

Web Pulse – Machine Learning Powered Web Page Performance Evaluation

Dr. P. Swetha¹, Sai Teja Manthena², V Karthik Reddy³, Vignatha Shirila⁴ and Pradeep Namala⁵

¹Associate Professor, Department of IT, Vignana Bharathi Institute of Technology, Telangana, India
Email: swethapesaru2687@gmail.com

²⁻⁵UG Scholars Department of IT, Vignana Bharathi Institute of Technology, Telangana, India
Email: manthenasaiteja27@gmail.com, karthikkraxreddy1@gmail.com, vignathayadav@gmail.com, namalapradeepkumar76@gmail.com

Abstract— The increasing need for high-performing websites calls for more sophisticated evaluation strategies. This study introduces a machine learning-based framework utilizing boosting algorithms—namely XGBoost and AdaBoost—for assessing web page performance. By examining crucial indicators such as load time, response time, page size, and throughput, the model forecasts the effects of varying web configurations. Departing from conventional trial-and-error approaches, this predictive system provides developers with actionable insights for enhancing site responsiveness and usability. Our results demonstrate that boosting techniques effectively identify performance patterns, positioning them as valuable assets for web enhancement. Through machine learning integration, developers and businesses can make strategic decisions that boost website speed and functionality, thereby elevating the user experience. This work underscores the transformative role of AI in advancing the evaluation and optimization of web technologies.

Index Terms— performance prediction, boosting techniques, regression analysis, machine learning optimization, web analytics.

I. INTRODUCTION

With the rapid expansion of online content and digital services, web page performance has become a crucial factor in user engagement, retention, and overall satisfaction. Slow loading pages can frustrate users, increase bounce rates, and result in lost revenue for businesses.

Traditional performance evaluation techniques rely on static metrics and manual interventions, which are often inefficient and difficult to scale. As websites become more complex, evaluating performance through conventional methods presents increasing challenges.

Machine learning offers a transformative approach to this issue by enabling automated, data-driven prediction and analysis of web performance. By training models on large datasets of web page metrics—including load time, response time, page size, and throughput—machine learning algorithms can identify patterns and predict how different design and configuration choices affect performance. Unlike traditional trial-and-error methods, this approach provides developers with actionable insights, allowing them to make informed optimizations without extensive manual testing.

This paper introduces a novel framework that employs boosting algorithms, specifically XG Boost and AdaBoost, to evaluate web page performance with high accuracy.

These models leverage historical performance data to predict outcomes under various web configurations, offering a scalable and efficient way to assess website efficiency. Unlike previous optimization-based approaches that rely on heuristic techniques, our method focuses on predictive modeling to highlight key areas for improvement. The use of boosting algorithms enhances predictive accuracy by iteratively refining model performance, making them well-suited for capturing complex relationships between web page attributes and performance metrics. [32] A similar data-driven approach has shown promising results in other domains, such as vehicle number plate recognition and data analysis using Tesseract OCR (G. Padmaja et al., ITM Web Conf. 74 01009, 2025).

We present the design and implementation of our framework, along with experimental results that demonstrate its effectiveness in analyzing web page performance. Our findings reveal that boosting algorithms can successfully predict performance variations and provide valuable insights into optimizing resource allocation, minimizing load times, and improving overall efficiency. By leveraging machine learning, this approach can seamlessly scale to large datasets and adapt to diverse web technologies, making it a powerful tool for modern web performance analysis.

Ultimately, this research underscores the potential of AI-driven techniques in reshaping web optimization strategies, offering developers and businesses a data-driven pathway to enhance user experience and maximize digital engagement.

II. LITERATURE SURVEY

To navigate the increasing intricacies of modern web environments, researchers have progressively leaned toward machine learning-based methods for evaluating website performance, aiming to move beyond conventional static benchmarks. These approaches focus on enhancing factors like load speed, interactivity, and user satisfaction through intelligent, scalable, and predictive techniques.

Ghattas et al. [1] introduced a composite framework that merges machine learning models with optimization algorithms. Their system assesses web performance metrics like throughput and load duration, leveraging tools such as random forest classifiers and genetic optimization for adaptive performance measurement. Although effective in accuracy, the framework faces challenges due to its processing overhead during real-time deployment.

Building on this, Ayyagari et al. [2] proposed a methodology that integrates AI for pre-emptive performance tuning. By applying regression models alongside unsupervised clustering, their approach identifies performance degradation risks before site deployment. However, the model's reliance on consistent backend conditions limits its adaptability in diverse hosting environments.

