computers

MDPI

Systematic Review

Predicting Website Performance: A Systematic Review of Metrics, Methods, and Research Gaps (2010–2024)

Mohammad Ghattas [1,2], Suhail Odeh [1,*] and Antonio M. Mora [2]

[1] Department of Software Engineering, Faculty of Science, Bethlehem University, Bethlehem P1520468, Palestine; mohamadghattas@correo.ugr.es
[2] Department of Signal Theory, Telematics and Communications, School of Computer Sciences and Telecommunications (ETSIIT), Research Center on Information and Communication Technologies (CITIC-UGR), University of Granada, 18071 Granada, Spain; amorag@ugr.es
* Correspondence: sodeh@bethlehem.edu

Abstract

Website performance directly impacts user experience, trust, and competitiveness. While numerous studies have proposed evaluation methods, there is still no comprehensive synthesis that integrates performance metrics with predictive models. This study conducts a systematic literature review (SLR) following the PRISMA framework across seven academic databases (2010–2024). From 6657 initial records, 30 high-quality studies were included after rigorous screening and quality assessment. In addition, 59 website performance metrics were identified and validated through an expert survey, resulting in 16 core indicators. The review highlights a dominant reliance on traditional evaluation metrics (e.g., Load Time, Page Size, Response Time) and reveals limited adoption of machine learning and deep learning approaches. Most existing studies focus on e-government and educational websites, with little attention to e-commerce, healthcare, and industry domains. Furthermore, the geographic distribution of research remains uneven, with a concentration in Asia and limited contributions from North America and Africa. This study contributes by (i) consolidating and validating a set of 16 critical performance metrics, (ii) critically analyzing current methodologies, and (iii) identifying gaps in domain coverage and intelligent prediction models. Future research should prioritize cross-domain benchmarks, integrate machine learning for scalable predictions, and address the lack of standardized evaluation protocols.

Keywords: website performance; systematic literature review; PRISMA; web evaluation; Core Web Vitals; prediction models; machine learning

Academic Editor: Paolo Bellavista

Received: 1 September 2025
Revised: 10 October 2025
Accepted: 11 October 2025
Published: 20 October 2025

Citation: Ghattas, M.; Odeh, S.; Mora, A.M. Predicting Website Performance: A Systematic Review of Metrics, Methods, and Research Gaps (2010–2024). *Computers* **2025**, *14*, 446. https://doi.org/10.3390/computers14100446

1. Introduction

In today's digital era, websites serve as primary platforms for communication between institutions and their audiences. They provide direct access to information, services, and engagement opportunities across various domains. A well-performing website significantly contributes to user trust, satisfaction, and organizational competitiveness, particularly in sectors like education, e-government, commerce, and healthcare. As a result, assessing website performance has emerged as a key area of focus for both researchers and practitioners. While traditional assessments rely on basic indicators such as loading time, interactivity, usability, and design quality, recent studies emphasize broader, more integrated evaluation models that consider multiple dimensions of user experience and technical efficiency. Several studies [1–6] have introduced various frameworks and tools for evaluating website

performance. However, most of these efforts have concentrated on specific domains such as e-government and education, with limited attention given to more dynamic and rapidly evolving fields like healthcare, online commerce, and financial services.

Related Surveys and Research Gap. Prior surveys have assessed website quality within specific domains or via limited metric sets. For example, comparative evaluations and method overviews emphasized e-government portals and general assessment frameworks without unifying prediction models [7]; TFN-AHP-based automated assessment targeted government websites [8] and recent domain-bound SLRs on university sites provided scope-specific insights without validating a cross-domain metric set or linking metrics to ML/DL prediction pipelines [9]. To date, there is no review that simultaneously (i) synthesizes metrics across domains, (ii) validates those metrics with expert input, and (iii) maps them to predictive (ML/DL) approaches.

Although machine learning and deep learning have significantly advanced performance prediction in various domains, their integration into website evaluation research remains relatively unexplored. Moreover, existing studies often fail to clearly identify the most critical factors influencing website performance. There is a noticeable gap in synthesizing the available metrics, techniques, and conceptual models that underpin current evaluation practices. To address these limitations, this study conducts a systematic literature review (SLR) aimed at integrating recent findings, identifying frequently used performance metrics, and highlighting both methodological and conceptual shortcomings in current research. By doing so, the study seeks to provide new perspectives on underrepresented areas and support the development of more robust and comprehensive prediction models for website performance. In this study, 59 quality metrics were identified, filtered, and refined into a final set of 16 key indicators, as discussed later in Section 3.3.

This paper conducts a systematic literature review to offer a comprehensive overview of the key studies concerning the assessment of website quality since 2010. The objective is to identify the existing evidence on this topic and pinpoint any research gaps in evaluating quality metrics, as well as provide an overview of various techniques or popular methodologies utilized to evaluate the quality of websites. Following the introduction, the structure of this paper is as follows: Section 2 details the research methodology, Section 3 summarizes the outcomes of the Systematic Literature Review (SLR), and Section 4 discusses the research questions about website quality evaluation issues. Four factors are included in the analysis of research trends: the country of the first author, the study context, the research emphasis, and the publication year, including approaches to evaluate website performance and factors influencing it. Section 5 presents the discussion, recommendations, and critical review, with conclusions outlined in Section 6.

To the best of our knowledge, this is the first systematic literature review that integrates both traditional website evaluation metrics and modern prediction techniques, including machine learning and deep learning approaches. Unlike previous reviews that focused primarily on single domains (e.g., e-government or education) or on isolated sets of metrics, our study consolidates 59 indicators into 16 validated core metrics through expert input. The reduction from the initial pool of 223 candidate metrics to the refined set of 59 was conducted through a structured filtering protocol, including duplicate removal, synonym consolidation, operationalize ability checks, and expert consensus, ensuring transparency and reproducibility. This dual focus on methodological synthesis and metric validation provides a novel and comprehensive perspective on website performance evaluation, setting the groundwork for future intelligent and domain-independent prediction models. Building on this positioning, the next section details our PRISMA-guided methodology and the procedures used to derive and validate the final set of 16 metrics.

Novelty and Contributions. This review advances the state of the art beyond prior descriptive surveys by explicitly framing website performance evaluation as a prediction problem and by unifying evidence from both research and practice into an expert-validated, prediction-oriented synthesis. Specifically, our contributions are fourfold: (i) we consolidate a broad corpus of 2010–2024 sources into a harmonized catalogue of 223 metrics, systematically refined to 59 and then prioritized via an expert survey into 16 core key performance indicators (KPIs); (ii) we articulate how these 16 KPIs can be applied as predictive features (classification/regression targets, typical label definitions, and practical feature-engineering notes); (iii) we provide a qualitative comparative analysis of the main predictive approaches reported in the literature (e.g., SVM, Random Forest, Logistic/Linear models, Decision Trees, Naïve Bayes, KNN, and ensemble methods), highlighting their strengths, limitations, and expected suitability across different website domains (e-commerce, e-government, education, media); and (iv) we distill method-level guidance for model selection and parameterization (e.g., kernel choice and C/γ considerations for SVMs, tree depth and number of estimators for ensembles), thus bridging the gap between metric reporting and actionable prediction workflows. By repositioning the findings around these contributions, the manuscript complements the statistical trends with a domain-aware, method-centric perspective that clarifies "what to use, when, and why" for predicting website performance in real-world settings.

2. Research Methodology

The study adopts a Systematic Literature Review (SLR) methodology to examine existing academic work related to website performance evaluation. It adheres to the PRISMA guidelines and incorporates established procedures outlined by Kitchenham et al. [10], commonly used in software engineering reviews. The process consists of several stages: defining research objectives, choosing relevant databases, setting inclusion and exclusion criteria, evaluating study quality, and extracting and synthesizing key data. A PRISMA flow diagram (Figure 1) illustrates the process of study identification, screening, eligibility assessment, and inclusion.

Based on this methodology, the following research questions were formulated.

2.1. Research Questions

The goal of this study is to explore the key elements that impact the evaluation of website performance. To support this objective, specific research questions were carefully formulated. With the increasing demand for high-performing websites, the interest in their design and development has grown notably. Nonetheless, traditional evaluation methods remain manual, time-intensive, and often inconsistent. As a result, there is a pressing need for automated tools or intelligent models that can support developers in accurately assessing website performance.

To directly support applied decision-making, each research question is framed to (i) identify concrete methods for specific challenges, (ii) clarify applicability across website domains, and (iii) distill practical guidance for model selection and hyperparameter configuration.

To identify the extent of studies conducted in this field, this systematic literature review (SLR) addresses the following research questions:

RQ1: What methodological challenges and threats to validity arise in predicting website performance, and how do they impact study design and evidence quality?

RQ2: Which predictive approaches (e.g., SVM, Random Forest, Logistic/Linear models, Decision Trees, Naïve Bayes, KNN, ensembles, deep learning) are applied to model website performance, and what are their main strengths and limitations?

RQ3: How applicable and transferable are these predictive methods and feature sets across different website domains (e-commerce, e-government, education, media, healthcare)?

RQ4: What practical configuration guidelines (feature engineering choices, model selection, and hyper-parameter settings) and evaluation protocols lead to reliable and reproducible prediction performance?

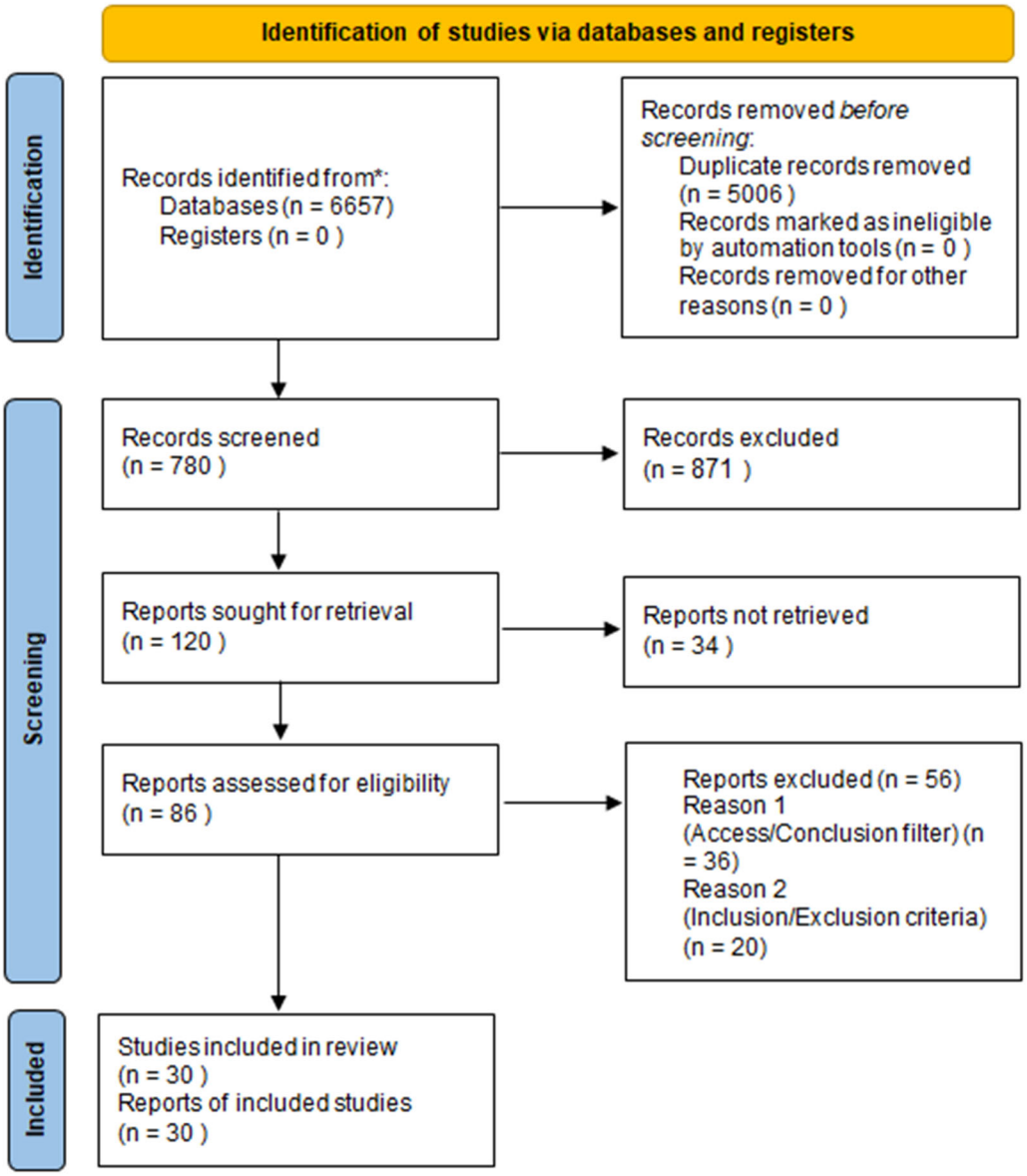

Figure 1. PRISMA flow diagram showing the identification, screening, eligibility, and inclusion process. * Databases searched: Scopus, Web of Science, IEEE Xplore, ScienceDirect, and SpringerLink.

2.2. Search Process

The literature search was carried out in two distinct phases. During the first phase, seven well-established academic databases were selected: Scopus, IEEE Xplore, SprinerLink, ResearchGate, ACM Digital Library, the Directory of Open Access Journals (DOAJ), and Google Scholar. These sources were chosen based on their broad accessibility to peer-reviewed publications in the fields of computer science, software engineering, and web technologies. Their inclusion helped ensure a balanced representation of both influential journals and newer studies addressing website performance. Table 1 summarizes the databases consulted in this review.

Table 1. Systematic literature review databases.

Online Database	URL
IEEE Xplore	http://ieeexplore.ieee.org/ (accessed on 10 October 2025)
Google Scholar	https://scholar.google.com/ (accessed on 10 October 2025)
Scopus	http://www.scopus.com/ (accessed on 10 October 2025)
SpringerLink	https://link.springer.com/ (accessed on 10 October 2025)
ResearchGate	https://www.researchgate.net/ (accessed on 10 October 2025)
ACM Digital Library	https://dl.acm.org/ (accessed on 10 October 2025)
Directory of Open Access Journals (DOAJ)	https://doaj.org/ (accessed on 10 October 2025)

The second phase involved manual screening of reference lists from initially selected studies to identify additional relevant papers. The search terms were derived from the research questions and refined using synonyms and Boolean operators (e.g., "website performance" AND "evaluation" OR "prediction" OR "quality metrics").

2.3. Study Selection

This review strictly followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines throughout the study selection process, as Analyses in Figure 1 [11]. The selection involved a multi-stage filtering approach. Initially, 6657 records were identified. After excluding duplicate entries and non-English publications, 1651 articles remained for preliminary screening. Titles and abstracts were reviewed to retain 780 studies. Further screening considering accessibility, methodological rigor, and alignment with the research questions narrowed the pool to 34 studies that were ultimately included in the review. A simplified numeric overview of the study selection stages, including QA filtering, is provided in Figure 2 to complement the PRISMA diagram (Figure 1).

