Classificação de Risco de Endereços IP Baseada em Inteligência de Ameaças e Aprendizado Ensemble

Francisco V. J. Nobre[1], **Rafael L. Gomes**[1] **(Orientador)**

[1] Universidade Estadual do Ceará (UECE)

valderlan.nobre@aluno.uece.br, rafa.lopes@uece.br

Abstract. *The accelerated dynamics of cyberattacks make static blocklists insufficient for effective network protection. This dissertation investigates Machine Learning models applied to IP address risk classification using Cyber Threat Intelligence (CTI) data. A dataset with 10,000 samples and 26 features was built from multiple CTI sources, including AbuseIPDB, VirusTotal, Pulsedive, and APIVoid. The research compares classical classifiers, neural networks, and ensemble methods under scenarios with and without SMOTE balancing. The results show that Stacking achieved the best predictive robustness, with Balanced Accuracy above 99.7% and superior probabilistic calibration, while Decision Tree reached the lowest inference time, 3.26 ms. The dissertation also proposes a hybrid edge-cloud architecture in which the selected classifier can support automated network response mechanisms.*

Resumo. *A dinâmica acelerada dos ataques cibernéticos torna as listas de bloqueio estáticas insuficientes para a proteção eficaz de redes. Esta dissertação investiga modelos de Aprendizado de Máquina aplicados à classificação de risco de endereços IP utilizando dados de Inteligência de Ameaças Cibernéticas. Para isso, foi construído um conjunto de dados com 10.000 amostras e 26 características extraídas de múltiplas fontes de CTI, incluindo AbuseIPDB, VirusTotal, Pulsedive e APIVoid. A pesquisa compara classificadores clássicos, redes neurais e métodos de ensemble em cenários com e sem balanceamento por SMOTE. Os resultados mostram que o Stacking obteve a maior robustez preditiva, com Acurácia Balanceada superior a 99,7% e melhor calibração probabilística, enquanto a Árvore de Decisão apresentou o menor tempo de inferência, 3,26 ms. A dissertação também propõe uma arquitetura híbrida borda-nuvem na qual o classificador selecionado pode apoiar mecanismos automatizados de resposta em redes.*

1. Introdução

O aumento da sofisticação dos ataques cibernéticos tem ampliado a necessidade de soluções de segurança proativas, inteligentes e adaptáveis. Em ambientes de rede, a identificação de endereços IP associados a atividades maliciosas é uma tarefa crítica, pois esses endereços podem estar relacionados a varreduras, tentativas de exploração, comunicação com infraestruturas de comando e controle, distribuição de malware e outras ações ofensivas. Tradicionalmente, a classificação de IPs maliciosos depende de listas estáticas de bloqueio, regras configuradas manualmente ou análises realizadas por especialistas, abordagens que apresentam limitações quanto à escalabilidade, ao tempo de resposta e à adaptação a novas ameaças [Wagner et al. 2019, Costa et al. 2024].

Nesse contexto, a Inteligência de Ameaças Cibernéticas, ou *Cyber Threat Intelligence* (CTI), fornece indicadores relevantes para apoiar a tomada de decisão em segurança. Bases públicas de reputação e análise de ameaças, como AbuseIPDB, VirusTotal, Pulsedive e APIVoid, disponibilizam informações sobre histórico de abuso, reputação, geolocalização, presença em listas de bloqueio e pontuações associadas a comportamento malicioso. Entretanto, essas fontes são heterogêneas e podem apresentar avaliações divergentes para um mesmo endereço IP, tornando necessário o uso de mecanismos capazes de consolidar evidências de forma mais robusta [Wagner et al. 2019].

Dentro desse contexto, a dissertação apresentada neste documento propôs uma abordagem para classificação de risco de endereços IP baseada em Aprendizado de Máquina e múltiplas fontes de CTI. A proposta inclui a formação de um conjunto de dados rotulado, uma estratégia de votação ponderada multicritério para classificação dos IPs em três níveis de risco e uma avaliação experimental de diferentes modelos supervisionados. Além disso, o trabalho apresenta uma arquitetura híbrida, organizada em camadas de borda e nuvem, para demonstrar como o mecanismo de classificação pode ser integrado a ambientes reais de segurança de redes.

