *Blue*, entre 0,8 e 1,0 Gbit/s; e *Green*, 1,0 Gbit/s ou mais.

Para avaliar a precisão, consideramos que uma predição está correta se o valor estimado e o valor real pertencem à mesma categoria. A acurácia então é calculada com base na comparação entre as categorias de valores predefinidas. Primeiro, os valores reais e previstos são mapeados para categorias específicas (como "r", "o", "y", "b" e "g"), de acordo com a definição dos intervalos. Em seguida, é verificado quantas vezes a categoria prevista corresponde à categoria real para cada registro. A acurácia é calculada como a proporção de acertos em relação ao total de registros, expressa como uma porcentagem. Isso facilita a avaliação do modelo, pois considera as predições dentro de intervalos de valor, em vez de comparar os números exatos.

#### 4. Resultados

Conforme ilustrado na Figura 2, os modelos LSTM e GRU apresentam desempenhos semelhantes na predição da Vazão, com valores de RMSE próximos na maioria dos pontos analisados. Esse resultado indica que ambas as arquiteturas são adequadas para a modelagem da série temporal. Observam-se, contudo, variações pontuais em que um modelo supera ligeiramente o outro, possivelmente em função das características locais dos dados, da presença de padrões sazonais e de diferenças na capacidade de retenção de dependências de longo prazo. Assim, embora ambos os modelos sejam viáveis, análises específicas por ponto de medição e ajustes de hiperparâmetros podem favorecer uma arquitetura em determinados contextos.

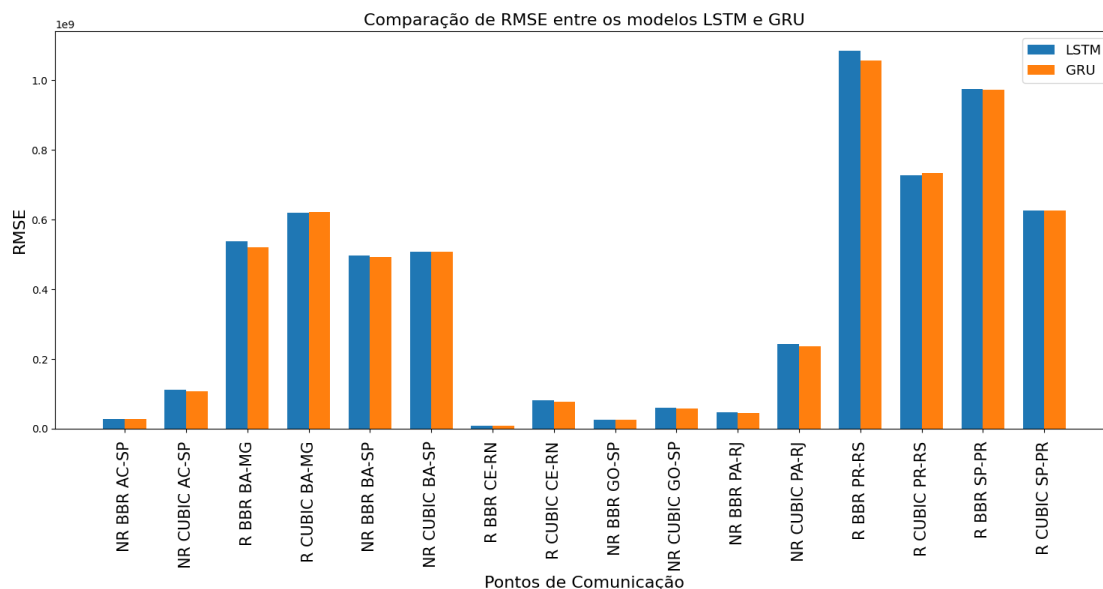


Figura 2. Resultados do RMSE

A análise da Figura 3 revela que, na maioria dos casos, os modelos LSTM e GRU apresentam desempenhos semelhantes, com acurácia superior a 85%, atingindo 100% em séries de alta previsibilidade. Contudo, discrepâncias pontuais são evidentes: em cenários como "NR CUBIC PA-RJ 2023", o GRU supera o LSTM, sugerindo que sua arquitetura de duas portas (atualização e redefinição) pode ser mais eficiente computacionalmente para modelar certas características do que a estrutura de três portas do LSTM. Tais variações são atribuídas à complexidade das séries temporais e à presença de ruídos nos dados.