To minimize subjectivity, screening decisions were independently verified by multiple reviewers, and any disagreements were resolved through discussion and consensus, ensuring consistency and reproducibility in the selection process.

It should be noted that a substantial number of potentially relevant articles were excluded at this stage due to lack of full-text access. While necessary to ensure methodological rigor and allow for consistent data extraction, this exclusion may have limited the overall coverage of the review and introduced a potential risk of passive selection bias.

2.4. Inclusion and Exclusion Criteria

To maintain the quality and focus of this review, inclusion and exclusion criteria were explicitly defined, as shown in Table 2. Eligible studies were those published between 2010 and 2024, written in English, and appearing in peer-reviewed journals or conference proceedings. Additionally, studies had to address website quality, performance evaluation, or prediction approaches and provide either a detailed methodology or empirical evidence. Studies were excluded if they were non-peer-reviewed (e.g., blogs, editorials), lacked methodological transparency, focused on unrelated areas such as SEO or social media without relevance to performance, or were inaccessible in full text.

To minimize selection bias, each article was independently reviewed by two researchers. Discrepancies were resolved through consensus. Additionally, borderline cases were discussed with a third reviewer to ensure objectivity and consistency.

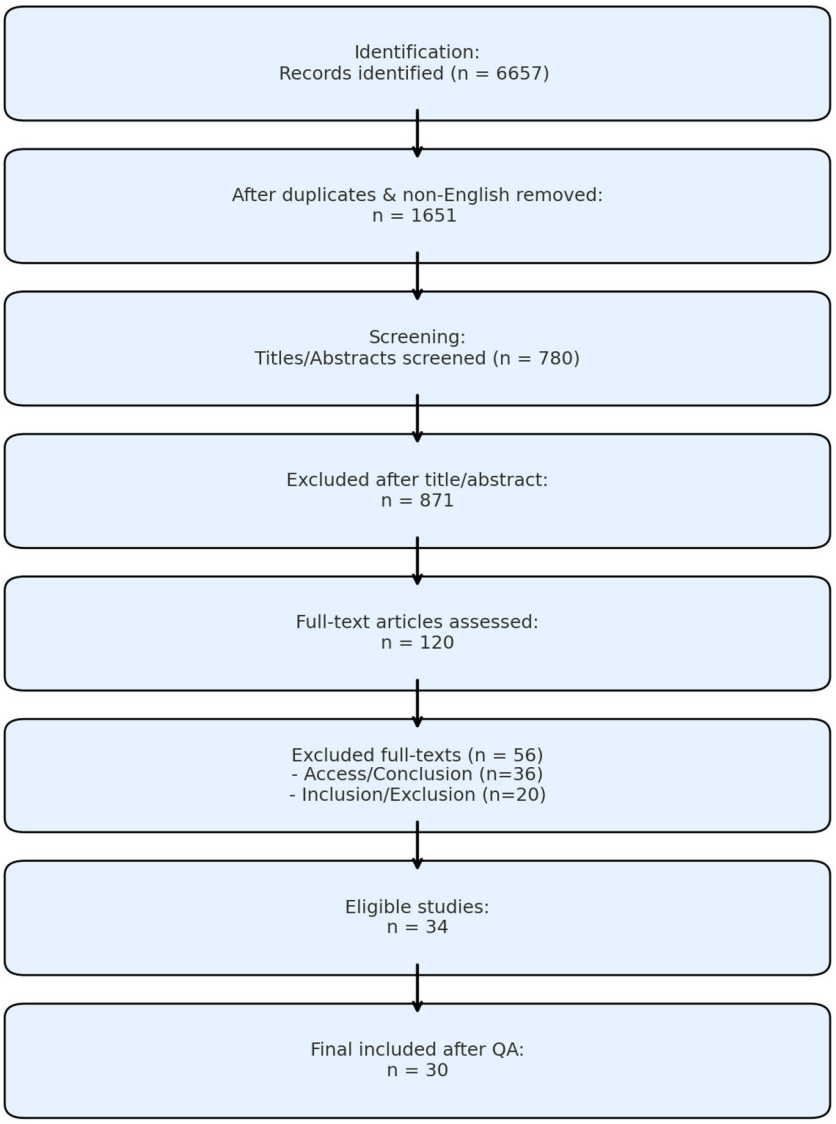

Figure 2. Simplified overview of the study selection process. Although 34 studies were initially eligible, 4 were excluded after QA assessment, resulting in 30 studies for the final synthesis (see Figure 1).

Table 2. Inclusion and exclusion criteria for study selection.

Inclusion Criteria	Exclusion Criteria
Published 2010–2024	Not peer-reviewed
English language	Lack of methodology
Web quality focus	Focus on unrelated topics
Peer-reviewed	Inaccessible full text

2.5. Quality Assessment

A structured quality assessment was implemented to evaluate the rigor and trustworthiness of the selected studies, using a nine item checklist adapted from Ghobadi et al. [12]. Each article was reviewed based on criteria covering research objectives, methodological clarity, data accuracy, and the validity of results. Responses were scored as 1 for "Yes," 0.5 for "Partially," and 0 for "No." The maximum score per study was 9 points. A minimum threshold of 4.5 was applied to filter out studies with insufficient quality, ensuring a

balance between inclusiveness and methodological soundness. Table 3 lists the assessment questions employed in this evaluation.

Table 3. Quality assessment questions used in the review.

QA Code	Assessment Question
Q1	Is the research objective clearly stated?
Q2	Is the context and scope of the study well-defined?
Q3	Is the methodology appropriate and clearly described?
Q4	Are the data sources valid and reliable?
Q5	Are the performance evaluation metrics clearly defined?
Q6	Are the results clearly presented and supported by data?
Q7	Are limitations discussed and addressed?
Q8	Does the study contribute new knowledge or findings?
Q9	Is the overall structure and academic quality of the article satisfactory?

To enhance objectivity and minimize subjective bias, two independent reviewers conducted the assessment. In cases of disagreement, a third reviewer was consulted to reach consensus. All scores were reviewed and corrected accordingly. Detailed QA results for all 34 included studies are presented in Table A1.

2.6. Data Extraction

A systematic procedure was followed to extract and consolidate relevant data from the selected studies. The objective was to obtain consistent, detailed, and comparable information across all sources. Key fields collected included article identifiers, reference data, the performance metrics under investigation, applied methodologies, and the study context. These extraction criteria are outlined in Table 4, while complete bibliographic references are provided in Table A2.

Table 4. Data extraction form.

Field	Description
ID	Unique identifier assigned to each article for referencing.
Title	The title of the article.
Authors	The author(s) of the article.
Publication Year	The publication year of the article.
Country	The country in which the research was conducted.
Performance Factors	Website performance aspects studied (e.g., load time, page size).
Methodologies	Research approaches used (e.g., classification, clustering, regression).
Techniques	Specific algorithms employed (e.g., SVM, Neural Networks, Decision Tree).
Context	Details about the research participants.

This structured data extraction enabled consistent comparison and analysis across all included studies. The extracted information formed the basis for the synthesis of patterns and findings discussed in the following section.

2.7. Data Synthesis

The extracted data from the 34 studies were analyzed using a qualitative synthesis strategy aimed at uncovering common patterns, methodologies, and key indicators associated with website performance prediction. The studies were grouped according to factors such as publication year, research method, evaluation framework, and application domain (e.g., education, government, healthcare).

To facilitate analysis, performance features and techniques were clustered using thematic coding, while frequency analysis highlighted the most frequently adopted algorithms and metrics. These insights are discussed in Section 3 and serve as a basis for identifying current limitations and suggesting future research pathways.

3. SLR Results

In this section, we provide an overview of the findings from the systematic literature review (SLR). The chosen articles, accompanied by pertinent details, are presented first, followed by an assessment of their quality.

3.1. Search Results

An initial search across seven academic databases resulted in the identification of 6657 articles. After eliminating duplicates and excluding non-English publications, 1651 records were eligible for preliminary screening. Based on titles and abstracts, 780 were removed, narrowing the list to 871 studies. A more detailed review considering full text availability and alignment with the study's research questions reduced the set to 120 candidates. Of these, 34 met all inclusion criteria. However, after applying the quality assessment framework (Section 2.5), four studies were excluded due to inadequate methodological rigor, leaving 30 high-quality studies for the final analysis.

The study selection process is illustrated in Figure 1 using the PRISMA flowchart methodology. Additionally, the detailed scores from the quality assessment phase, which contributed to the inclusion decisions, are summarized in Table 3 (see Section 2.5). The detailed distribution of articles retrieved from each database, and the filtering decisions taken at each phase, are presented in Table 5.

Table 5. Overview of the search and selection process across databases.

Database	Initial Results	After Duplicate Removal	After Title/Abstract Filter	After Access Filter	After Abstract/Conclusion Filter	After Inclusion/Exclusion	After QA
IEEE Xplore	243	17	17	17	12	5	5
Google Scholar	1043	431	178	23	15	9	7
Scopus	104	60	60	13	8	5	5
ResearchGate	2690	496	268	9	18	3	2
SpringerLink	705	23	23	23	8	2	2
ACM	661	266	73	24	16	2	2
DOAJ	1211	358	161	11	9	8	7
Total	6657	1651	780	120	86	34	30

Table 5 provides a detailed breakdown of the number of articles retrieved and filtered at each stage of the review process across the selected databases. The figures correspond to those presented in the PRISMA flowchart (Figure 1), ensuring consistency and methodological transparency.

As part of the synthesis phase, 59 distinct website performance metrics were extracted from the selected 34 studies. These indicators were organized into eight thematic categories,

covering aspects such as performance, accessibility, usability, design, SEO, content, and technical characteristics. Table 6 outlines these groupings, while full metric descriptions are provided in Table A3.

Table 6. Categorized website quality metrics identified from the literature (# = number of metrics).

# of Metrics	Sample Metrics	Category
12	Load time, TTFB, Page size	Performance
8	Alt text, Color contrast	Accessibility
9	Navigation, Readability	Usability
7	Meta tags, Link structure	SEO
6	Layout, Mobile responsiveness	Design Quality
7	Relevance, Freshness	Content Quality

Note: a complete list of the 59 extracted metrics is provided in Table A3.

Furthermore, the annual publication trend of the selected articles from 2010 to 2024 is presented in Figure 3, indicating a significant increase in research interest starting in 2016.

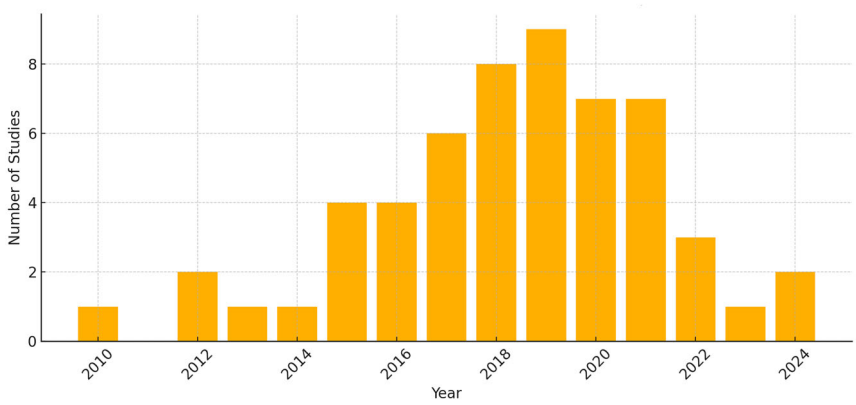

Figure 3. Annual distribution of the 34 studies initially included before QA filtering (2010–2024).

3.2. Quality Assessment Results

To ensure the validity and methodological rigor of the included studies, each of the 34 initially selected articles was assessed using a standardized nine-question quality assessment (QA) framework (see Section 2.5 and Table 3). Each question was scored as Yes = 1, Partial = 0.5, or No = 0, resulting in a total possible score of 9 per study. Based on the predefined inclusion threshold of 4.5 points, 30 studies met the required quality level and were included in the final synthesis. The remaining four studies were excluded due to low scores, insufficient methodological detail, or unclear data reporting.

The QA evaluation was performed independently by two reviewers, with any disagreements resolved by consensus or consultation with a third reviewer. All scores were reviewed and recalculated to ensure accuracy. For instance, the total QA score for study P2 was corrected from 6 to 6.5 after manual verification.

A full breakdown of QA scores for each study is presented in Table A1.

3.3. Key Quality Factors and Evaluation Methods

Initially, 223 quality factors were extracted based on a comprehensive review of literature, standards, and practical performance considerations. This initial pool was refined by eliminating duplicates and applying memoing and filtering techniques, resulting

in a more focused set of 59 relevant metrics (see Table 7, Table A4 for the extraction criteria, and Table A3 for the complete list of metrics).

Table 7. Final set of 16 selected web performance quality metrics with operational definitions and measurement methods. (Answers RQ4).

No.	Selected Metric	Operational Definition	Measurement Method
1	Load Time	Total duration required for a webpage to fully load all resources (HTML, CSS, JS, images).	Measured in milliseconds using tools such as Google Lighthouse, GTmetrix, or WebPageTest.
2	Time to First Byte (TTFB)	The time between initiating a request and receiving the first byte from the server.	Measured in ms via browser dev tools or performance testing tools.
3	Page Size	The total size of the webpage including all assets (HTML, CSS, scripts, images).	Measured in KB/MB using performance testing tools.
4	Number of Requests	The total number of HTTP(S) requests made to load a webpage.	Counted via dev tools or testing tools like WebPageTest.
5	Time to Interactive (TTI)	The time it takes for a page to become fully interactive for the user.	Measured in ms using Lighthouse.
6	Largest Contentful Paint (LCP)	Time taken for the largest visible content element (image/text block) to render in the viewport.	Measured in ms using Lighthouse/Core Web Vitals.
7	Total Link	The number of hyperlinks included in the webpage.	Counted using HTML parsers or crawler tools.
8	Byte In	The total amount of data transferred from the server to load the page.	Measured in KB/MB using WebPageTest or network monitors.
9	Start Render Time	Time when the browser starts painting the first pixels on the screen.	Measured in ms using WebPageTest or Lighthouse.
10	Document Complete Time	The time until the document and resources are fully loaded.	Measured in ms using WebPageTest or GTmetrix.
11	Speed Index	A user-centric metric showing how quickly page content is visually displayed.	Measured in ms using Lighthouse or WebPageTest.
12	Compression	The use of resource compression (e.g., GZIP, Brotli) to reduce file size.	Checked via HTTP headers or Lighthouse audits.
13	Broken Links Detection	Identifies invalid or non-functioning hyperlinks on the page.	Evaluated using crawler tools (e.g., ScreamingFrog, W3C Link Checker).
14	Markup Validation (HTML Errors)	Detects errors in HTML structure affecting compatibility and rendering.	Measured using W3C Validator or similar tools.
15	Response Time	The time a server takes to respond to a client request.	Measured in ms using dev tools or monitoring platforms.
16	Design Optimization	Assessment of layout efficiency, visual hierarchy, and responsive design practices.	Evaluated qualitatively and with tools (e.g., Lighthouse best-practice audits).