1.1. Motivação

A motivação desta dissertação decorre da limitação de abordagens tradicionais de defesa, baseadas em listas estáticas de bloqueio e regras mantidas manualmente, para acompanhar a velocidade e a escala dos ataques cibernéticos atuais. Embora fontes de CTI forneçam indicadores relevantes sobre reputação, histórico de abuso e atividades suspeitas, essas bases são heterogêneas, adotam critérios próprios de pontuação e podem produzir avaliações divergentes para um mesmo endereço IP. Dessa forma, o trabalho foi motivado pela necessidade de combinar evidências provenientes de múltiplas fontes de CTI para classificar o risco de endereços IP de forma mais robusta, escalável e adequada à automação de respostas em redes, apoiando decisões operacionais como permitir, monitorar, restringir ou bloquear conexões [Costa et al. 2024, Wagner et al. 2019].

1.2. Objetivos

O objetivo geral da dissertação foi desenvolver e avaliar modelos de classificação de risco de endereços IP baseados em Aprendizado de Máquina, utilizando dados provenientes de múltiplas fontes de Inteligência de Ameaças Cibernéticas, com a finalidade de automatizar e aumentar a precisão da detecção de atividades maliciosas em redes de computadores.

Para atingir esse objetivo, o trabalho buscou construir um conjunto de dados rotulado e representativo de endereços IP, desenvolver e treinar modelos supervisionados para classificação multiclasse, avaliar comparativamente diferentes algoritmos por meio de métricas de desempenho e robustez, investigar o impacto do balanceamento por SMOTE e propor uma arquitetura capaz de integrar o classificador a mecanismos automatizados de resposta, como atualização dinâmica de regras de firewall.

1.3. Inovação e Contribuições

A principal originalidade da dissertação está na integração de três elementos que normalmente são tratados de forma separada: a fusão de sinais provenientes de múltiplas fontes de CTI, a classificação supervisionada de risco de endereços IP e a integração do resultado do modelo a uma arquitetura operacional de resposta em rede. Diferentemente de

abordagens baseadas apenas em listas estáticas ou em classificações binárias, a proposta adota uma classificação ternária, considerando os níveis *allowlist*, *suspicious* e *denylist*, permitindo associar diferentes ações de segurança a cada nível de risco.

Como contribuições científicas, este trabalho constrói um conjunto de dados (público em IEEEDataPort[1]) com 10.000 amostras e 26 características extraídas de fontes públicas de CTI, propõe uma estratégia de rotulação por votação ponderada multicritério, avalia diferentes paradigmas de Aprendizado de Máquina aplicados ao mesmo problema de cibersegurança e analisa a relação entre qualidade preditiva, calibração probabilística, generalização, balanceamento de classes e tempo de inferência. Os resultados indicam que o Stacking apresenta a melhor robustez preditiva, enquanto modelos mais simples, como Árvore de Decisão, oferecem menor latência para cenários de alta vazão.

2. Proposta

A Figura 1 apresenta uma visão geral da arquitetura proposta nesta dissertação. A solução é organizada como uma arquitetura híbrida e distribuída, composta por uma camada de borda, responsável pela interação direta com a infraestrutura de rede, e uma camada de nuvem, responsável pelo processamento intensivo, pela consulta a fontes externas de CTI e pela classificação de risco de endereços IP. Essa separação busca combinar escalabilidade e baixa latência: enquanto a nuvem executa tarefas de análise, enriquecimento e inferência, a borda atua próxima aos gateways e firewalls, permitindo que decisões de segurança sejam aplicadas de forma mais rápida no ambiente operacional [Costa et al. 2024].

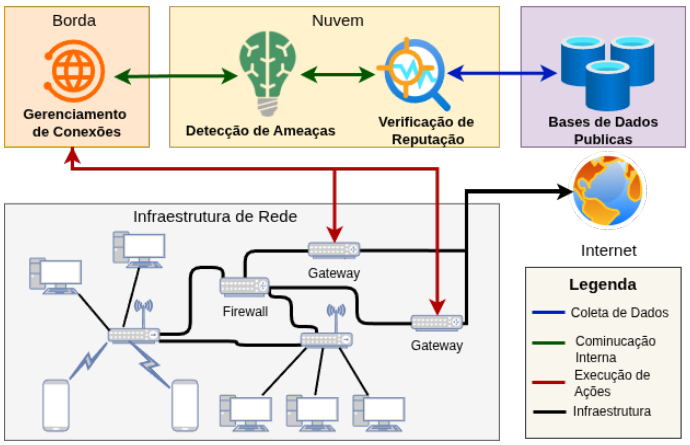

Figura 1. Visão geral da proposta.