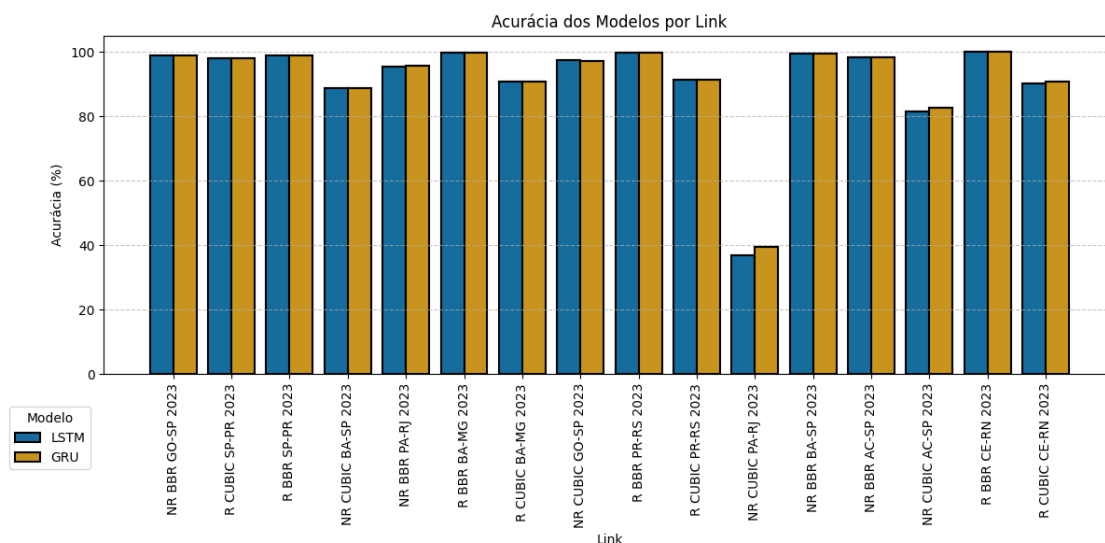


Figura 3. Resultados Acurácia.

A análise individual dos resultados destaca o impacto dos protocolos TCP no RMSE. Em pontos como *PR-RS* (Figura 4), o protocolo TCP BBR gera erros de predição significativamente maiores que o CUBIC, sugerindo que o BBR introduz maior variabilidade no comportamento do throughput.

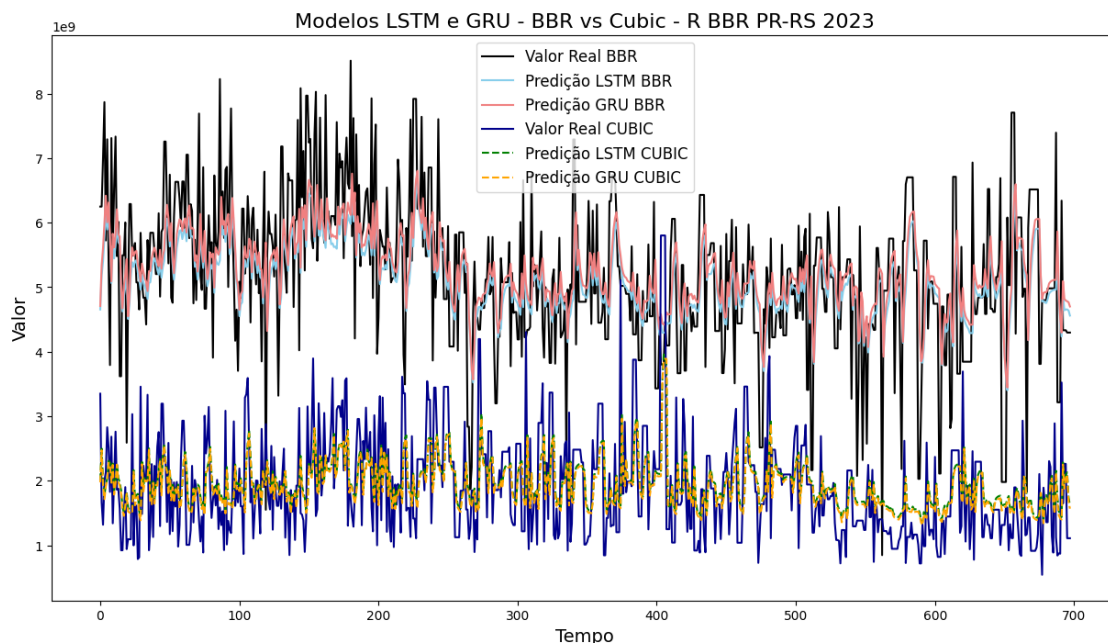
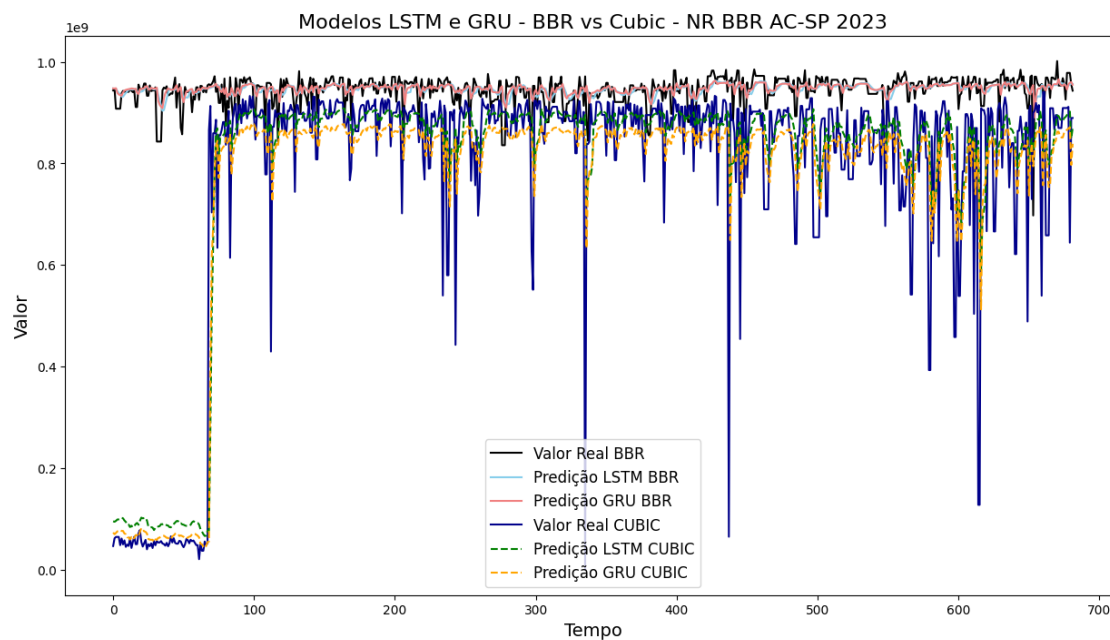