Reduction Protocol (223 $\rightarrow$ 59). From an initial pool of 223 candidate metrics, we followed a structured reduction pipeline. (i) Duplicate handling: exact and near-duplicate items were removed after normalizing names and consolidating synonyms (e.g., "page weight" and "page size"). (ii) Operationalizability: metrics without a clear operational definition or that could not be objectively measured were excluded. (iii) Scope relevance: items not aligning with web-performance constructs were filtered out according to the extraction criteria (Table A4). (iv) Cross-study support: metrics reported in $\geq$2 independent studies or standards were retained; singletons were kept only when strongly justified. (v) Memoing and consensus: two memoing rounds were used to collapse overlapping concepts and resolve borderline cases, leading to a final set of 59 metrics (complete list in Table A3).

To further validate and prioritize these metrics, we conducted an online survey targeting 35 web developers, performance experts, and industry professionals. The survey (see Figure 4) included the refined list of 59 metrics, and participants were asked to rate each one on a scale of 1 (poor) to 3 (excellent).

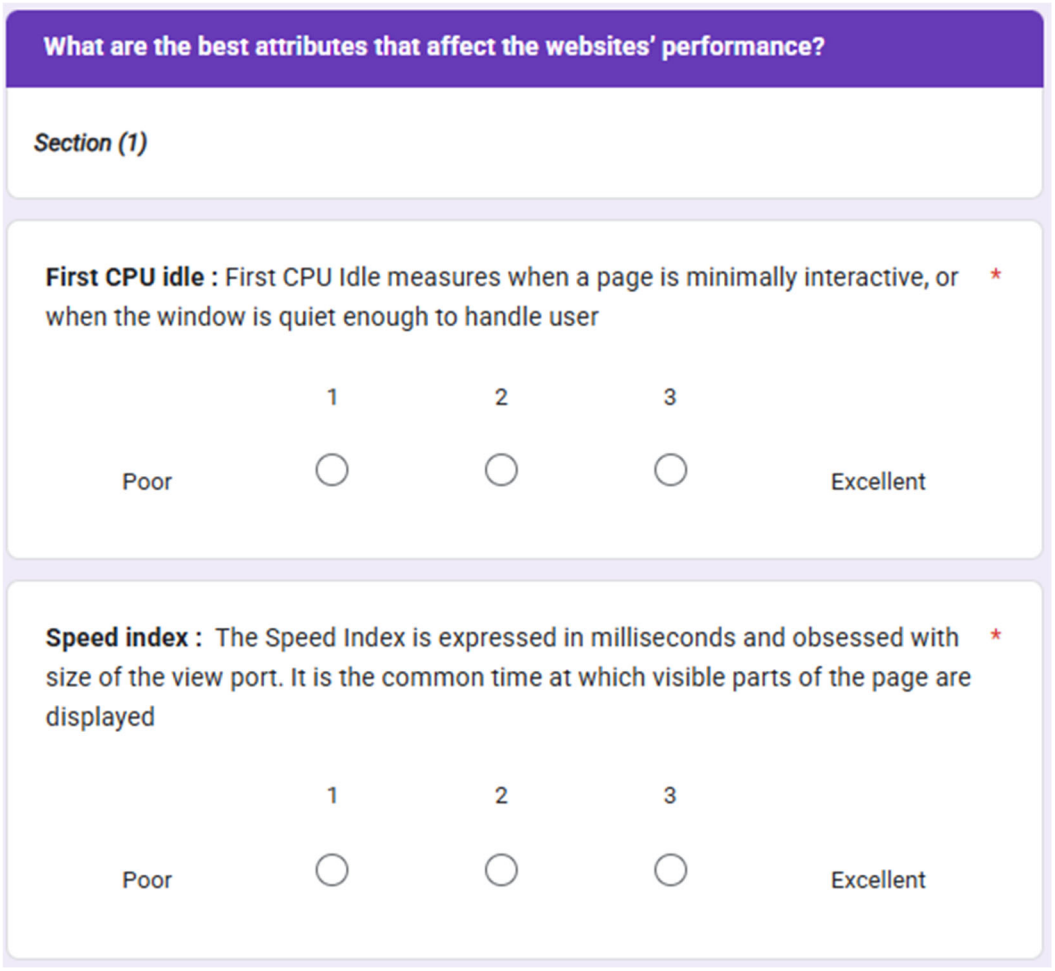

Figure 4. Example of the online survey questionnaire (* = required question).

Although broader scales (e.g., 5- or 7-point) are commonly used to capture more nuanced expert opinions, the reduced 3-point scale was deliberately chosen for this study. The primary objective was not to measure subtle differences in perception, but rather to filter and prioritize a large pool of candidate metrics into a concise set of key indicators. A simplified scale minimized respondent fatigue, ensured consistency across a heterogeneous group of experts, and facilitated transparent threshold-based decisions (i.e., metrics rated "3" by more than 50% of experts were directly included in the final list). This trade-off was considered acceptable given the exploratory and reduction-oriented purpose of the survey. The full questionnaire used in this survey, including all 59 evaluated metrics, is provided in Appendix F for reference.

The 35 participants represented diverse demographic and professional backgrounds to ensure the credibility and relevance of the collected feedback. Specifically, the sample included 21 males and 14 females from five countries (Palestine, Jordan, Egypt, Lebanon, and Spain). Age groups ranged from 25 to 50 years, and professional roles were categorized as junior developers (10), senior developers (12), and technical leads or researchers (13). Most participants held at least a bachelor's degree, with 15 holding master's or doctoral degrees. Years of experience varied from 2 to 20 years, ensuring that the survey results reflected both academic knowledge and practical industry expertise. Figure 5 illustrates the demographic distribution of the survey participants by gender, age, professional role, years of experience, and education level, while Table 8 provides a detailed summary of their demographics. This diversity of perspectives strengthens the reliability and representativeness of the expert validation process.

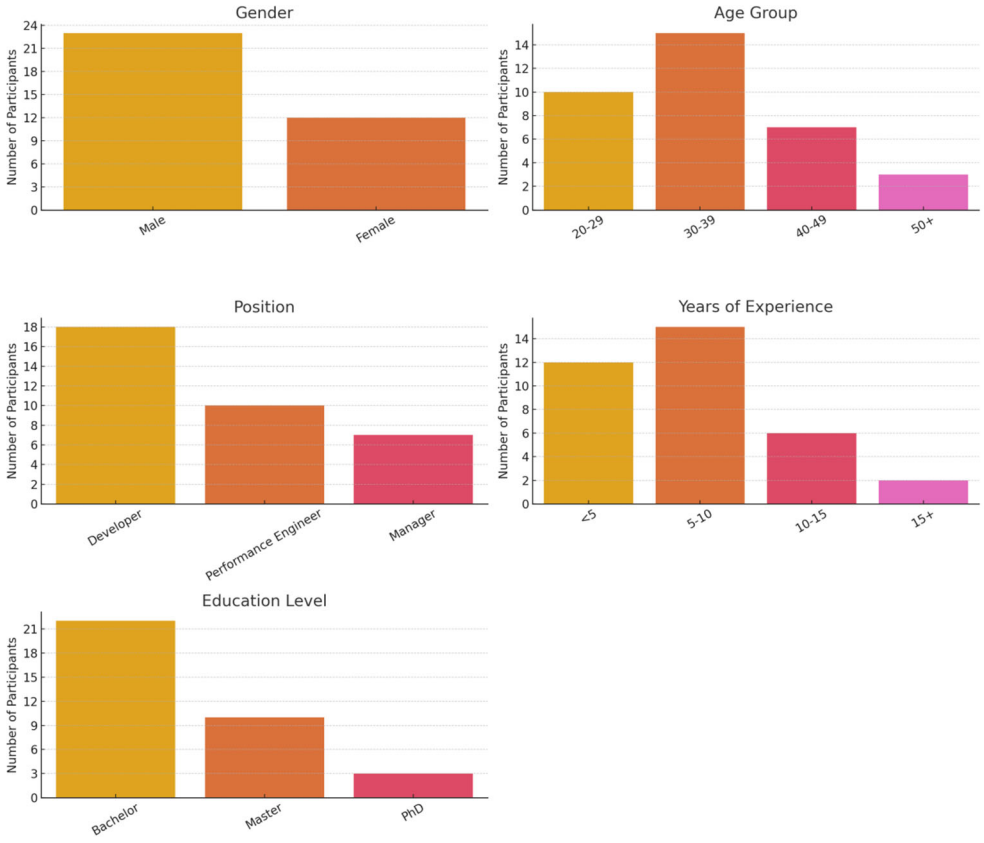

Figure 5. Distribution of survey participants by demographic categories (gender, age group, position, years of experience, education level).

Table 8. Summary of survey participant demographics.

Attribute	Categories
Gender	21 Male, 14 Female
Age Groups	25–34 (15), 35–44 (13), 45–50 (7)
Professional Role	Junior Developer (10), Senior (12), Researcher/Lead (13)
Years of Experience	2–5 (9), 6–10 (14), 11–20 (12)
Education Level	Bachelor's (20), Master's (10), PhD (5)
Country of Residence	Palestine, Jordan, Egypt, Lebanon, Spain

A threshold-based selection approach was employed to derive the final set of metrics. Metrics that received a score of 3 (excellent) from more than 50% of participants were immediately included in the final list. For metrics that received ratings between 40% and 50%, a weighted evaluation was applied, factoring in the metric's real-world applicability and significance in web performance contexts. For instance, metrics such as load time and responsiveness were prioritized due to their direct impact on user experience.

This multi-step filtering and expert validation process resulted in the selection of 16 core performance quality factors, which are used in the subsequent analysis. The final 16 quality metrics selected through expert validation and statistical filtering are presented in Table 7, which now includes their operational definitions and measurement methods to ensure transparency and practical applicability. To enhance clarity and replicability, the final set of 16 selected web performance quality metrics is presented below, together with their operational definitions and measurement methods.

4. Discussion

This section addresses the four research questions as follows: RQ1 is covered in Section 4.1 (evaluation challenges), RQ2 in Sections 4.2 and 4.3 (research trends) and Section 4.7 (predictive methods), RQ3 in Sections 4.5 and 4.7 (domain applicability), and RQ4 in Sections 4.8 and 4.9 (key performance indicators and configuration guidelines).

For clarity, figures and tables are explicitly tagged in the captions with the research question(s) they answer (e.g., 'Answers RQ2') to make the linkage between evidence and each RQ immediately visible.

4.1. Challenges in Evaluating Website Performance

Despite the increasing number of tools and frameworks for assessing website performance, several persistent challenges remain in the evaluation process [7,13]. Based on the systematic analysis of selected studies, these challenges can be categorized into three key perspectives:

Researcher-based: Difficulties in standardizing evaluation criteria, inconsistencies in methodology, and limited access to robust datasets often hinder the comparability and generalizability of results.

Developer-based: Web developers frequently face challenges related to the complexity of evaluation tools, limited awareness of standardized metrics, and restricted resources to implement advanced analyses.

Website-based: The diversity of website types, the dynamism of web content, and the variation in user expectations and technical environments complicate the development of universal performance indicators.

Table 9 summarizes these challenges alongside suggested mitigation strategies and references from the reviewed literature.

Table 9. Challenges in website performance evaluation categorized by stakeholder perspective.

Aspect	Problem	Potential Solution	References
Researchers' Aspects	Variations in evaluation methodologies	Standardize protocols and metrics	[14,15]
	Focus on e-government and education domains	Encourage interdisciplinary studies and funding	[7,13]
	Subjective evaluation criteria	Use objective, standard-based measures	[7]
	Issues of validity and reliability	Apply statistical validation and peer review	[16]
Developers' Aspects	Limited time, budget, and expertise	Use cost-effective tools with minimal requirements	[15–17]
	Complexity of existing tools	Provide user-friendly interfaces and guidance	[6,7]
	Dynamic nature of websites	Employ agile and continuous evaluation	[18,19]
Website Evaluation Aspects	Domain diversity makes generalization hard	Customize evaluation per domain	[7,9,13]
	Include usability testing and feedback	User evaluation complexity	[14,15]
	Create standardized benchmarks	Cross-domain comparability is limited	[9,13]

4.2. Distribution of Research Focus

The selected studies (n = 30) were analyzed to determine their primary research focus. As shown in Figure 6, the majority of the studies (49%) focused on identifying and analyzing performance-related quality factors, highlighting the growing academic interest in categorizing measurable website features that influence performance.

A further 42% of the studies emphasized evaluation approaches, aiming to benchmark or assess existing websites using specific metrics or tools. These works contributed to understanding practical applications but often lacked generalizability due to limited datasets and narrow case studies.

Only 9% of the studies incorporated machine learning or deep learning techniques to predict website performance or classify quality levels. This highlights a significant research gap in leveraging AI-based methods for early-stage, automated web performance prediction. (see Table A5 for the research focus of the selected articles).

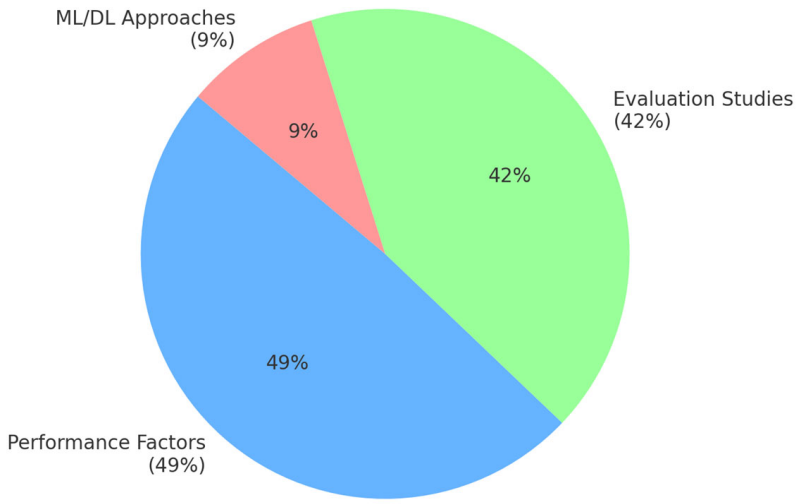

Figure 6. Distribution of research focus among selected studies.

This distribution suggests that while the field is mature in identifying performance factors and conducting structured evaluations, the integration of advanced prediction models remains limited, representing a clear opportunity for innovation in future research.