A arquitetura é composta por três módulos principais. O primeiro é o módulo de *Verificação de Reputação*, localizado na camada de nuvem, responsável por consultar bases públicas de CTI, como AbuseIPDB, VirusTotal, Pulsedive e APIVoid. Esse módulo coleta informações sobre reputação, presença em listas de bloqueio, geolocalização, relatos de abuso e pontuações de comportamento malicioso. As informações obtidas são utilizadas para enriquecer os dados associados a cada endereço IP e alimentar o processo de classificação.

[1]https://dx.doi.org/10.21227/ys2m-9d61

O segundo módulo é o de *Detecção de Ameaças*, também localizado na nuvem. Esse componente utiliza modelos de Aprendizado de Máquina para classificar os endereços IP em três níveis de risco: *allowlist*, *suspicious* e *denylist*. O processo envolve a consolidação dos dados coletados, a extração de características e a inferência do modelo treinado. Dessa forma, a proposta não se limita a verificar se um endereço IP aparece em uma lista de bloqueio, mas combina diferentes evidências para produzir uma decisão de risco mais robusta.

O terceiro módulo é o de *Gerenciamento de Conexões*, localizado na camada de borda. Esse componente recebe as decisões produzidas na nuvem e as transforma em ações operacionais na infraestrutura de rede, como atualização de regras de firewall, bloqueio de endereços classificados como maliciosos, monitoramento de endereços suspeitos e aplicação de políticas intermediárias, como degradação temporária da conexão por meio de *tarpit*. Por estar próximo aos dispositivos de rede, esse módulo reduz o tempo necessário para aplicar decisões de segurança em gateways, firewalls e demais componentes da infraestrutura.

O fluxo da proposta inicia-se com a coleta de metadados de conexão na infraestrutura de rede. Esses dados são encaminhados para a camada de nuvem, onde ocorre a consulta às bases de CTI e o enriquecimento das informações relacionadas ao endereço IP analisado. Em seguida, o módulo de detecção de ameaças utiliza o modelo de Aprendizado de Máquina para classificar o risco do IP. A decisão produzida é então enviada à camada de borda, onde o módulo de gerenciamento de conexões executa a ação correspondente, como permitir, monitorar, degradar ou bloquear a conexão.

Por fim, a proposta contempla um ciclo de reavaliação e aprendizado contínuo, no qual novos dados coletados com alto grau de confiança podem ser reincorporados ao conjunto de treinamento. Esse mecanismo busca mitigar a degradação de desempenho causada pela volatilidade das ameaças cibernéticas. Assim, a arquitetura proposta conecta a análise baseada em CTI e Aprendizado de Máquina à aplicação prática de políticas de segurança em infraestruturas reais de rede.

3. Resultados

3.1. Configuração Experimental

A avaliação experimental foi conduzida com o objetivo de analisar a capacidade dos modelos de Aprendizado de Máquina de classificar endereços IP em três níveis de risco. Para isso, foi utilizado um conjunto de dados construído a partir das quatro fontes de CTI já citadas. Os modelos avaliados incluem métodos baseados em árvores (Decision Tree (DT), Random Forest (RF) e Extra Trees (ET)), classificadores por distância (K-Nearest Neighbors (KNN)), métodos de margem (Support Vector Machine (SVM)), redes neurais (Neural Network (NN) e Convolutional Neural Network (CNN)) e técnicas de ensemble (Voting, Stacking e AdaBoost).