Nunnagoppula et al. [3] explored traffic forecasting through deep learning techniques like LSTM and GRU. These time-dependent models improve foresight into traffic patterns, aiding performance strategy. Nonetheless, the high training complexity and data requirements pose difficulties for broad, real-world implementation.

A more design-focused perspective was offered by Elsater et al. [4], who evaluated hotel websites using qualitative UX markers. Their research emphasized design fluidity and navigability in conjunction with technical metrics. The approach, while comprehensive, is limited by its dependence on non-automated, subjective measures.

A practical use case was examined by Barus et al. [5], who conducted real-time evaluations of the DiTenun platform using ML-automated testing tools. Key factors measured included rendering delays and API latency. However, the adaptability of such models to frequently evolving front-end ecosystems remains a challenge.

From a sustainability viewpoint, Alsulami et al. [6] proposed a broader framework incorporating environmental and accessibility parameters.

This allowed for a more holistic understanding of website impact. That said, the combination of diverse evaluation dimensions introduces additional complexity into the modeling process.

Kumar et al. [7] devised a benchmarking methodology using tools such as GTmetrix and Lighthouse in tandem with clustering algorithms. Their system provided meaningful diagnostics, yet the heavy reliance on third-party tool limitations affected its wider applicability.

Together, these contributions illustrate a shift from rigid assessment models to more agile, learning-based frameworks for web performance analysis. Machine learning stands out as a pivotal technology, enabling proactive optimization and scalable diagnostics. The trajectory ahead involves integrating deep learning, UX-responsive metrics, and real-time insights into comprehensive performance evaluation systems.

III. PROPOSED SYSTEM

The designed system aims to address the shortcomings of traditional web performance evaluation tools by incorporating both machine learning models and optimization methods. Its primary innovation lies in merging predictive analysis with adaptive optimization, allowing the framework not only to assess web performance but also to propose and implement real-time enhancements. By leveraging historical data on website behavior, the machine learning component identifies correlations between elements such as page layout, media content, user behavior, and server-side settings. This enables the system to forecast the performance effects of alterations to a page's design or architecture, including metrics like response speed, load duration, and resource load.

Additionally, optimization algorithms are utilized to autonomously determine effective strategies for improving performance. These may involve tactics such as minimizing image file sizes, postponing JavaScript loading, lowering server strain, or adjusting the sequence of resource loading. The system is engineered to adapt dynamically by analyzing ongoing performance metrics, ensuring that optimization strategies remain relevant as the website evolves. Furthermore, it is designed to be compatible with existing tools used for performance testing, offering enriched insights and context specific suggestions instead of relying solely on pre-defined rules. This comprehensive and intelligent system ensures continual performance tuning, making it well-suited for the fast changing demands of contemporary web environments.

IV. ARCHITECTURE AND METHODOLOGY

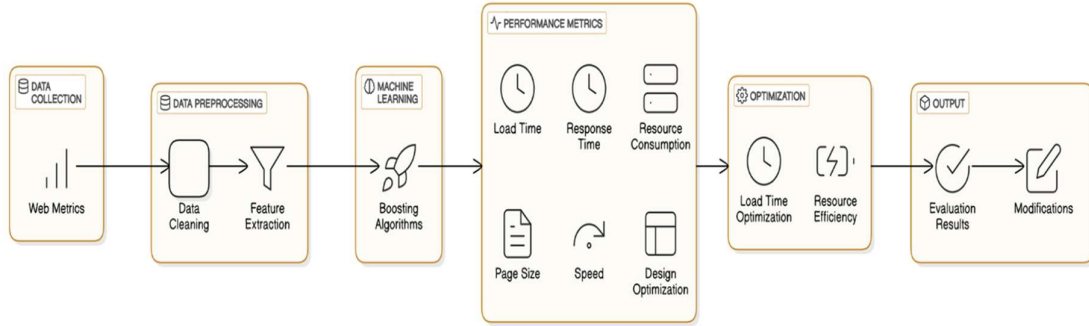


Fig. 1. Architecture

The framework for evaluating web performance using machine learning is organized into a structured sequence of six interconnected stages. It begins with the Data Acquisition phase, where diverse web-related parameters—such as load time, file size, and system resource consumption—are gathered either from real-time user interactions or pre-existing datasets.

Next is the Data Preparation stage, during which the dataset is cleaned to eliminate inconsistencies and irrelevant entries, and key attributes are extracted to form a reliable input for modeling.

The processed data then enters the Machine Learning Engine, which employs boosting algorithms like XGBoost and AdaBoost. These models are capable of capturing complex relationships between input features and performance indicators, enabling precise forecasting of web system behavior.