4.3. Trends by Publication Year

The temporal distribution of the selected studies (n = 30) from 2010 to early 2024 is presented in Figure 7. The data shows a gradual increase in research interest, particularly after 2015, with a peak observed in 2019. From 2020 onward, the number of publications remained relatively stable, although a slight decline was noted during 2022–2024.

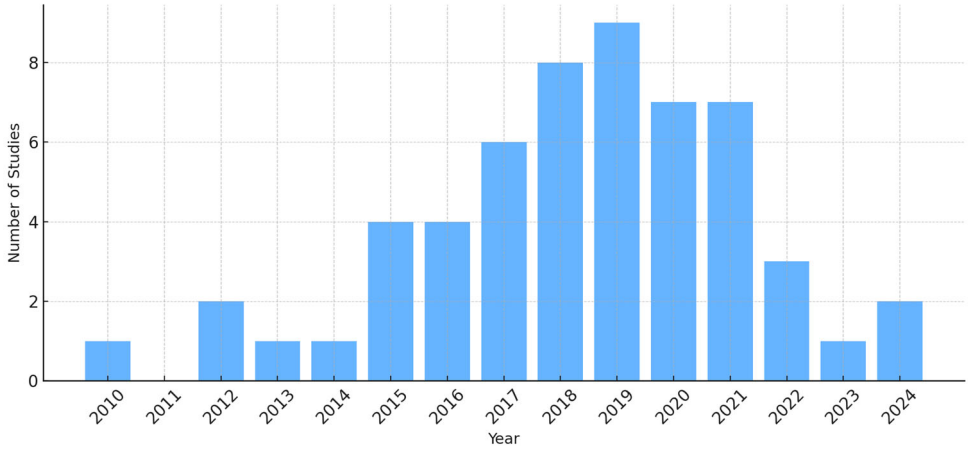

Figure 7. Annual distribution of the final 30 high-quality studies after QA filtering (2010–2024).

This upward trend reflects the growing the growing academic and industrial concern with website performance and user experience over the past decade. The surge in publications around 2018–2020 may be linked to the widespread adoption of mobile-first development practices and the emergence of performance-centric frameworks such as Google Lighthouse and Core Web Vitals.

In addition, a comparison between early and recent studies shows were more exploratory, later works tend to incorporate structured evaluations and, in some cases, predictive modeling, indicating a shift towards data-driven methodologies.

4.4. Research Geography

The geographic distribution of the final 30 selected studies, based on the first author's country, is presented in Figure 8. The majority of contributions originated from India (5 studies), Turkey (4), and China (4), indicating strong academic interest in website performance evaluation within these regions.

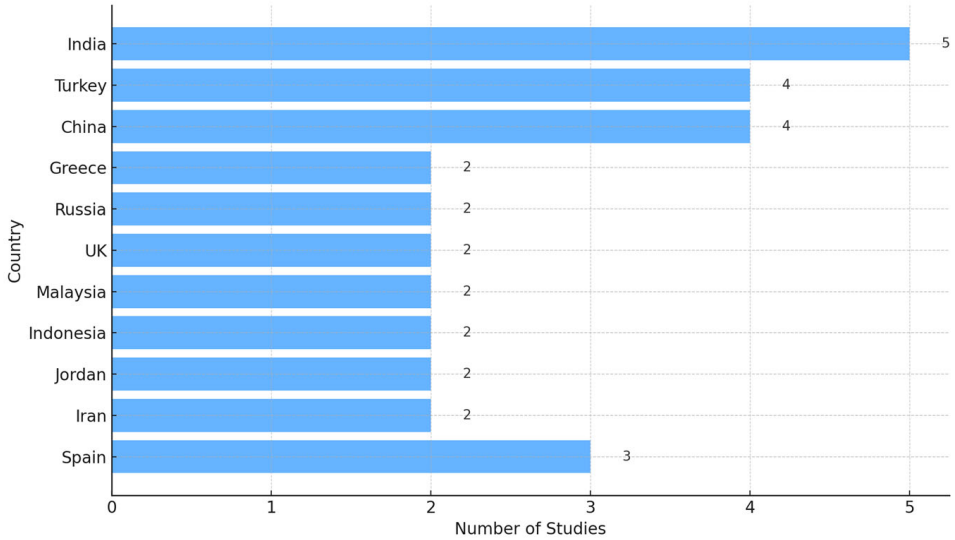

Figure 8. Distribution of selected studies by first author's country.

Additional contributions came from countries including Spain (three studies), and Greece, Russia, the United Kingdom, Malaysia, Indonesia, Jordan, and Iran with two studies each. This diverse yet regionally concentrated distribution suggests that research in this area is particularly active in Asia and parts of Europe, while remaining underrepresented in North America, Africa, and Latin America. These findings highlight the need for broader geographical inclusion and the development of globally applicable benchmarks, as well as opportunities for cross-regional collaboration and knowledge transfer.

However, it is important to note that the relatively small sample size included in this review may limit the representativeness of the geographic distribution. The apparent concentration of research in Asia and parts of Europe could partly reflect the limited dataset rather than the actual global state of research activity. Nonetheless, the findings still provide useful insights into regional research patterns. Therefore, these results should be interpreted with caution. Future reviews with broader coverage may provide a more balanced global perspective.

4.5. Study Context and Website Types

The selected studies were further categorized according to the types of websites analyzed. As shown in Figure 9, government and educational websites were the most frequently examined, with 10 and 8 studies, respectively. Other investigated domains included healthcare and e-commerce (3 studies each), nonprofit organizations and library systems (2 each), along with a small number focusing on tourism and municipal websites.

This pattern highlights a strong emphasis on public and academic web platforms, potentially due to the ease of access to such sites. Conversely, areas like finance, industry, and entertainment are noticeably less represented, pointing to gaps in the existing body of research.

These gaps present opportunities for future studies to address diverse performance contexts, user expectations, and technical challenges specific to these neglected domains.

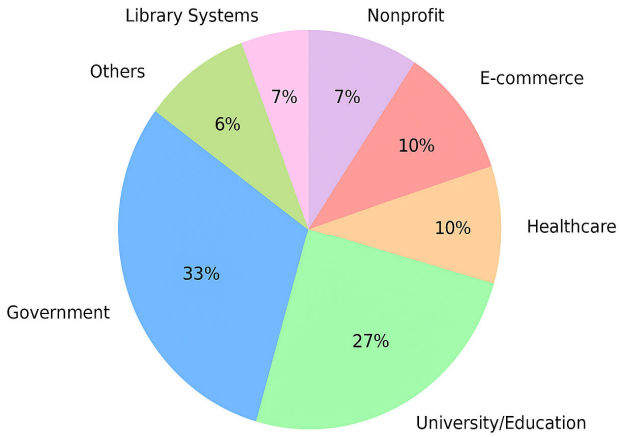

Figure 9. Classification of selected studies by website type.

4.6. Summary of Research Trends

This systematic review reveals several key research trends in the field of website performance evaluation:

Research Focus: Nearly half of the studies focused on identifying quality factors that affect performance, while only a small portion investigated predictive modeling using ML/DL techniques, which indicates a noticeable gap in automation and intelligent performance prediction.

Publication Growth: The number of studies increased steadily after 2015, reaching a peaking in 2019, reflecting a growing interest in performance optimization, possibly driven by technological advances and rising user expectations.

Geographic Distribution: Most studies originated from a limited set of countries (e.g., India, Turkey, China), which highlights strong regional involvement but also underscores underrepresentation from other global regions, such as North America and Africa.

Website Types: Research was predominantly targeted at government and educational platforms. Business, healthcare, and entertainment websites received less attention, which suggests an imbalance in domain representation.

These insights highlight the need for more balanced, cross-regional, and cross-domain research. Moreover, future studies are encouraged to integrate advanced machine learning frameworks, conduct comparative cross-country investigations, and examine performance trade-offs across different industries to build more generalized and scalable evaluation models.

4.7. What Are the Approaches Used to Predict Website Performance?

Among the 30 selected studies, a subset focused on predicting website performance using various analytical and computational approaches. As illustrated in Figure 10, the most commonly used methods fall into three major categories:

Machine Learning Algorithms: These include supervised models such as Decision Trees (DTs), Support Vector Machines (SVMs), Random Forest (RF), and Naive Bayes (NB). These models were applied in several studies to predict key performance outcomes (e.g., loading speed, responsiveness). Ensemble methods including AdaBoost were also employed to improve classification accuracy.

Statistical and Heuristic Methods: Several studies used regression analysis, fuzzy logic, or rule-based scoring models to estimate performance scores based on selected metrics. These methods, while less flexible, offer interpretability and are useful when datasets are small or incomplete.

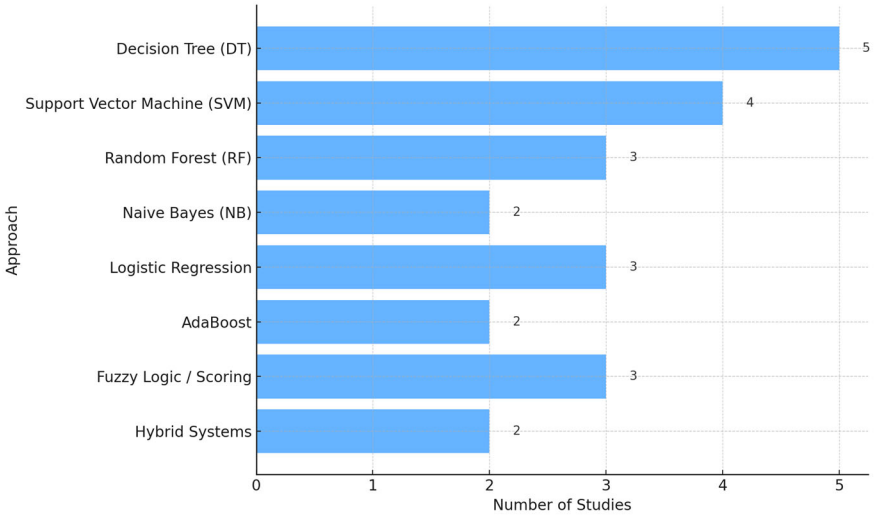

Figure 10. Distribution of prediction approaches (ML, statistical, hybrid), providing evidence for RQ3 on methods used to predict website performance. (Answers RQ2).

Hybrid and Intelligent Systems: A limited number of studies implemented neuro-fuzzy models, hybrid AHP-ML pipelines, or knowledge-based expert systems that combine rule engines with data learning. These approaches demonstrate innovation but are not yet widely adopted.

Table 10 summarizes the comparative strengths, limitations, and domain applicability of the main predictive approaches identified in the reviewed studies.

Table 10. Comparative analysis of predictive approaches for website performance evaluation (answers RQ2 and RQ3).

Approach/Method	Strengths	Limitations	Suitable Domains/Contexts
Support Vector Machine (SVM)	High accuracy with small-to-medium datasets; effective for classification; robust against overfitting with proper kernel choice.	Sensitive to parameter tuning (C, γ); computationally expensive with large datasets.	E-commerce (traffic prediction), Education (content-heavy sites).
Random Forest (RF)	Handles high-dimensional data; robust to noise and imbalance; provides feature importance ranking.	Less interpretable; slower training with very large datasets.	E-government portals; Healthcare (multi-factor performance).
Decision Trees (DTs)	Easy to interpret and visualize; fast training; suitable for categorical features.	Prone to overfitting; limited generalization without pruning/ensembles.	Educational sites; Small-scale organizational portals.
Naïve Bayes (NB)	Extremely fast and efficient; works well with text/content features; low data requirement.	Assumes independence among features; lower accuracy in complex scenarios.	News/media sites; Content-driven platforms.
Logistic/Linear Regression	Simple, interpretable; effective baseline for binary outcomes.	Limited in capturing non-linear relationships; lower predictive power.	Benchmarking studies; Simple performance classification.

Table 10. *Cont.*

Approach/Method	Strengths	Limitations	Suitable Domains/Contexts
K-Nearest Neighbors (KNN)	Non-parametric; intuitive; adapts easily to new data.	Inefficient with large datasets; sensitive to noisy/irrelevant features.	Social media; User-interaction heavy sites.
Ensemble Methods (AdaBoost, Gradient Boosting, XGBoost)	High predictive accuracy; reduces variance and bias; robust in complex data scenarios.	Higher complexity; harder to interpret; longer training times.	Cross-domain applications; Large heterogeneous datasets.
Statistical/Heuristic (Regression, Fuzzy, Rule-based)	Interpretable; useful with limited/incomplete data; simple implementation.	Limited adaptability; lower accuracy with complex/large datasets.	Early-stage studies; Benchmarking frameworks.
Hybrid/Intelligent (Neuro-fuzzy, AHP-ML, Expert Systems)	Combine strengths of multiple paradigms; innovative; context-aware.	Limited adoption; higher complexity; lack of standardized frameworks.	Specialized domains (finance, healthcare, smart systems).

Figure 10 further illustrates the number of studies that used each predictive approach. While ML techniques are gaining momentum, traditional models still dominate due to their simplicity and accessibility.

In terms of specific performance aspects, our synthesis suggests that SVM and ensemble methods are particularly effective for predicting latency-related indicators such as the Load Time, Response Time, and Time to First Byte, where classification boundaries are sharp and parameter tuning allows robust accuracy [18,20,21]. Random Forest demonstrates strong applicability in handling multi-factor aspects such as the Page Size, Number of Requests, and composite performance scores due to its ability to manage high-dimensional data and rank feature importance [21,22]. Naïve Bayes, though less accurate for complex scenarios, is promising for content-driven aspects such as usability or SEO-related indicators, where text features or categorical distributions dominate [20,22]. Decision Trees provide a lightweight alternative for smaller datasets and simple categorical aspects (e.g., presence of Broken Links or Markup Validation) [20,22]. Regression models remain useful as interpretable baselines for binary or threshold-based outcomes, such as distinguishing between acceptable and unacceptable loading speeds [18,20].

Method–KPI suitability synthesis. Across the reviewed studies, latency-related indicators (Load Time, TTFB, Response Time) are most effectively modeled with margin-based classifiers and ensembles (e.g., SVM with RBF kernel; Gradient-Boosted Trees), which cope well with non-linear boundaries and heterogeneous signals [18,20,21]. Multi-factor aspects (Page Size, Number of Requests, composite scores) benefit from Random Forest due to its robustness to high-dimensional inputs and the availability of feature-importance diagnostics [21,22]. Content-driven or categorical facets (e.g., link integrity, markup validity) admit lightweight baselines (Decision Trees, Naïve Bayes) [20,22] complemented by interpretable Logistic/Linear models for thresholding (acceptable vs. unacceptable) [18]. These mappings indicate that choosing the right learner depends on the dominant performance facet—temporal latency vs. structural complexity vs. content signals thus answering RQ2 (methods) and RQ3 (applicability across website types).

4.8. What Are the Key Performance Indicators (KPIs) Used in Studies?

While Section 3.3 presented the final set of 16 key quality indicators derived through expert consultation and data filtering, this section examines their frequency of use across the 30 selected studies.