Figura 4. Exemplo: Ponto de Comunicação PR-RS.

Em contraste, nos pontos *AC-SP* (Figura 5), os valores de RMSE são baixos, especialmente com o TCP BBR. Isso indica estabilidade e previsibilidade na Vazão, facilitando o desempenho de ambos os modelos.



**Figura 5. Ponto de Comunicação AC-SP.**

## 5. Conclusão

Com isso, vimos que a predição de desempenho confirmou-se como uma ferramenta essencial para a gestão resiliente de infraestrutura de redes. Motivada pela carência de análises em cenários reais e dinâmicos, esta dissertação propôs um modelo de predição de Vazão que considera explicitamente o impacto dos algoritmos de controle de congestionamento TCP. O estudo preencheu uma lacuna importante na literatura ao detalhar como diferentes estratégias de controle influenciam o comportamento estatístico da rede e, conseqüentemente, a precisão dos modelos preditivos.

Os resultados evidenciaram que a escolha do algoritmo afeta significativamente a alocação de recursos e a acurácia das predições. Essa constatação oferece subsídios vitais para que administradores evitem gargalos e otimizem a infraestrutura, validando a viabilidade de modelos que incorporam características específicas do TCP. O impacto prático estende-se a redes acadêmicas e comerciais, fornecendo ferramentas para um gerenciamento mais proativo frente ao crescimento constante da demanda.

Por fim, o trabalho estabelece uma base sólida para investigações futuras, abrindo caminho para a análise de outros algoritmos TCP e a exploração de novas abordagens de aprendizado de máquina. Diante da contínua evolução das redes de comunicação, os insights aqui apresentados contribuem para o desenvolvimento de soluções cada vez mais precisas e adaptáveis, fundamentais para a garantia da qualidade de serviço e eficiência operacional.

## 6. Produção Científica e Tecnológica

Como resultado direto da produção científica gerada por este trabalho, diversos artigos foram publicados em conferências de renome, o que não apenas atesta a qualidade e relevância da pesquisa realizada, mas também contribui para a disseminação dos resultados e o compartilhamento de insights com a comunidade acadêmica em geral.

1. IEEE/IFIP Network Operations and Management Symposium (NOMS 2023 - Qualis: A2)<sup>1</sup>
2. Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2023 - Qualis: A4)<sup>2</sup>
3. IEEE Transactions on Network and Service Management (TNSM - Qualis: A2)<sup>3</sup>
4. Latin-American Symposium on Dependable Computing (LADC 2024 - Qualis: A4)<sup>4</sup>
5. Latin-American Symposium on Dependable Computing (LADC 2025 - Qualis: A4)<sup>5</sup>

Adicionalmente, como parte deste trabalho, vale mencionar dois registros de software, onde a ferramenta "T-For"(baseada na proposta da dissertação) conquistou o 2º Lugar na Competição de Projetos do Telecom Summit 2023<sup>6</sup>, evento promovido pelo Sindicato das Indústrias e Empresas de Instalação e Manutenção de Redes, Equipamentos e Sistema de Telecomunicações do Estado do Ceará (Sindimest Ceará), em parceria com a Federação das Indústrias do Estado do Ceará (FIEC).