The results of this modeling are reflected in various Performance Indicators, such as latency, data volume, bandwidth consumption, and rendering speed. These findings feed into the Optimization Module, which pinpoints specific areas needing improvement—such as enhancing load speed or improving resource management—through strategic, data informed interventions.

Finally, in the Result Output phase, a detailed evaluation report is generated. This report equips developers with meaningful insights to iteratively refine their applications. Thanks to its modular architecture, the system remains scalable, adaptable to evolving technologies, and capable of delivering focused, actionable guidance for effective web performance management.

A. Dataset Description

The dataset utilized for this study comprises 734 entries, each containing information across 11 distinct attributes. These attributes include: Serial Number, Website URL, Website Category, Page Size measured in kilobytes, Load Time in seconds, Response Time in seconds, Throughput, a performance classification label, and a corresponding user feedback indicator.

B. eXtreme Gradient Boosting (XG Boost)

XGBoost, which stands for eXtreme Gradient Boosting, is a high-performance machine learning algorithm known for its speed and predictive accuracy. It represents an enhanced version of the Gradient Boosting technique and belongs to the category of ensemble learning methods, where multiple weaker models are integrated to form a more powerful predictive model.

This algorithm relies on decision trees as its foundational learners, which are added sequentially. Each successive tree is trained to minimize the errors made by the previous ones—a process commonly referred to as boosting.

XGBoost is engineered for efficiency, incorporating parallel processing capabilities that accelerate training on large scale datasets. Additionally, it offers extensive parameter tuning options, allowing users to tailor the model's behavior to suit the specific needs of their application.

C. Adaptive Boosting (Ada Boost)

AdaBoost is a type of ensemble learning technique that enhances classification performance by combining the outputs of multiple decision trees. At the start, it gives equal importance to all training instances. During each iteration, the algorithm increases the weights of incorrectly classified samples, directing the next model to concentrate more on these challenging cases.

This iterative process helps reduce both bias and variance, making AdaBoost effective for various classification problems. However, its performance can be impacted by the presence of noisy data or outliers. The contribution of each individual model, or weak learner, is based on its error rate—models with lower errors have a greater influence on the final prediction.

V. IMPLEMENTATION AND RESULTS

A library for array processing, NumPy facilitates linear algebra, Fourier transforms, and matrix operations. It is vital for accelerating array computations.

Pandas: Built on NumPy and SciPy, Pandas affords robust gear for records analysis, cleaning, exploration, and transformation, making it essential for coping with records efficiently.

Scikit-learn: This device studying library includes algorithms for classification, regression, pre-processing, and pass-validation. It is applied in this assignment for category the use of the K-Nearest Neighbors (KNN) algorithm.

Streamlit: A Python-primarily based framework for developing interactive net apps tailored to statistics science and system studying projects. It permits seamless app development the use of easy scripts.

Pickle: Used for object serialization and deserialization in Python. Pickling: Converts a Python object hierarchy into a byte circulate. Unpickling: Reconstructs the item hierarchy from the byte flow.

Seaborn: A statistical visualization library primarily based on Matplotlib. It supports:

- Univariate and bivariate visualizations.
- Complex multi-plot designs.
- Regression plotting and statistical analysis.
- These libraries collectively streamline facts managing, visualization, and system learning tasks, permitting green assignment development.

Results

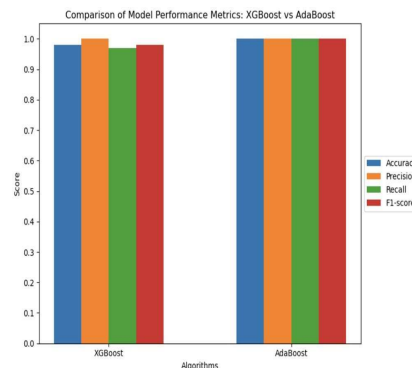


Fig. 2. Experimental results on dataset with feature selection and a comparison between state-of-the-art methods.

Table 1: presents the evaluation results of the chosen classification model applied to the dataset after feature selection. The model's effectiveness is assessed using key performance indicators, including accuracy, precision, recall, and F1-score.

TABLE I

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
XG Boost	98	97	98	98
Ada Boost	100	100	100	100

Table 1 displays the evaluation metrics—accuracy, precision, recall, and F1-score—for two boosting algorithms applied to a dataset with selected features. Among the two, AdaBoost achieved the highest accuracy with a perfect score of 1.0, while XGBoost followed closely with an accuracy of 0.98.