Based on the systematic review, traditional performance metrics—such as Load Time, Page Size, and Response Time—were the most widely adopted, appearing in more than 70% of the studies. Other commonly reported indicators included First Byte, Number of Requests, and Document Complete Time, particularly in earlier studies focused on network-level optimization.

By contrast, modern, user-centric indicators such as Time to Interactive, Largest Contentful Paint, and Speed Index were more prominent in recent publications, especially those aligned with Google Lighthouse and Core Web Vitals.

On the other hand, indicators like Markup Validation, Design Optimization, and Compression appeared less frequently, likely due to challenges in automating their measurement or the lack of universally accepted standards.

Figure 11 illustrates the distribution of the 16 KPIs across the reviewed studies. This visualization highlights a clear imbalance in metric adoption and underscores the need for more holistic frameworks that extend beyond load speed to also address accessibility, maintainability, and user experience.

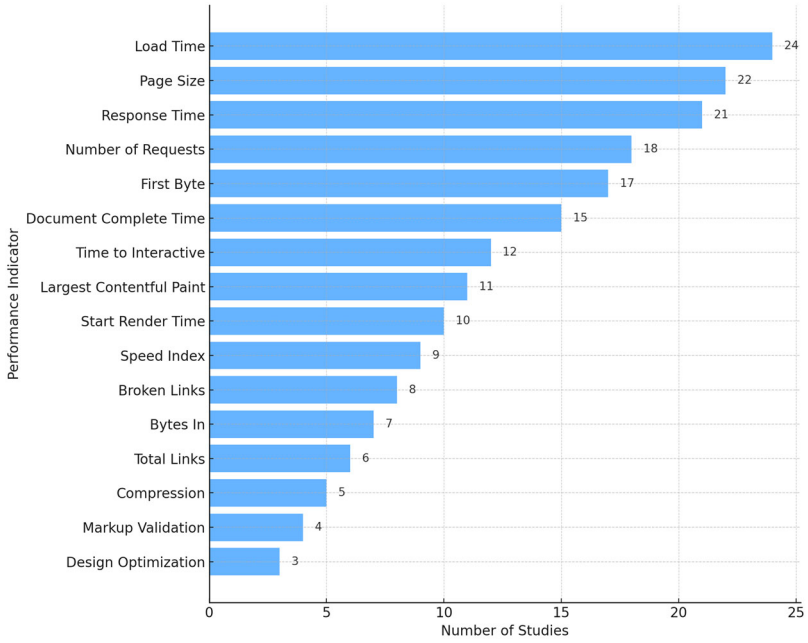

Figure 11. Frequency of adoption of the 16 validated KPIs across reviewed studies, directly supporting RQ4 on the aspects affecting website performance.

4.9. Technological Advances and Development of Performance Criterion

Over the past decade, website performance evaluation has evolved significantly in response to rapid technological changes. Early studies primarily relied on traditional technical metrics such as Load Time, Page Size, and Response Time, often using custom scripts or browser-based tools for measurement.

However, recent years have seen a shift toward more user-centric and standardized indicators, driven by tools like Google Lighthouse and Core Web Vitals. Metrics such as Largest Contentful Paint (LCP), Time to First Byte (TTFB), and Document complete time have emerged as key components of performance assessment, focusing on perceived speed and usability rather than raw loading efficiency.

Additionally, the increasing integration of machine learning techniques has contributed to the development of intelligent evaluation systems that adapt to various performance scenarios, user behaviors, and platform types. These shifts reflect a

broader trend from static, one-size-fits-all models toward context-aware and dynamic evaluation frameworks.

Despite these advancements, there remains a lack of consistency in applying modern performance criteria across research studies, and many still rely solely on traditional indicators. This underscores the need for updated, standardized evaluation protocols that incorporate both technical efficiency and user experience dimensions.

The synthesis of these findings motivates a deeper discussion of their implications and practical significance.

5. Discussion and Recommendation

5.1. Discussions

This review systematically examined 30 studies published between 2010 and 2024, aiming of identifying trends, quality indicators, and methods used in website performance evaluation. The analysis revealed several important insights regarding the evolution and current state of the field.

First, while the number of studies has increased significantly since 2015, there is a notable concentration in specific countries (e.g., India, Turkey, China), with few contributions from regions such as North America and Africa. This highlights the need for greater geographical diversity and global benchmarking in website performance research.

Second, the types of websites analyzed remain heavily skewed toward government and educational platforms. This may be due to the public accessibility of such websites, but it also creates a gap in understanding the performance of commercial, financial, and entertainment platforms, which are increasingly critical in today's digital ecosystem.

Third, the results show that although numerous quality factors have been proposed in the literature, a small subset of performance metrics primarily load time, page size, and response time are repeatedly used. User-centric indicators such as time to interactive, largest contentful paint, and Document complete time are gaining traction but are still underutilized.

Fourth, machine learning techniques have begun to be applied in performance prediction, yet their use is still limited. Most studies continue to rely on rule-based, statistical, or manual scoring systems. This reflects both a technological gap and restricted access to robust datasets needed to train predictive models.

Robustness of Key Metrics. Among the validated 16 metrics, indicators such as load time, time to first byte, and page size proved to be the most robust across domains, consistently influencing user experience and technical efficiency. By contrast, factors like design optimization and markup validation appeared less reliable due to variability in implementation practices and lack of standardized measurement protocols. This imbalance underscores the need to prioritize universally applicable metrics while refining context-sensitive ones.

Interactions among Validated KPIs. While the 16 validated KPIs provide distinct perspectives on website performance, several of them interact or correlate in real-world scenarios. For instance, Load Time is directly influenced by Page Size and Number of Requests [21,23,24], as larger assets and excessive requests prolong overall page loading. Similarly, Time to First Byte (TTFB) is closely tied to Response Time, both reflecting server efficiency and network latency [18,25]. User-centric indicators such as Largest Contentful Paint (LCP) and Start Render Time are strongly affected by front-end optimization practices (e.g., compression, design optimization) [22,26,27], demonstrating that visual performance often depends on underlying technical efficiencies. Moreover, structural quality metrics such as Markup Validation and Broken Link Detection indirectly affect user experience and SEO [6,28,29], reinforcing the interconnected nature of performance, usability, and

accessibility. Acknowledging these interdependencies is critical for practitioners, as improvements in one KPI (e.g., reducing page size) can simultaneously enhance multiple others (e.g., faster load time, better LCP).

Temporal Relevance. Although the review systematically covered the period 2010–2024, the majority of the 30 included studies were published before 2020. These earlier works provided the foundational definitions, metrics, and evaluation frameworks that shaped subsequent research in web performance. More recent studies (2020–2024), however, emphasize modern indicators such as Core Web Vitals (e.g., LCP, TTI, TTFB) and explore the integration of machine learning and deep learning approaches for predictive evaluation. This combination ensures that the review reflects both the historical evolution of methodologies and their current relevance to contemporary web technologies.

Limited Use of ML/DL. Despite the proven potential of machine learning and deep learning in predictive analytics, their adoption in website performance studies remains limited. Several interrelated barriers explain this phenomenon. First, there is a shortage of publicly available benchmark datasets, which constrains model training and cross domain validation [18,21]. Second, the complexity of model development, including hyper-parameter tuning and computational costs, restricts their practical applicability in large-scale studies [20,22]. Third, the interpretability challenge of black-box models makes it difficult for practitioners to trust and integrate such approaches in real world settings [30,31]. Consequently, most existing works continue to rely on rule-based, heuristic, or statistical methods, which, although less adaptive, offer greater simplicity and transparency.

Challenges of Standardization. Another challenge lies in the lack of cross-domain benchmarks. Current evaluations are often confined to government or educational websites, leaving sectors such as finance, healthcare, and entertainment underrepresented. Without broader datasets and internationally agreed-upon standards, it is difficult to generalize findings across industries. This gap highlights the urgency of developing shared frameworks and cross-domain repositories to ensure more consistent and scalable evaluation practices.

This reliance on traditional methods highlights a fundamental gap in the field: although user-centric indicators and intelligent prediction models are available, their adoption remains minimal due to data scarcity, lack of standardized benchmarks, and limited cross-domain validation. This suggests that much of the current literature addresses symptoms of performance issues rather than developing scalable, predictive frameworks.

Collectively, these findings emphasize the need for more comprehensive, automated, and user-centered evaluation frameworks. There is also a pressing need to consolidate performance indicators across domains and to encourage the use of intelligent systems that can adapt to different web contexts and user profiles.

Moreover, by providing clear operational definitions and measurement methods for the 16 final KPIs (see Section 3.3 and Table 8), this study improves the transparency and replicability of the evaluation framework. Such clarification ensures that each indicator is not only conceptually identified but also practically measurable using standardized tools (e.g., Google Lighthouse, WebPageTest). This step bridges the gap between theoretical selection of performance factors and their real-world applicability, thereby increasing the practical value of the proposed framework for both researchers and practitioners.

An additional strength of this study lies in the structured reduction process that refined the initial pool of 223 candidate metrics into 59 and subsequently into 16 expert-validated KPIs. By explicitly applying duplicate removal, synonym consolidation, operationalize ability checks, and consensus rounds, the study ensured both transparency and replicability in metric selection. This systematic distinguishes differentiates our review from prior surveys, which often presented fragmented or domain-specific metric lists with-

out a reproducible methodology. As a result, the final set of KPIs can be considered not only comprehensive but also robust and reliable, reinforcing the practical value of the proposed framework.

Limitations of Study

Although this systematic review provides a comprehensive overview of website performance evaluation, it is not without limitations. First, the selection of studies was limited to seven databases and English-only publications, which may have excluded relevant research in other languages or grey/non-indexed literature. Second, while the quality assessment followed a rigorous checklist, the scoring process still involved subjective judgments that could influence inclusion decisions. Third, the study focused on synthesizing reported metrics and methodologies without performing a quantitative meta-analysis, which may limit the generalizability of certain findings. Acknowledging these limitations provides transparency and sets directions for more inclusive future reviews.

Additionally, a significant number of potentially relevant studies (approximately 85%) were excluded due to the lack of full-text access. While this exclusion ensured methodological transparency and rigorous data extraction, it also introduces a risk of selection bias, potentially limiting the comprehensiveness of the review. Future research should adopt broader access strategies, such as institutional subscriptions or direct author requests, to minimize this limitation and enhance coverage of the literature. This limitation highlights the broader challenge of accessibility in systematic reviews, particularly in web performance research, and underscores the need for open access publishing to mitigate potential bias.

5.2. Recommendations for Practitioners

Based on the systematic review of 30 studies and the refined set of 16 quality indicators, several practical recommendations can be made for web developers, performance analysts, and digital experience teams. Table 11 can be regarded as a practical roadmap for practitioners, since it organizes the findings into actionable recommendations across performance dimensions. This enhances the applicability of the review by bridging academic insights with real-world performance engineering practices.

Table 11. Summary of practical recommendations by performance area. (Answers RQ4).

Area	Recommendations	References
Core Web Vitals (CWVs)	- Improve LCP: Optimize image sizes, minify and combine resources, utilize browser caching, and consider lazy loading.	[32,33]
	- Enhance FID: Minimize JavaScript execution, prioritize critical JS, avoid render-blocking resources.	[33,34]
	- Minimize CLS: Use fixed dimensions for images and videos, and avoid third-party layouts, and pre-load content.	[13,35,36]
Content & Design	- Compress images: Use efficient formats (WebP), and optimize sizes without quality loss.	[37,38]
	- Minify and combine resources: Reduce HTTP requests, minify HTML/CSS/JS, and combine when possible.	[6,39,40]
	- Implement lazy loading: Load non-critical elements only when needed, improve initial page load.	[21,23]

Table 11. *Cont.*

Area	Recommendations	References
Browser Caching	- Enable caching for static assets: Set appropriate headers for local storage, reduce load and improve experience. - Consider CDN: Distribute content across servers, reduce latency, and improve global performance. -Optimized server response times are crucial for efficient performance.	[41,42] [21,23,41]
Mobile Responsiveness	- Use responsive design: Ensure seamless adaptation to different screen sizes and devices. - Test for mobile usability: Use tools like Google's Mobile-Friendly Test to identify and fix issues.	[17,28,33,38,42]
Monitoring & Analysis	- Use website analytics: Track key metrics (page load, bounce rate, conversion) to identify improvement areas. - Conduct regular performance audits: Use tools like Google PageSpeed Insights and Lighthouse to detect technical issues and optimization opportunities.	[19,33,39]

These suggestions are grouped by performance dimension and informed by both the frequency of use in prior studies and their relevance to modern web evaluation. Table 11 provides a categorized summary of these recommendations, organized by performance areas such as Core Web Vitals, content design, browser caching, and mobile responsiveness. The table also includes supporting references to aid implementation and guide further research.

1. Adopt Core Performance Metrics Early in Development
    - Prioritize foundational metrics such as Load Time, Time to First Byte, and Page Size, as they directly impact user experience and are supported by nearly all performance testing tools.
    - Tools such as Google PageSpeed Insights and WebPageTest be used to help continuously monitor these metrics throughout development.
2. Implement User-Centric Indicators
    - Incorporate Largest Contentful Paint (LCP), and Time to Interactive (TTI) for a more realistic evaluation of perceived performance.
    - These should be particularly emphasized in dynamic, content-heavy websites.
3. Optimize Design and Front-End Assets
    - Reduce the number of requests and overall page weight through efficient asset management.
    - Apply compression techniques, lazy loading, and minified JavaScript/CSS to improve rendering times.
4. Ensure Code and Accessibility Quality
    - Regularly validate HTML structure using tools such as W3C Markup Validator to detect errors and improve maintainability.
    - Check for broken links, missing ALT tags, and other accessibility issues that degrade both SEO and usability.
5. Apply Predictive and Intelligent Tools Where Possible
    - Use machine learning-based evaluations to predict site performance, especially in high-traffic applications where small inefficiencies can scale.
    - Integrate automated performance testing pipelines into CI/CD environments.

Guidelines for model selection and parameterization (Answers RQ4).

For SVM, standardize features and prefer the RBF kernel as a default; tune C and γ via grid or Bayesian search over logarithmic ranges (e.g., $C \in [10^{-2}, 10^{3}]$, $\gamma \in [10^{-4}, 10^{0}]$) using 5–10-fold cross-validation.

For Random Forest, start with n_estimators $\geq$ 300, tune max_depth (e.g., 6–20) and min_samples_leaf (1–5) while monitoring out-of-bag error; use permutation importance for interpretability.

For Gradient-Boosted Trees/XGBoost, adopt learning_rate 0.05–0.2 with early stopping (validation patience 20–50 rounds), tune n_estimators (200–800), max_depth (4–8), and subsample/colsample (0.7–1.0).

For Logistic Regression, standardize inputs and select L2 regularization with C tuned on a log scale; report calibrated probabilities when thresholding KPIs.