Os experimentos foram conduzidos em dois cenários: sem balanceamento e com balanceamento por SMOTE. A avaliação considerou métricas de desempenho e robustez, incluindo Acurácia, Acurácia Balanceada, F1-Macro, Recall, Matthews Correlation Coefficient (MCC, métrica que avalia a qualidade da classificação considerando simultaneamente verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos),

Log Loss e Overfit Gap (diferença entre o desempenho obtido no treinamento e no teste, utilizada para indicar possível sobreajuste do modelo). Além disso, foi realizado um benchmark de tempo de inferência para avaliar a viabilidade de implantação dos modelos em ambientes reais de rede [Géron 2017].

3.2. Resultados de Classificação

A Tabela 1 apresenta os resultados obtidos no cenário sem SMOTE. Observa-se que os métodos de ensemble apresentaram os melhores desempenhos globais, com destaque para o Stacking. Esse modelo alcançou Acurácia de 0,9970, Acurácia Balanceada de 0,9964, F1-Macro de 0,9958 e Log Loss de 0,0092, indicando elevada capacidade de classificação e melhor calibração probabilística entre os modelos avaliados.

Tabela 1. Desempenho dos Modelos

Modelo	Acc	Acc. Bal.	F1-Macro	Recall	MCC	Log Loss	Overfit Gap
Stacking	**0,9970**	**0,9964**	**0,9958**	**0,9970**	**0,9952**	**0,0092**	0,0030
Voting	0,9960	0,9938	0,9947	0,9960	0,9936	0,0274	0,0034
AdaBoost	0,9925	0,9934	0,9899	0,9925	0,9881	0,9265	0,0018
RF	0,9660	0,9728	0,9553	0,9660	0,9477	0,0798	0,0046
CNN	0,9565	0,9639	0,9435	0,9565	0,9336	0,0842	0,0063
DT	0,9410	0,9519	0,9252	0,9410	0,9117	0,1489	0,0065
ET	0,9355	0,9481	0,9190	0,9355	0,9041	0,2040	0,0093
KNN	0,9435	0,9236	0,9236	0,9435	0,9096	0,2421	0,0083
NN	0,9400	0,9144	0,9182	0,9400	0,9046	0,2261	**0,0017**
SVM	0,9145	0,9039	0,8902	0,9145	0,8663	0,1979	0,0029

No cenário com balanceamento por SMOTE, apresentado na Tabela 2, o Stacking também obteve o melhor desempenho geral, alcançando Acurácia de 0,9980, Acurácia Balanceada de 0,9978, F1-Macro de 0,9975 e MCC de 0,9968. Esses resultados demonstram que o modelo manteve elevada robustez mesmo após a aplicação de amostras sintéticas para balanceamento das classes.

Tabela 2. Desempenho dos Modelos com SMOTE

Modelo	Acc	Acc. Bal.	F1-Macro	Recall	MCC	Log Loss	Overfit Gap
Stacking	**0,9980**	**0,9978**	**0,9975**	**0,9980**	**0,9968**	**0,0217**	0,0051
Voting	0,9950	0,9936	0,9930	0,9950	0,9920	0,0392	0,0048
CNN	0,9920	0,9876	0,9899	0,9920	0,9872	0,0353	0,0047
AdaBoost	0,9795	0,9835	0,9726	0,9795	0,9680	0,9298	0,0046
RF	0,9730	0,9784	0,9641	0,9730	0,9582	0,0706	0,0062
DT	0,9605	0,9647	0,9480	0,9605	0,9390	0,1162	0,0100
NN	0,9470	0,9515	0,9325	0,9470	0,9183	0,1258	**0,0018**
ET	0,9390	0,9492	0,9231	0,9390	0,9083	0,2070	0,0123
KNN	0,9345	0,9324	0,9153	0,9345	0,8985	0,2762	0,0084
SVM	0,9135	0,9019	0,8887	0,9135	0,8646	0,2085	0,0177

A comparação entre os dois cenários mostra que o SMOTE não atua como solução universal para classificação de ameaças. Embora tenha melhorado o desempenho de

alguns modelos, como a CNN, que passou de Acurácia Balanceada de 0,9639 para 0,9876, também introduziu efeitos negativos em modelos sensíveis. O AdaBoost apresentou valores elevados de Log Loss nos dois cenários, indicando baixa calibração probabilística, enquanto o Extra Trees apresentou aumento do Overfit Gap no cenário balanceado. Esses resultados reforçam que técnicas de balanceamento devem ser escolhidas de acordo com o comportamento de cada modelo e os requisitos operacionais do ambiente [Chawla et al. 2002].