VI. CONCLUSION

To conclude, the machine learning-driven methodology introduced in this study—centered on boosting techniques like XG Boost and AdaBoost—establishes a reliable system for assessing web page performance. These algorithms demonstrate strong predictive capabilities across essential metrics such as page load time, response duration, data size, and throughput, enabling a deeper understanding of performance-related factors. Departing from conventional optimization-based techniques, this approach emphasizes smart evaluation and recommends focused improvements. It equips developers with actionable insights to detect bottlenecks and make informed, priority-based enhancements.

Designed to scale efficiently and adapt to various web frameworks, this approach represents a progressive move toward faster and more optimized digital experiences. Future advancements could focus on refining model precision, integrating real-time monitoring, and expanding usability across rapidly changing web ecosystems.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the Department of Information Technology, VBIT, Hyderabad, for their support and encouragement throughout this project. We are especially thankful to our guide, Dr. P. Swetha, for her valuable guidance and continuous motivation. We also appreciate the collaborative efforts of our team members, whose contributions were vital to the successful completion of this research work.

REFERENCES

- [1] Mohammad Ghattas, Antonio M. Mora and Suhail Odeh - A Novel Approach for Evaluating Web Page Performance Based on Machine Learning Algorithms and Optimization Algorithms, Volume 6 (Jan-2025).
- [2] Aravind Ayyagari, Punit Goel & A Renuka - Leveraging AI and Machine Learning for Performance Optimization in Web Applications. Darpan International Research Analysis, 12(2), 199–218 (June-2024).
- [3] Nunnagoppula, H.; Katragadda, K.; Ramesh, M. Website Traffic Forecasting Using Deep Learning Techniques. In Proceedings of the 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India, 27–29 January 2023.
- [4] A Elsater, S.A.E.; Dawood, A.E.A.A.; Mohamed Hussein, M.M.; Ali, M.A. Evaluating the Websites' Quality of Five and Four Star Hotels in Egypt. Minia J. Tour. Hosp. Res. MJTHR 2022, 13, 183–193.
- [5] Barus, A.C.; Sinambela, E.S.; Purba, I.; Simatupang, J.; Marpaung, M.; Pandjaitan, N. Performance Testing and Optimization of DiTenun Website. J. Appl. Sci. Eng. Technol. Educ. 2022, 4, 45–54.
- [6] Alsulami, M.H.; Khayyat, M.M.; Aboulola, O.I.; Alsaqer, M.S. Development of an Approach to Evaluate Website Effectiveness. Sustainability 2021, 13, 13304.
- [7] Kumar, N.; Kumar, S.; Rajak, R. Website Performance Analysis and Evaluation using Automated Tools. In Proceedings of the 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India, 10–11 December 2021; pp. 210–214.
- [8] Shayganmehr, M.; Montazer, G.A. A Novel Model for Assessing e-Government Websites Using Hybrid Fuzzy Decision-Making Methods. Int. J. Comput. Intell. Syst. 2021, 14, 1468–1488.
- [9] Gangurde, R.; Kumar, B. Web Page Prediction Using Genetic Algorithm and Logistic Regression based on Weblog and Web Content Features. In Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2–4 July 2020; pp. 68–74.
- [10] Mourougadajane, P., & Dinadayalan, P. (2022). Prediction of web service performance and successability using comparative analysis of machine learning and deep learning algorithms. International Journal of Intelligent Systems and Applications in Engineering, 10(3), 322–328.