For KNN, scale features, select k in 3–15 via cross-validation, and prefer distance weighting when class imbalance exists.

Across models, adopt nested CV for fair model comparison, stratify splits when classes are imbalanced, and report both threshold-free (AUC) and threshold-based metrics (F1, accuracy) for classification or MAE/RMSE for regression.

When deployment interpretability is required, complement ensembles with SHAP-based post-hoc explanations to relate predictions to specific KPIs.

To complement these guidelines, Table 12 summarizes the recommended hyper-parameter ranges and best practices for major machine learning algorithms used in website performance prediction.

Table 12. Recommended parameter configurations for major ML algorithms in web performance prediction (Answers RQ4).

Algorithm	Key Parameters and Ranges	Notes/Best Practices	Ref
SVM	$C \in [10^{-2}, 10^{3}]$, $\gamma \in [10^{-4}, 10^{0}]$ (log scale)	Standardize features; prefer RBF kernel as default; tune via grid/Bayesian search.	[18,22]
Random Forest	n_estimators $\geq$ 300; max_depth = 6–20; min_samples_leaf = 1–5	Monitor out-of-bag error; use permutation importance for feature interpretability.	[20,39]
Gradient-Boosted Trees/XGBoost	learning_rate = 0.05–0.2; n_estimators = 200–800; max_depth = 4–8; subsample/colsample = 0.7–1.0	Apply early stopping (20–50 rounds patience); balance bias/variance with tuned depth.	[18,26]
Logistic Regression	Regularization: L2; C tuned on log scale	Standardize inputs; report calibrated probabilities for thresholding KPIs.	[6,40]
KNN	k = 3–15; weighting = distance-based	Scale features; cross-validate k; prefer distance weighting under class imbalance.	[16,29]
All models	Validation: Nested CV; Metrics: AUC, F1, Accuracy, MAE/RMSE	Stratify splits if imbalance exists; use SHAP for interpretability in deployment.	[15,22,43]

These recommendations aim to bridge the gap between academic research and real-world performance engineering. By applying these practices, web teams can not only meet technical performance standards but also deliver improved user experiences across various platforms.

6. Conclusions

This study conducted a systematic review of research conducted between 2010 and 2024 on website performance evaluation. Drawing on 30 high-quality studies, we identified recurring quality factors, assessment methods, predictive approaches, and key performance indicators (KPIs) used across different domains and timeframes.

The review highlighted a set of 16 critical quality metrics that have been consistently used or recommended in the literature. It also revealed that while traditional metrics (e.g., load time, page size, and response time) remain dominant, there is a growing shift toward more user-centric indicators (e.g., LCP, TTI, and TTFB), especially in more recent studies.

In terms of methodology, most studies still rely on descriptive or rule-based evaluations. However, some have begun to incorporate machine learning techniques for performance prediction, though this remains an underdeveloped area with significant potential for growth.

Furthermore, the review identified several research gaps: a limited geographic distribution of studies, underrepresentation of certain domains (e.g., e-commerce, healthcare, and entertainment), and lack of standardized performance benchmarks. These gaps represent clear opportunities for future investigation.

Ultimately, this study contributes to the field by synthesizing recent research trends and tools, providing a validated list of key performance metrics, offering recommendations for practitioners, and highlighting future research directions in web performance evaluation.

By bridging the gap between theory and practice, this work can support both academics and professionals in enhancing the efficiency, usability, and reliability of modern web applications.

Building on the findings of this study, future research will aim to develop a comprehensive model for website performance evaluation. Specifically, a dataset will be constructed based on the 16 validated performance metrics identified in this review. This dataset will be used to train and evaluate machine learning and deep learning models. The goal is to develop an accurate, scalable, and adaptive model capable of predicting website performance across different domains. Such a model could assist researchers and practitioners in diagnosing performance issues and recommending targeted improvements.

By addressing the four research questions, this review moves beyond descriptive statistics to provide actionable insights on predictive approaches, their strengths and limitations, and their applicability across different domains. The consolidated set of 16 KPIs and qualitative method comparison lay the groundwork for reliable and domain-independent models for predicting website performance. Building on these findings, our planned future work will focus on constructing a dataset derived from the validated KPIs and leveraging frameworks such as Core Web Vitals. We also envision "WebPulse AI" as a practical application that operationalizes these insights into a real-time tool for website performance prediction.

Nevertheless, the present review should be interpreted within its methodological constraints. Future studies are encouraged to expand the database coverage, include non-English sources, and employ quantitative meta-analysis to strengthen evidence synthesis. Addressing these aspects would ensure broader generalizability and stronger empirical grounding for subsequent predictive models.

Author Contributions: Conceptualization, M.G. and S.O.; methodology, M.G.; software, M.G.; validation, M.G., S.O. and A.M.M.; formal analysis, M.G.; investigation, M.G.; resources, S.O. and A.M.M.; data curation, M.G.; writing—original draft preparation, M.G.; writing—review and editing, S.O. and A.M.M.; visualization, M.G.; supervision, S.O. and A.M.M.; project administration, S.O.; funding acquisition, S.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially funded by the Spanish Ministry of Science, Innovation and Universities MICIU/AEI/10.13039/501100011033 (https://www.ciencia.gob.es/en/Convocatorias.html, accessed on 10 October 2025) under project/grant PID2023-147409NB-C21, and by ERDF, EU. It has also been funded by the European UnionNextGenerationEU/PRTR (https://next-generation-eu.europa.eu/index_en, accessed on 10 October 2025), under projects/grants TED2021-131699B-I00 and TED2021-129938B-I00.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors would like to express their sincere gratitude to the Deanship of Graduate Studies and Scientific Research at Bethlehem University for their continued support and encouragement. This research was conducted with the support of Research Group #RG-BU006, whose guidance and collaboration have been instrumental in advancing this work. Their contributions are deeply appreciated.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Quality Assessment

Table A1. Quality assessment results of the reviewed studies based on evaluation questions (Q1–Q9).

Studies	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Sum
P1	Y	Y	Y	P	P	P	Y	P	P	6.5
P2	Y	Y	Y	P	P	P	Y	P	P	6.5
P3	Y	Y	Y	P	P	P	Y	Y	Y	7.5
P4	Y	P	Y	N	N	N	Y	P	P	4.5
P5	Y	Y	P	P	P	P	Y	P	Y	6.5
P6	P	Y	Y	P	Y	Y	P	P	Y	7.5
P7	P	Y	Y	P	P	P	Y	P	P	6
P8	P	Y	Y	Y	P	P	Y	P	P	6.5
P9	Y	Y	Y	Y	Y	Y	Y	Y	Y	9
P10	Y	Y	P	P	P	P	Y	P	P	6
P11	Y	Y	Y	P	Y	P	Y	Y	N	7
P12	Y	Y	P	Y	P	P	P	P	N	5.5
P13	N	Y	Y	P	N	N	P	P	Y	4.5
P14	P	P	P	P	P	N	Y	P	P	4.5
P15	P	Y	Y	P	P	N	P	P	N	4.5
P16	Y	P	Y	N	N	N	Y	Y	Y	5.5
P17	Y	Y	Y	N	P	Y	P	Y	N	6
P18	P	Y	P	P	P	P	Y	Y	N	5.5
P19	Y	Y	Y	N	P	P	Y	Y	N	6
P20	Y	Y	Y	P	N	N	P	N	P	4.5
P21	P	Y	P	P	P	P	Y	P	P	5.5
P22	Y	N	Y	N	N	N	Y	P	Y	4.5
P23	Y	P	Y	P	P	P	Y	Y	P	6.5

Table A1. *Cont.*

Studies	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Sum
P24	Y	N	Y	P	P	P	Y	P	P	5.5
P25	Y	Y	Y	P	Y	P	Y	P	Y	7.5
P26	Y	Y	Y	Y	Y	P	Y	Y	N	7.5
P27	Y	P	P	P	P	P	Y	Y	N	5.5
P28	Y	P	P	P	P	P	P	P	P	5
P29	Y	Y	Y	P	P	P	P	P	P	6
P30	P	Y	P	P	P	P	P	Y	N	5
P31	Y	N	Y	P	P	P	P	Y	N	5
P32	Y	Y	Y	N	P	P	P	Y	Y	4.5
P33	Y	Y	P	N	P	P	P	P	P	5
P34	P	Y	P	P	N	P	P	P	Y	5

Appendix B. Studies

Table A2 provides details of the 34 studies that initially met the inclusion/exclusion criteria. Four of these were excluded during quality assessment, leaving 30 studies in the final synthesis.

Table A2. Summary of studies meeting the inclusion and exclusion criteria.

Paper ID	Title	Journal	Author	Country	Year	Context	Ref
P1	Web Application Performance Analysis of E-Commerce Sites in Bangladesh: An Empirical Study	Modern Education and Computer Science Press(MECS Press)	Mahfida Amjad, Md. Tutul Hossain, Rakib Hassan, Md. Abdur Rahman	Bangladesh	2021	e-commerce sites	[23]
P2	Evaluating the performance of government websites: An automatic assessment system based on the TFN-AHP methodology	Journal of Information Science	Xudong Cai, Shengli Li, Gengzhong Feng	China	2020	e-government	[8]
P3	The Performance Evaluation of a Website using Automated Evaluation Tools	Technology Innovation Management and Engineering Science International	Achaporn Kwangsawad; Aungkana Jattamart; Paingruthai Nusawat	Thailand	2019	herbal cosmetic	[33]
P4	Performance evaluation of websites using entropy and grey relational analysis methods: The case of airline companies	Decision Science Letters	Kemal Vatansever, Yakup Akgűl	Turkey	2018	airlines	[26]
P5	An Intelligent Method to Assess Webpage Quality using Extreme Learning Machine	International Journal of Computer Science and Network Security	Jayanthi, B., Krishnakumari, P	India	2016	education, finance, news and health	[32]
P6	Analytic Hierarchy Process (AHP) Based Model for Assessing Performance Quality of Library Websites	Information Technology Journal	Harshan, R. K., Chen, X., and Shi, B.	China	2017	library	[25]

Table A2. *Cont.*

Paper ID	Title	Journal	Author	Country	Year	Context	Ref
P7	An empirical performance evaluation of universities website	International Journal of Computer Applications	KAUR, Sukhpuneet; KAUR, Kulwant; KAUR, Parminder	India	2016	education	[24]
P8	Predicting web page performance level based on web page characteristics	International Journal of Web Engineering and Technology	Junzan Zhou, Yun Zhang, Bo Zhou and Shanping Li	China	2015	education	[21]
P9	Measuring Quality of Asian Airline Websites Using Analytical Hierarchy Process: A Future Customer Satisfaction Approach	Information Systems International	Humera Khan, P.D.D.Dominic	Malaysia	2013	airline	[29]
P10	A comparison of Asian e-government websites quality: using a non-parametric test	International Journal of Business Information Systems	P.D.D. Dominic and Handaru Jati	Malaysia	2011	e-government	[44]
P11	Quality Ranking of E-Government Websites: PROMETHEE II Approach	International Conference for Informatics for Development	Handaru Jati	Indonesia	2011	e-government	[28]
P12	Evaluation of Usage of University Websites in Bangladesh	DESIDOC Journal of Library & Information Technology	ISLAM, Anwarul; TSUJI, Keita	Bangladesh	2011	university	[40]
P13	Measuring the quality of e-commerce websites using analytical hierarchy process	TELKOMNIKA (Telecommunica-tion Computing Electronics and Control)	Aziz, U. A., Wibisono, A., and Nisafani	Indonesia	2019	e-commerce	[27]
P14	Measuring website quality of the Indian railways	International Journal of Entrepreneurial Knowledge	Jain, R. K., and Rangnekar	Indian	2015	railways	[45]
P15	Evaluation of Nigeria Universities Websites Quality: A Comparative Analysis	Library Philosophy and Practice	Sunday Adewale Olaleye, Ismaila Temitayo Sanusi, Dandison C. Ukpabi, Adekunle Okunoye	Nigeria	2018	university	[46]
P16	A comparative approach to web evaluation and website evaluation methods	International Journal of Public Information Systems	Zahran, D. I., Al-Nuaim, H. A., Rutter, M. J., and Benyon, D	Scotland, UK	2014	government	[7]
P17	A comparison of Asian airlines websites quality: using a non-parametric test	International Journal of Business Innovation and Research	Dominic, P. D. D., and Jati, H	Malaysia	2011	airline	[47]
P18	A filter-wrapper based feature selection for optimized website quality prediction	Amity International Conference on Artificial Intelligence (AICAI)	Akshi Kumar, Anshika Arora	India	2019	commercial, organization, government	[22]

Table A2. *Cont.*

Paper ID	Title	Journal	Author	Country	Year	Context	Ref
P19	A neuro-fuzzy classifier for website quality prediction	International Conference on Advances in Computing, Communications and Informatics	Malhotra, R., and Sharma, A	India	2013	NA	[30]
P20	A Novel Model for Assessing e-Government Websites Using Hybrid Fuzzy Decision-Making Methods	International Journal of Computational Intelligence Systems	Shayganmehr, M., and Montazer, G. A	Iran	2021	e-government	[48]
P21	A proposal for a quality model for e-government website	International Conference on Data and Software Engineering (ICoDSE)	HENDRADJAYA, Bayu; PRAPTINI, Rina	Indonesia	2015	government	[49]
P22	Performance Evaluation of Websites Using Machine Learning	EIMJ	MM Ghattas, PDB Sartawi	Palestine	2020	NA	[18]
P23	Analysis and modelling of websites quality using fuzzy technique	Second International Conference on Advanced Computing & Communication Technologies	MITTAL, Harish; SHARMA, Monika; MITTAL, J. P	India	2012	NA	[50]
P24	Analytic hierarchy process for website evaluation	Intelligent Decision Technologies	KABASSI, Katerina	Greece	2018	government, health	[51]
P25	Application of mathematical simulation methods for evaluating the websites effectiveness	Systems of Signals Generating and Processing in the Field of on-Board Communications	Erokhin, A. G., Vanina, M. F., and Frolova, E. A	Russia	2019	e-commerce	[52]
P26	Empirical validation of website quality using statistical and machine learning methods.	International Conference-Confluence the Next Generation Information Technology Summit (Confluence)	Poonam Dhiman, Anjali	India	2014	NA	[20]
P27	Evaluating the Websites' Quality of Five- and Four-Star Hotels in Egypt	Minia Journal of Tourism and Hospitality Research MJTHR	Elsater, S. A. E., Dawood, A. E. A. A., Mohamed Hussein, M. M., and Ali, M. A	Egypt	2022	hotel	[16]
P28	A review of website evaluation using web diagnostic tools and data envelopment analysis	Bulletin of Electrical Engineering and Informatics	Najadat, H., Al-Badarneh, A., and Alodibat	Jordan	2021	e-government	[6]
P29	Empirical and Automated Analysis of Web Applications	International Journal of Computer Applications	KULKARNI, R. B.; DIXIT, S. K	India	2012	e-commerce, banking, and e-governance	[53]

Table A2. *Cont.*

Paper ID	Title	Journal	Author	Country	Year	Context	Ref
P30	Website Performance Analysis and Evaluation using Automated Tools.	International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques	Kumar, N., Kumar, S., and Rajak, R	India	2021	organization	[39]
P31	Framework for evaluation of academic website	International Journal of Computer Techniques	Devi, K., and Sharma, A	India	2016	academic	[35]
P32	Brief analysis on Website performance evaluation	IET Digital Library	Li Peng; YueMing Lu; Dongbin Wang	China	2015	NA	[54]
P33	Web page prediction using genetic algorithm and logistic regression based on weblog and web content features	International Conference on Electronics and Sustainable Communication Systems	Gangurde, R., and Kumar	India	2020	organization	[31]
P34	Performance Testing and Optimization of DiTenun Website	Journal of Applied Science, Engineering, Technology, and Education	Barus, A. C., Sinambela, E. S., Purba, I., Simatupang, J., Marpaung, M., and Pandjaitan, N.	Indonesia	2022	industry	[55]

Appendix C. Metrics

Table A3. Description of the 59 website performance metrics identified in the reviewed studies.