O bom desempenho do Stacking está associado à sua capacidade de combinar as predições de diferentes classificadores de base por meio de um meta-modelo, explorando a complementaridade entre algoritmos. Essa característica é particularmente relevante em cenários de classificação de risco, nos quais diferentes modelos podem capturar padrões distintos nos dados de CTI [Wolpert 1992].

3.3. Tempo de Inferência

Além da qualidade preditiva, a viabilidade de implantação em redes reais depende do tempo de inferência. Em ambientes de alta vazão, a latência do classificador pode impactar diretamente o processamento de conexões e a tomada de decisão em tempo real. Por isso, a dissertação também avaliou o custo computacional dos modelos, conforme apresentado na Figura 2.

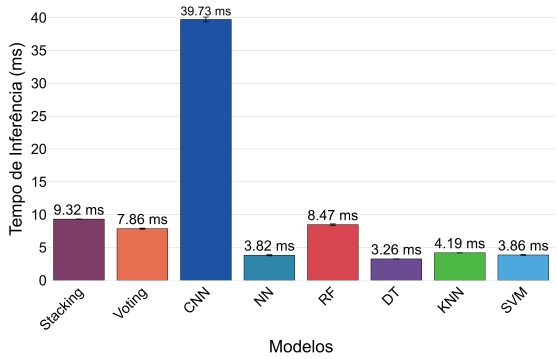

Figura 2. Comparativo de tempo de inferência.

Os resultados revelam um trade-off importante entre qualidade preditiva e eficiência operacional. DT apresentou o menor tempo de inferência, com 3,26 ms, consolidando-se como a alternativa mais eficiente para cenários nos quais a velocidade de decisão é o principal requisito. Em contraste, a CNN apresentou o maior custo computacional, com 39,73 ms por inferência.

Os modelos Voting e Stacking posicionaram-se em uma zona intermediária de equilíbrio. O Voting apresentou tempo de inferência de 7,86 ms, enquanto o Stacking apresentou 9,32 ms. Apesar de ser ligeiramente mais custoso, o Stacking obteve os melhores resultados preditivos e a melhor calibração probabilística, tornando-se mais adequado para decisões críticas de bloqueio. O Voting, por sua vez, mostrou-se uma alternativa competitiva quando se busca reduzir a complexidade sem perda expressiva de desempenho.

De forma geral, os resultados indicam que a classificação de risco de endereços IP deve ser tratada como um problema dependente de contexto. Para ambientes nos quais fal-

sos negativos são especialmente críticos, modelos mais robustos e bem calibrados, como Stacking, são mais indicados. Para ambientes nos quais a resposta precisa ocorrer com latência mínima, modelos simples e interpretáveis, como DT, podem ser mais apropriados.

4. Produção Científica e Tecnológica

Com relação à produção científica, têm-se os seguintes artigos diretamente associados à dissertação: [Nobre et al. 2026b, Nobre et al. 2026a, Nobre et al. 2025c, Nobre et al. 2026c, Nobre et al. 2025b, Nobre et al. 2025a]. Adicionalmente, tem-se a seguinte produção científica complementar desenvolvida durante o mestrado: [Santos et al. 2026, Trajano et al. 2026, Souza et al. 2026, Portela et al. 2024, Ferreira et al. 2024b, Silva et al. 2024b, Silva et al. 2024a, Ferreira et al. 2024a].

Produção Tecnológica (Registro de Software no INPI): **SIGMA-IP:** Sistema Inteligente de Gestão e Monitoramento de Ameaças para Redes IPs; **CTI-IPGuard:** Sistema de Classificação de Endereços IP baseado em Inteligência de Ameaças Cibernéticas.

Referências

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357.

Costa, M. A., Costa, Y. M., Almeida, Y. O., Cardoso, F. J., and Gomes, R. L. (2024). Connection management using automated firewall based on threat intelligence. In *Proceedings of the 2024 Latin America Networking Conference*, LANC '24, page 32–37, New York, NY, USA. Association for Computing Machinery.