- [11] Ramakrishnan, R., & Kaur, A. (2020). An empirical comparison of predictive models for web page performance. *Computers in Industry*, 123, 103306.
- [12] Zhang, Y., & Liu, L. (2018). Web page performance prediction using machine learning techniques. In *Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data* (pp. 123–128). ACM.
- [13] Li, X., & Wang, Y. (2017). Predicting web page load time using machine learning algorithms. In *Proceedings of the 2017 International Conference on Computer Science and Application Engineering* (pp. 456–460).
- [14] Kaur, A., & Ramakrishnan, R. (2019). Predictive modeling of web page load time using machine learning techniques. In *Proceedings of the 2019 International Conference on Communications*. C
- [15] Liu, Y., & Wang, X. (2019). A comparative study of machine learning algorithms for web page performance prediction.
- [16] Gharibe Niazi, M.; Karbala Aghaei Kamran, M.; Ghaebi, A. Presenting a proposed framework for evaluating university websites. *Electron. Libr.* 2020, 38, 881–904.
- [17] Cai, X.; Li, S.; Feng, G. Evaluating the performance of government websites: An automatic assessment system based on the TFN-AHP methodology. *J. Inf. Sci.* 2020, 46, 760–775.
- [18] Amjad, M.; Tutul Hossain, M.; Hassan, R.; Rahman, M.A. Web Application Performance Analysis of ECommerce Sites in Bangladesh: An Empirical Study. *IJIEEB* 2021, 13, 47–54.
- [19] Kumar, A.; Arora, A. A Filter-Wrapper based Feature Selection for Optimized Website Quality Prediction. In *Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI)*, Dubai, United Arab Emirates, 4–6 February 2019; pp. 284–291.
- [20] Butkiewicz, M.; Madhyastha, H.V.; Sekar, V. Understanding website complexity: Measurements, metrics, and implications. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, Berlin, Germany, 2–4 November 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 313–328.
- [21] Zhou, J.; Zhang, Y.; Zhou, B.; Li, S. Predicting web page performance level based on web page characteristics. *IJWET* 2015, 10, 152.
- [22] Nagarajan, S.N.; Ravikumar, S. Model for Predicting End User Web Page Response Time. *arXiv* 2012, arXiv:1204.6304.
- [23] Li, Z.; Zhang, M.; Zhu, Z.; Chen, Y.; Greenberg, A.G.; Wang, Y.-M. WebProphet: Automating Performance Prediction for Web Services. In *Proceedings of the 7th USENIX Symposium on Networked Systems Design and Implementation*, NSDI 2010, San Jose, CA, USA, 28–30 April 2010; pp. 143–158.
- [24] Allison, R.; Hayes, C.; McNulty, C.A.M.; Young, V. A Comprehensive Framework to Evaluate Websites: Literature Review and Development of GoodWeb. *JMIR Form. Res.* 2019, 3, e14372.
- [25] Armaini, I.; Dar, M.H.; Bangun, B. Evaluation of Labuhanbatu Regency Government Website based on Performance Variables. *Sink. J. Dan Penelit. Tek. Inform.* 2022, 7, 760–766.
- [26] Devi, K.; Sharma, A.K. Framework for Evaluation of Academic Website. *Int. J. Comput. Tech.* 2016, 3, 234–239.
- [27] Erokhin, A.G.; Vanina, M.F.; Frolova, E.A. Application of Mathematical Simulation Methods for Evaluating the Websites Effectiveness. In *Proceedings of the 2019 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, 20–21 March 2019; pp. 1–5.
- [28] Harshan, R.K.; Chen, X.; Shi, B. Analytic Hierarchy Process (AHP) Based Model for Assessing PerformanceQuality of Library Websites. *Inf. Technol. J.* 2016, 16, 35–43.
- [29] [29] Hidayah, N.A.; Subiyakto, A.; Setyaningsih, F. Combining Webqual and Importance Performance Analysis for Assessing a Government Website. In *Proceedings of the 2019 7th International Conference on Cyber and IT Service Management (CITSM)*, Jakarta, Indonesia, 6–8 November 2019; pp. 1–6.
- [30] Kabassi, K. Analytic Hierarchy Process for website evaluation. *Intell. Decis. Technol.* 2018, 12, 137–148.
- [31] Kwangsawad, A.; Jattamart, A.; Nusawat, P. The Performance Evaluation of a Website using Automated Evaluation Tools. In *Proceedings of the 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES iCON)*, Bangkok, Thailand, 11–13 December 2019; pp. 1–5.
- [32] G. Padmaja, Swetha Pesaru, Desidi Narsimha Reddy, D. Anitha Kumari, Shiva Prasad Maram - Robust Vehicle Number Plate Text Recognition and Data Analysis Using Tesseract Ocr, *ITM Web Conf.* 74 01009 (2025).
- [33] Tripti Sharma, Desidi Narsimha Reddy, Chamandeep Kaur, Sanjiv Rao Godla, R. Salini, Adapa Gopi and Yousef A.Baker El-Ebiary, “Federated Convolutional Neural Networks for Predictive Analysis of Traumatic Brain Injury: Advancements in Decentralized Health Monitoring” *International Journal of Advanced Computer Science and Applications(IJACSA)*, 15(4), 2024.
- [34] Tripti Sharma, Desidi Narsimha Reddy, Chamandeep Kaur, Sanjiv Rao Godla, R. Salini, Adapa Gopi and Yousef A.Baker El-Ebiary, “Federated Convolutional Neural Networks for Predictive Analysis of Traumatic Brain Injury: Advancements in Decentralized Health Monitoring” *International Journal of Advanced Computer Science and Applications(IJACSA)*, 15(4), 2024. <http://dx.doi.org/10.14569/IJACSA.2024.0150494>.
- [35] Zhang, L., Zhao, K. (2022). Boosting Time Series Forecasting Accuracy through Hybrid Data Augmentation and Ensemble Learning. *Journal of Forecasting*, 41(3), 462–481.