No	Name	Description
1	fully loaded (requests)	This is the quantity of requests that the browser has to make for pieces of material on the page (images, JavaScript, CSS, etc.). It is an information request message sent from a client to a server using the hypertext transfer protocol (HTTP). To transmit images, text, or pages to the user's browser, it must first request that data, which it does via an HTTP request.
2	first CPU idle	First CPU Idle measures when a page is minimally interactive, or when the window is quiet enough to handle user input.
3	speed index	A metric that measures how quickly the contents of a webpage are visually displayed during loading.
4	start render	The time when the first non-white content is painted on the screen, indicating the beginning of the webpage rendering process
5	load time	The time it takes for a webpage to load completely, including all resources and rendering
6	mobile optimization	The optimization of the website for mobile devices, including responsive design, mobile-friendly layouts, and fast loading times on mobile networks, to enhance user experience for mobile users
7	document complete (time)	Indicates the point at which the browser's on load event appears, indicating that all of the static page content has been loaded to some extent.
8	last painted hero	Function as an artificial indicator, shows the user when the final critical piece of content is visually rendered on the screen.

Table A3. *Cont.*

No	Name	Description
9	first content-full paint	The time when the first content element (such as text or images) is rendered on the screen.
10	first byte	Measures the time between when an internet user makes an HTTP request, such as loading a webpage, and when the client's browser receives the first byte of data.
11	bytes in	The amount of data that the browser needs download in order to fully load a webpage.
12	time to interactive	Denoting the moment when the last prolonged task concludes, succeeded by 5 s of network and main thread dormancy. TTI offers users a comprehensive understanding of the website's responsiveness from the perspective of the site visitor.
13	max potential first input delay	The period between when a user first interacts with your site, such as clicking a button, and when the browser is fully ready to respond to that interaction.
14	first meaningful paint	The first significant paint is the amount of time it takes for the main material of a page to appear on the screen. Although it frequently identifies non-meaningful paints like headers and navigation bars, it is utilized as an approximation of the first meaningful paint.
15	largest content-full paint	Largest content-full paint (LCP) is a crucial, user-centric metric for measuring perceived load speed as it marks the page load timeline when the page's main content has been loaded.
16	cumulative layout shift	A metric that measures the amount of unexpected layout shifts that occur during the loading process, affecting visual stability and user experience
17	first input delay (FID)	Measures the delay between the user's first interaction (like clicking a link or button) and the time the browser begins processing that interaction. It reflects real interactivity performance and responsiveness from the user's perspective.
18	availability of hyperlinks	Verifies if users of websites can access the pages without any problems.
19	updatability of information	Refers to the information updates on websites. It is measured by the percentage of updated hyperlinks on websites during the assessment cycle.
20	richness of content	Determines if webpages have a variety of information resources.
21	security of website	Verify whether websites are protected from Trojans by robust security measures.
22	impacts on search engines	Refers to the performance of websites on search engines.
23	impacts on social media	Checks the influence of websites on the social media.
24	impacts on network	Measures the effect of websites on the Internet by reflecting their popularity and importance through the use of tools like PageRank.
25	page size	Refers to a specific web page's overall size. Every file that makes up a webpage is included in the page size. The HTML document, any images that are included, style sheets, scripts, and other material are all included in these folders.
26	page requests	A request for a web page, in its whole or in part (including requests for additional frames), results from user actions such typing a URL, clicking a link, sending out a "refresh" command, or moving across the page.
27	browser cache	A temporary storage area in memory or on disk that holds the most recently downloaded Web pages.
28	page redirects	Page redirects add to the loading cycle, increasing the time to display a page.
29	compression	JavaScript and CSS ensure proper compressed this makes the website run much faster.
30	render-blocking JavaScript	Render-blocking resources are portions of code in website files, usually CSS and JavaScript, that prevent a web page from loading quickly.

Table A3. *Cont.*

No	Name	Description
31	traffic	The browser gathers data, which is then transmitted to the Alexa website. At the website, this collected data is stored and analyzed, forming the foundation for the company's web traffic reporting.
32	page rank	It is used to calculate and display the PageRank for each Website.
33	browser compatibility	Ensuring that the website is compatible with different web browsers and devices, optimizing performance and user experience across various platforms.
34	content delivery network	The use of CDN services to distribute webpage content across multiple servers located geographically closer to users, improving load times and reducing latency.
35	response time	A website server is expected to respond to a browser request within specific parameters.
36	markup validation	It is employed to evaluate and compute the quantity of HTML errors present on the website, including orphan codes, coding errors, missing tags, and similar issues.
37	broken links	Links on websites might be internal or external. When a visitor clicks on a link, they trust that the page will load successfully.
38	total link	Total Link on webpage.
39	text link	Total Text Link.
40	word count	Total words on page.
41	total body words	Number of words in sentence.
42	total sentence	Number of sentences in paragraph.
43	total paragraph	Number of paragraphs in body text.
44	total cluster count	Number of text cluster on page.
45	total image	Total Image on page.
46	alt image count	Number of images with ALT clause.
47	no alt image count	Number of images without ALT clause.
48	animation count	Number animated element.
49	unique image count	Number of unique images.
50	image map count	Number of image maps on page.
51	Un-sized image count	Number of images without size definition.
52	total color	Total color on page.
53	reading complexity	Overall Page Readability.
54	number of components	The amount of request/response between a client and a host
55	design optimization	The scripts, HTML, or CSS codes are optimized to enhance loading speed. This optimization concurrently reduces the quantity of website elements, including images, scripts, HTML, CSS codes, or videos.
56	availability	Is a website that is accessible.
57	the frequency of update	Check frequently website is updated with new content.
58	html page sizes	The size of all the HTML code on your web page—this size does not include images, external JavaScript's or external CSS files.
59	download time	The average time to download any page related to the services, including all content contained therein.

Appendix D. Data Extraction

Table A4. Summary of factors examined and algorithms/approaches used in the reviewed studies.

Paper ID	Factors Examined	Algorithms/Approaches
P1	fully loaded (requests), first CPU idle, speed index, start render, load time, fully loaded (time), document complete (time), last painted hero, first contentful paint, and first byte	automated evaluation tools
P2	availability of hyperlinks, updatability of information, loading speed of web pages, richness of content, security of website, construction of columns, impacts on search engines	TFNs and AHP
P3	page size, page requests, pages peed, browser cache, page redirects, compression, and render-blocking JavaScript	automated evaluation tools
P4	Traffic, page rank, design optimization, load time, response time, markup, and broken links	Entropy and Grey Relational Analysis
P5	Total Link, Text Link, Word Count, Total Body Words, Total Sentence, Total paragraph, Total cluster count, Total Image, Alt Image Count, Unique Image Count, Image map Count, Unsized Image count, Total Color, Reading Complexity	Extreme Learning Machine (ELM), SVM
P6	Load time, number of components, page speed, page size, response time, mark-up validation, broken links, and design optimization	AHP and FAHP
P7	No. of Requests, Load time and Page size	automated evaluation tools
P8	Number of servers contacted, Number of origins contacted, Number of object requests median, Object request size median, Number of JavaScript objects median, Size of JavaScript objects median, Number of image objects median, Size of image objects median, Number of flash objects median, Size of flash objects median, Number of CSS objects median, Size of CSS objects median, Maximum size of objects normalized median	RF, AdaBoost, Logistic Regression, SVM, NB, BN
P9	Load time page size, response time, page speed, availability, broken links, no. of component, response time, markup validation	Analytical Hierarchy Process
P10	Load time, response time, page rank, the frequency of update, traffic, design optimization, page size, number of the item, accessibility error, markup validation	LWM, AHP, FAHP, NHM
P11	Load time, response time, page rank, the frequency of update, traffic, design optimization, size, no of items, accessibility error, markup validation, and broken link	PROMETHEE II and AHP
P12	Total no of HTML files, HTML page sizes, composition, total number of images, and download time	web diagnostic tools
P13	Load Time, Page Size, Number of Item, Page Speed Score, Availability, Page Rank, Traffic, Design Optimization, Markup Validation	AHP
P14	Continuous Connectivity, Quick Response, Ease of Access, Options to Pay, Content Usefulness, Ease of Navigation, Clarity of Data, Privacy and Security, Aesthetics, Customization	Statistical tools (ANOVA)
P15	ease of use, processing speed, aesthetic design, interactive responsiveness, entertainment, trust and usefulness	web analytical tools
P16	Usability, maintainability, reliability, efficiency, navigation, content	web analytical tools

Table A4. *Cont.*

Paper ID	Factors Examined	Algorithms/Approaches
P17	Load time, response time, page rank, frequency of update, traffic, design optimization, size, number of items, accessibility error, broken links	LWM, AHP, FAHP
P18	Relevance, Updating, Accuracy, Total size, Broken Links, Loading Time, Communication, Social Media Connectivity, Browser Compatibility, Typography & Font, Color Scheme, Overall Theme	NB, KNN, DT, RF
P19	Word Count, Body Text Words, Page Size, Table Count, Graphics Count, Division Count, List Count, Number of Links, Page Title length	ANFIS clustering algorithms
P20	Speed of servers' responsiveness, Compatibility with social networks, Document downloading time, Bandwidth, File size, Picture size, Server location, Security, Content quality	Hybrid Fuzzy Decision-Making Methods
P21	Responsiveness, Service Availability, Multi-lingual, Service Accuracy, User Satisfaction, Security, Trust, Information Accuracy, System availability, Access Time, Browser Usage, Usability	automated evaluation tools
P22	Page size, load time, design optimization, markup validation, response time, speed, broken links	Linear regression, SVM
P23	load time, response time, mark-up validation, broken link, accessibility error, size, page rank, frequency of update, traffic and design optimization	Fuzzy logic
P24	Content and appearance, Information quality, Navigability, Graphic design, FAQs, Interactivity, Satisfaction, Usability, Reliability, Privacy, Web Services, Technology, Functionality	AHP, fuzzy AHP
P25	conversion metric, time spent on site, number of refusals, number of pages viewed	mathematical simulation methods
P26	Total words length, Body text length, Title text length, Total links, Internal links, Size of page in KB, Emphasize text, HTML Lines, JS Lines, Complexity, Tables, Graphics	statistical and ML methods
P27	Informational content, Design, Ease of use, Interactivity, Marketing Image, Online processes	Statistical tools
P28	Ambiguity, uncertainty, time, Usefulness, satisfaction, Download time, help features, dynamic content, response time, average page size, hits, visitors	automated evaluation tools
P29	Page load, response time, optimal navigation, HTML, maintainability, security, functionality, usability, efficiency, creditability	automated evaluation tools
P30	User Friendliness, Accessibility, Security, SEO, Social	automated evaluation tools
P31	Usability, Content, Presentation, Functionality, and Reliability	automated evaluation tools
P32	Query DNS, Response to request, Establish connection	automated evaluation tools
P33	Web log	automated evaluation tools
P34	response time and service availability	Logistic Regression (LR)

Appendix E. Research Focus

Table A5. Classification of the reviewed studies according to their main research focus.

Research Topics	Paper ID
Identifying factors influencing website performance	P1,P2,P3,P4,P5,P6,P7,P8,P9,P10,P11,P12,P14,P15,P16,P17,P18,P19, P20,P21,P22,P23,P24,P25,P26,P27,P28,P29,P30,P31,P32,P34
The state-of-the-art in performance evaluation of websites	P2,P4,P6,P9,P10,P11,P14,P17,P19,P20,P23,P24,P25,P28,P32
ML and Deep learning	P5,P8,P18,P19,P26,P33,P22

Appendix F. Full Survey Questionnaire

Title: A Novel Approach for Evaluating Websites' Performance Based on Deep Learning and Optimization Algorithms—Survey

Researcher Introduction:

My name is Mohammad Ghattas, a PhD student at the University of Granada. This survey aims to identify key web attributes that affect website performance from the perspective of experts (developers and webmasters) in the State of Palestine.

Confidentiality Notice:

No identifiable information is collected. Participation is voluntary. Completing and submitting the survey implies consent.

Estimated Duration:

Approximately 30 min.

Participant Demographics Questions:

What is your gender? (Male/Female)

What is your age? (Open-ended or Range Selection: 20–29, 30–39, 40–49, 50+)

What is your country of residence? (e.g., Palestine, Jordan, Egypt, Lebanon, Spain)

What is your current job position? (Junior Developer/Senior Developer/Technical Lead/Researcher/Other)

How many years of experience do you have in web development or performance-related roles? (e.g., 0–2, 3–5, 6–10, 10+)

What is your highest level of education? (Bachelor's/Master's/PhD/Other)

Survey Structure:

The online questionnaire includes 59 web attributes organized in five sections. Participants are asked to rate each attribute on a scale from 1 (Poor) to 3 (Excellent), based on its impact on web performance.

1. First CPU Idle

Measures when a page becomes minimally interactive.

Options: Poor, Excellent

2. Speed Index

Measures how quickly the contents of a page are visibly populated.

Options: Poor, Excellent

3. Traffic

Browser collects data and transmits it for web traffic reporting.

Options: Poor, Excellent

4. PageRank

Used to calculate and display the PageRank for each website.

Options: Poor, Excellent

5. Design Optimization

Optimized scripts, HTML, and CSS for quicker loading.