Ferreira, M., Ribeiro, S., Nobre, F., Linhares, M., Araújo, T., and Gomes, R. (2024a). Predição de desempenho de rede resiliente a falhas de medição. In *Anais do XXIX Workshop de Gerência e Operação de Redes e Serviços*, pages 29–42, Porto Alegre, RS, Brasil. SBC.

Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024b). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*, pages 1–7.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.

Nobre, F., Araujo, R., Alves, D., Santos, J., Nobre, J., and Gomes, R. (2026a). Gerenciamento adaptativo de conexões e classificação de endereços ip na borda da rede usando proxies preditivos aiops. In *Anais do XXXI Workshop de Gerência e Operação de Redes e Serviços*, pages 127–140, Porto Alegre, RS, Brasil. SBC.

Nobre, F. J., da Costa, Y. M., Alves, D. O., Araujo, R. S., Rodrigues, L. S., Neto, A. B., Jr., L. A. P., and Gomes, R. L. (2025a). Sigma-ip: Sistema inteligente de gestão e monitoramento de ameaças para redes ips. In *Anais do XXX Workshop de Gerência e Operação de Redes e Serviços*, pages 43–56, Porto Alegre, RS, Brasil. SBC.

Nobre, F. V., Urbano, A. C., Araújo, R. D. S., Melo, Y. C., and Gomes, R. L. (2026b). Detection of malicious ip based on artificial intelligence and threat intelligence. *International Journal of Security and Networks*. Forthcoming.

Nobre, F. V. J., Alves, D. O., Araujo, R. S., Campos, G. A., and Gomes, R. L. (2026c). Risk classification of ip addresses using machine learning with weighted voting approach. In Rodrigues, L. A. and Oliveira, R., editors, *Dependable and Secure Computing*, pages 320–328, Cham. Springer Nature Switzerland.

Nobre, F. V. J., Araujo, R. S., Alves, D. O., Nascimento, E. S., Rodrigues, L. S., and Gomes, R. L. (2025b). Network connection management based on the integration of artificial intelligence for it operations and threat intelligence databases. In *2025 23rd International Symposium on Network Computing and Applications (NCA)*, pages 95–102.

Nobre, F. V. J., Silva, D. d. S., Ferreira, M. C. M. M., Brito, M. L. M. L., de Araújo, T. P., and Gomes, R. L. (2025c). Time-weighted correlation approach to identify high delay links in internet service providers. *Journal of Internet Services and Applications*, 16(1):419–430.

Portela, A., Linhares, M. M., Nobre, F. V. J., Menezes, R., Mesquita, M., and Gomes, R. L. (2024). The role of tcp congestion control in the throughput forecasting. In *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*, LADC '24, page 196–199, New York, NY, USA. Association for Computing Machinery.

Santos, J., Nobre, F., Castro, I., Linhares, M., Ferreira, M., and Gomes, R. (2026). Predição de atraso alto em infraestruturas de rede através de redes neurais de grafos. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1150–1163, Porto Alegre, RS, Brasil. SBC.

Silva, D., Nobre, F., Ferreira, M., Araújo, T., and Gomes, R. (2024a). Correlacionando dados de monitoramento de rede para identificação de causas de problemas de desempenho. In *Anais do XXIX Workshop de Gerência e Operação de Redes e Serviços*, pages 15–28, Porto Alegre, RS, Brasil. SBC.

Silva, D., Nobre, F., Ferreira, M., Portela, A., Araújo, T., and Gomes, R. (2024b). Identificação das causas de situações de alto atraso em provedores de internet. In *Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 111–120, Porto Alegre, RS, Brasil. SBC.

Souza, P., Santos, J., Neto, A., Nobre, F., Trajano, A., and Gomes, R. (2026). Detecção de ataques de phishing por meio de modelos de linguagem. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 589–602, Porto Alegre, RS, Brasil. SBC.

Trajano, A., Costa, C., Nobre, F., and Gomes, R. (2026). Melhorando a eficiência energética na execução de llms via controle de potência de gpus orientado por slas. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 968–981, Porto Alegre, RS, Brasil. SBC.

Wagner, T. D., Mahbub, K., Palomar, E., and Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers Security*, 87:101589.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.