1. IoT-ID: Identificação de Dispositivos IoT usando Dados de Rede Anonimizados. 2023. Patente: Programa de Computador, data de registro: 17/05/2023, Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial.
2. T-For: Um modelo de previsão de Vazão de Rede. 2024. Patente: Programa de Computador, data de registro: 04/02/2024, Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial.

## Referências

- Bayne, L., Schepis, D., and Purchase, S. (2017). A framework for understanding strategic network performance: Exploring efficiency and effectiveness at the network level. *Industrial Marketing Management*, 67:134–147.
- Brito, M. L. L., Ferreira, M. C. M., Portela, A. L. C., and Gomes, R. L. (2026). Ai-based estimation of bandwidth availability for data offloading in edge-cloud computing. *IEEE Networking Letters*, 8:69–73.
- Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73.
- Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*, pages 1–7.
- Gomes, R. L., Bittencourt, L. F., Madeira, E. R., Cerqueira, E., and Gerla, M. (2016). A combined energy-bandwidth approach to allocate resilient virtual software defined networks. *Journal of Network and Computer Applications*, 69:98–106.

<sup>1</sup><https://doi.org/10.1109/NOMS56928.2023.1015426>

<sup>2</sup><https://doi.org/10.5753/sbrc.2023.508>

<sup>3</sup><https://doi.org/10.1109/TNSM.2024.3349701>

<sup>4</sup><https://doi.org/10.1145/3697090.3699869>

<sup>5</sup>[https://doi.org/10.1007/978-3-032-11539-3\\_21](https://doi.org/10.1007/978-3-032-11539-3_21)

<sup>6</sup><https://www.uece.br/agin/noticias/laboratorio-larces-uece-vence-competicao-telecom-summit-2023/>

- Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. M. (2013). A virtual network allocation algorithm for reliability negotiation. In *2013 22nd International Conference on Computer Communication and Networks (ICCCN)*, pages 1–7.
- Kanaya, T., Tabata, N., and Yamaguchi, S. (2020). A study on performance of cubic tcp and tcp bbr in 5g environment. In *2020 IEEE 3rd 5G World Forum (5GWF)*, pages 508–513.
- Lara-Benítez, P., Carranza-García, M., and Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31(03):2130001. PMID: 33588711.
- Miyazawa, K., Yamaguchi, S., and Kobayashi, A. (2020). Mechanism of cyclic performance fluctuation of tcp bbr and cubic tcp communications. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1139–1144.
- Redhu, P., Kumar, K., et al. (2023). Short-term traffic flow prediction based on optimized deep learning neural network: Pso-bi-lstm. *Physica A: Statistical Mechanics and its Applications*, 625:129001.
- Ribeiro, S. E., Menezes, R. A., Portela, A. L., Araújo, T. P., and Gomes, R. L. (2023). Aplicando redes neurais e análise temporal para predição adaptativa de desempenho de rede. In *Anais do XLI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 490–503. SBC.
- Sepasgozar, S. S. and Pierre, S. (2022). Network traffic prediction model considering road traffic parameters using artificial intelligence methods in vanet. *IEEE Access*, 10:8227–8242.
- Silveira, M. M., Portela, A. L., Menezes, R. A., Souza, M. S., Silva, D. S., Mesquita, M. C., and Gomes, R. L. (2023). Data protection based on searchable encryption and anonymization techniques. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–5. IEEE.
- Sone, S. P., Lehtomäki, J. J., and Khan, Z. (2020). Wireless traffic usage forecasting using real enterprise network data: Analysis and methods. *IEEE Open Journal of the Communications Society*, 1:777–797.
- Wang, Y., Ke, S., An, C., Lu, Z., and Xia, J. (2024). A hybrid framework combining lstm nn and bnn for short-term traffic flow prediction and uncertainty quantification. *KSCE Journal of Civil Engineering*, 28(1):363–374.
- Yamak, P. T., Yujian, L., and Gadosey, P. K. (2019). A comparison between arima, lstm, and gru for time series forecasting. In *Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence*, pages 49–55.
- Yang, H., Li, X., Qiang, W., Zhao, Y., Zhang, W., and Tang, C. (2021). A network traffic forecasting method based on sa optimized arima–bp neural network. *Computer Networks*, 193:108102.