Options: Poor, Excellent

6. Fully Loaded (Requests)

The quantity of requests the browser makes for pieces of material on the page.
Options: Poor, Excellent
7. Start Render
Time when the first non-white content is painted on the screen.
Options: Poor, Excellent
8. Load Time
Time it takes for a webpage to load completely.
Options: Poor, Excellent
9. Mobile Optimization
Website optimization for mobile devices.
Options: Poor, Excellent
10. Document Complete (Time)
Point at which the browser's onload event appears.
Options: Poor, Excellent
11. Last Painted Hero
Shows the user when the final critical content is visually rendered.
Options: Poor, Excellent
12. First Content-Full Paint
Time when the first content element is rendered on screen.
Options: Poor, Excellent
13. First Byte
Time between an HTTP request and receiving the first byte of data.
Options: Poor, Excellent
14. Bytes In
Amount of data the browser downloads to fully load the page.
Options: Poor, Excellent
15. Time to Interactive
Moment when the last prolonged task ends and responsiveness starts.
Options: Poor, Excellent
16. Max Potential First Input Delay
Time between user interaction and the browser's readiness.
Options: Poor, Excellent
17. First Meaningful Paint
Time for the main material of a page to appear on screen.
Options: Poor, Excellent
18. Largest Content-Full Paint
Measures perceived load speed of the main content.
Options: Poor, Excellent
19. Cumulative Layout Shift
Measures unexpected layout shifts during loading.
Options: Poor, Excellent
20. User Session Duration
Average amount of time a user spends on the site in one session.
Options: Poor, Excellent
21. Availability of Hyperlinks
Checks if users can access pages without problems.
Options: Poor, Excellent
22. Updatability of Information
Percentage of updated hyperlinks during assessment.
Options: Poor, Excellent

23. Richness of Content
Determines if pages have a variety of information.
Options: Poor, Excellent
24. Security of Website
Verifies whether websites are protected from threats.
Options: Poor, Excellent
25. Impacts on Search Engines
Refers to website performance on search engines.
Options: Poor, Excellent
26. Impacts on Social Media
Checks website influence on social media.
Options: Poor, Excellent
27. Impacts on Network
Effect of websites on the Internet's popularity and importance.
Options: Poor, Excellent
28. Page Size
Overall size of a web page including all resources.
Options: Poor, Excellent
29. Page Requests
Requests resulting from user actions like typing URLs or clicking links.
Options: Poor, Excellent
30. Browser Cache
Temporary storage for recently downloaded web pages.
Options: Poor, Excellent
31. Page Redirects
Redirects that increase loading time.
Options: Poor, Excellent
32. Compression
JavaScript and CSS compression to improve speed.
Options: Poor, Excellent
33. Render-Blocking JavaScript
Code portions that prevent quick loading.
Options: Poor, Excellent
34. Browser Compatibility
Ensures compatibility across different browsers and devices.
Options: Poor, Excellent
35. Content Delivery Network
Use of CDN services to improve load times and reduce latency.
Options: Poor, Excellent
36. Response Time
Time for a website server to respond to browser requests.
Options: Poor, Excellent
37. Markup Validation
Evaluates quantity of HTML errors on the site.
Options: Poor, Excellent
38. Broken Links
Detects links that fail to load successfully.
Options: Poor, Excellent
39. Total Number of Hyperlinks
Total count of clickable links on a webpage.

Options: Poor, Excellent
40. Text-Based Hyperlinks
Number of hyperlinks embedded in text.
Options: Poor, Excellent
41. Word Count
Total number of words on the page.
Options: Poor, Excellent
42. Total Body Words
Number of words in the main body.
Options: Poor, Excellent
43. Total Sentence
Number of sentences in a paragraph.
Options: Poor, Excellent
44. Total Paragraph
Number of paragraphs in body text.
Options: Poor, Excellent
45. Total Clusters Count
Number of text clusters on a page.
Options: Poor, Excellent
46. Total Images
Total number of images.
Options: Poor, Excellent
47. Alt Image Count
Number of images with ALT text.
Options: Poor, Excellent
48. No Alt Image Count
Number of images without ALT text.
Options: Poor, Excellent
49. Animation Count
Number of animated elements.
Options: Poor, Excellent
50. Unique Image Count
Number of unique images.
Options: Poor, Excellent
51. Image Map Count
Number of image maps on a page.
Options: Poor, Excellent
52. Un-sized Image Count
Number of images without defined size.
Options: Poor, Excellent
53. Total Number of Distinct Colors
Total unique color values used across the webpage.
Options: Poor, Excellent
54. Reading Complexity
Overall page readability.
Options: Poor, Excellent
55. Number of HTTP Elements Requested
Total number of resources requested to load the page.
Options: Poor, Excellent
56. Availability

Accessibility of the website.
Options: Poor, Excellent
57. The Frequency of Update
How often the website is updated.
Options: Poor, Excellent
58. HTML Page Sizes
Size of all HTML code on the page.
Options: Poor, Excellent
59. Download Time
Average time to download any page.
Options: Poor, Excellent

References

1. Faustina, F.; Balaji, T. Evaluation of universities websites in Chennai city, India using analytical hierarchy process. In Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 3–5 March 2016; IEEE: New York, NY, USA; pp. 112–116. [CrossRef]
2. 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. [CrossRef]
3. Shayganmehr, M.; Montazer, G.A. Identifying Indexes Affecting the Quality of E-Government Websites. In Proceedings of the 2019 5th International Conference on Web Research (ICWR), Tehran, Iran, 24–25 April 2019; pp. 167–171. [CrossRef]
4. Joyami, E.N.; Salmani, D. Assessing the Quality of Online Services (Website) of Tehran University. 2019. Available online: https://un-pub.eu/ojs/index.php/pntsbs/article/view/4519 (accessed on 28 August 2025).
5. Fogli, D.; Guida, G. Evaluating Quality in Use of Corporate Web Sites: An Empirical Investigation. *ACM Trans. Web* **2018**, *12*, 1–35. [CrossRef]
6. Najadat, H.; Al-Badarneh, A.; Alodibat, S. A review of website evaluation using web diagnostic tools and data envelopment analysis. *Bull. Electr. Eng. Inform.* **2021**, *10*, 258–265. [CrossRef]
7. Zahran, D.I.; Al-Nuaim, H.A.; Rutter, M.J.; Benyon, D. A Comparative Approach To Web Evaluation And Website Evaluation Methods. *Int. J. Public Inf. Syst.* **2014**, *10*. Available online: http://www.ijpis.net/index.php/IJPIS/article/view/126 (accessed on 28 August 2025).
8. 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. [CrossRef]
9. Saleh, A.H.; Yusoff, R.C.M.; Bakar, N.A.A.; Ibrahim, R. Systematic literature review on university website quality. *Indones. J. Electr. Eng. Comput. Sci.* **2022**, *25*, 511. [CrossRef]
10. Kitchenham, B. *Procedures for Performing Systematic Reviews*; Technical Report TR/SE-0401; Keele University: Keele, UK, 2004.
11. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Ann. Intern. Med.* **2009**, *151*, W-65. [CrossRef]
12. Ghobadi, S. What drives knowledge sharing in software development teams: A literature review and classification framework. *Inf. Manag.* **2015**, *52*, 82–97. [CrossRef]
13. Niazi, M.G.; Kamran, M.K.A.; Ghaebi, A. Presenting a proposed framework for evaluating university websites. *Electron. Libr.* **2020**, *38*, 881–904. [CrossRef]
14. Adepoju, S.A.; Oyefolahan, I.O.; Abdullahi, M.B.; Mohammed, A.A. Integrated Usability Evaluation Framework for University Websites. *i-Manager* **2019**, *8*, 40–48. [CrossRef]
15. 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. [CrossRef]
16. Elsater, S.A.E.A.; Dawood, A.E.A.A.; Hussein, M.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. [CrossRef]
17. Alsulami, M.H.; Khayyat, M.M.; Aboulola, O.I.; Alsaqer, M.S. Development of an Approach to Evaluate Website Effectiveness. *Sustainability* **2021**, *13*, 13304. [CrossRef]
18. Ghattas, M.M. Performance Evaluation of Websites Using Machine Learning. Master's Thesis, Al-Quds University, Jerusalem Governorate, Palestine, 2020.
19. Kinnunen, M. Evaluating and Improving Web Performance Using Free-to-Use Tools. laturi.oulu.fi. Available online: https://oulurepo.oulu.fi/handle/10024/15601 (accessed on 18 February 2024).

20. Dhiman, P.; Anjali. Empirical validation of website quality using statistical and machine learning methods. In Proceedings of the 2014 5th International Conference—Confluence The Next Generation Information Technology Summit (Confluence), Noida, India, 25–26 September 2014; pp. 286–291. [CrossRef]
21. Zhou, J.; Zhang, Y.; Zhou, B.; Li, S. Predicting web page performance level based on web page characteristics. *Int. J. Web Eng. Technol.* **2015**, *10*, 152. [CrossRef]
22. 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. [CrossRef]
23. Amjad, M.; Hossain, M.T.; Hassan, R.; Rahman, M.A. Web Application Performance Analysis of ECommerce Sites in Bangladesh: An Empirical Study. *Int. J. Inf. Eng. Electron. Bus.* **2021**, *13*, 47–54. [CrossRef]
24. Kaur, S.; Kaur, K.; Kaur, P. An Empirical Performance Evaluation of Universities Website. *Int. J. Comput. Appl.* **2016**, *146*, 10–16. [CrossRef]
25. Harshan, R.K.; Chen, X.; Shi, B. Analytic Hierarchy Process (AHP) Based Model for Assessing Performance Quality of Library Websites. *Inf. Technol. J.* **2016**, *16*, 35–43. [CrossRef]
26. Vatansever, K.; Akgűl, Y. Performance evaluation of websites using entropy and grey relational analysis methods: The case of airline companies. *Decis. Sci. Lett.* **2018**, *7*, 119–130. [CrossRef]
27. Aziz, U.A.; Wibisono, A.; Nisafani, A.S. Measuring the quality of e-commerce websites using analytical hierarchy process. TELKOMNIKA Telecommun. *Comput. Electron. Control* **2019**, *17*, 1283–1290. [CrossRef]
28. Jati, H. Quality Ranking of E-Government Websites—PROMETHEE II Approach. In Proceedings of the International Conference on Informatics for Development 2011 (ICID 2011), Yogyakarta, Indonesia, 26 November 2011; pp. 39–45. Available online: https://www.semanticscholar.org/paper/Quality-Ranking-of-E-Government-Websites-PROMETHEE-Jati/75baad420698797cfca91b7fd1278a512cdecb6b (accessed on 19 February 2024).
29. Khan, H.; Dominic, P.D.D. Measuring Quality of Asian Airline Websites Using Analytical Hierarchy Process: A Future Customer Satisfaction Approach. In Proceedings of the Information Systems International Conference, Bali, Indonesia, 2–4 December 2013.
30. Malhotra, R.; Sharma, A. A neuro-fuzzy classifier for website quality prediction. In Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, India, 22–25 August 2013; pp. 1274–1279. [CrossRef]
31. 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. [CrossRef]
32. Jayanthi, B.; Krishnakumari, P. An Intelligent Method to Assess Webpage Quality using Extreme Learning Machine. *Int. J. Comput. Sci. Netw. Secur. (IJCSNS)* **2016**, *16*, 81.
33. 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. [CrossRef]
34. Massaro, A.; Giannone, D.; Birardi, V.; Galiano, A.M. An Innovative Approach for the Evaluation of the Web Page Impact Combining User Experience and Neural Network Score. *Future Internet* **2021**, *13*, 145. [CrossRef]
35. Devi, K.; Sharma, A.K. Framework for Evaluation of Academic Website. *Int. J. Comput. Tech.* **2016**, *3*, 234–239.
36. Wang, X.S.; Balasubramanian, A.; Krishnamurthy, A.; Wetherall, D. Demystifying Page Load Performance with WProf. In Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13), Lombard, IL, USA, 2–5 April 2013; pp. 473–485. Available online: https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/wang_xiao (accessed on 18 February 2024).
37. De Fausti, F.; Pugliese, F.; Zardetto, D. Towards Automated Website Classification by Deep Learning. *arXiv* **2021**, arXiv:1910.09991. [CrossRef]
38. Rasheed, K.; Noman, M.; Imran, M.; Iqbal, M.; Khan, Z.M.; Abid, M.M. Performance comparison among local and foreign universities websites using seo tools. ICTACT J. *SOFT Comput.* **2018**, *8*, 1559–1564.
39. 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. [CrossRef]
40. Islam, A.; Tsuji, K. Evaluation of Usage of University Websites in Bangladesh. *DESIDOC J. Libr. Inf. Technol.* **2011**, *31*, 469–479. [CrossRef]
41. 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*, 760–776. [CrossRef]
42. Pandya, S. Review paper on web page prediction using data mining. *Int. J. Comput. Eng. Intell. Syst.* **2015**, *6*, 760–766.

43. Dominic, P.D.D.; Jati, H.; Kannabiran, G. Performance evaluation on quality of Asian e-government websites—An AHP approach. *Int. J. Bus. Inf. Syst.* **2010**, *6*, 219–239. [CrossRef]
44. Dominic, P.D.D.; Jati, H.; Sellappan, P.; Nee, G.K. A comparison of Asian e-government websites quality: Using a non-parametric test. *Int. J. Bus. Inf. Syst.* **2011**, *7*, 220–246. [CrossRef]
45. Jain, R.K.; Rangnekar, S. Measuring website quality of the Indian railways. *Int. J. Entrep. Knowl.* **2015**, *3*, 57–64. [CrossRef]
46. Olaleye, S.A.; Sanusi, I.T.; Ukpabi, D.C.; Okunoye, A. Evaluation of Nigeria Universities Websites Quality: A Comparative Analysis. jultika.oulu.fi. Available online: https://oulurepo.oulu.fi/handle/10024/23263 (accessed on 17 February 2024).
47. Dominic, P.D.D.; Jati, H. A comparison of Asian airlines websites quality: Using a non-parametric test. *Int. J. Bus. Innov. Res.* **2011**, *5*, 599–623. [CrossRef]
48. 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. [CrossRef]
49. Hendradjaya, B.; Praptini, R. A proposal for a quality model for e-govemment website. In Proceedings of the 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, Indonesia, 25–26 November 2015; pp. 19–24. [CrossRef]
50. Mittal, H.; Sharma, M.; Mittal, J.P. Analysis and Modelling of Websites Quality Using Fuzzy Technique. In Proceedings of the 2012 Second International Conference on Advanced Computing & Communication Technologies, Rohtak, Haryana India, 7–8 January 2012; pp. 10–15. [CrossRef]
51. Kabassi, K. Analytic Hierarchy Process for website evaluation. *Intell. Decis. Technol.* **2018**, *12*, 137–148. [CrossRef]
52. 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. [CrossRef]
53. Kulkarni, R.B.; Dixit, D.S.K. Empirical and Automated Analysis of Web Applications. *Int. J. Comput. Appl.* **2012**, *38*, 1–8. [CrossRef]
54. Peng, L.; Lu, Y.; Wang, D. Brief analysis on Website performance evaluation. In Proceedings of the Third International Conference on Cyberspace Technology, Beijing, China, 17–18 October 2015; p. 4. [CrossRef]
55. 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. [CrossRef]