iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78736

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

KNN-Based Machine Learning Techniques to Predict Website Hits and Increase Internal Page Traffic Through User Behavior Analysis

Swati Sain[1], Vikas Malik[2]
[1]*Student, M.Tech, CSE,* [2]*Associate Professor, Computer Science & Engineering and information Technology, Bhagat Phool Singh Mahila Vishwavidyalaya, Khanpur kalan, Sonipat (Haryana) 131 301*

Abstract: ***In today's online world, website generate large amount of user interaction, data , including click stream, browsing patterns, and engagement metrics. Proper analysis of this information is essential for predicting website visits and improving internal pages navigation. Machine learning techniques helps automate the process of analyzing user behavior and discovering browsing patterns.***
One widely used most commonly used method is the K-Nearest Neighbor (KNN) algorithm, known for its simplicity and effectiveness in similarity-based prediction. KNN compares a current user session with and finds the most similar patterns to estimate future page visits and traffic distribution.
This review paper explains how KNN models can be used to predict website hits and increase internal page traffic. It discusses web usage mining, feature selection, prediction processes, recommendation method, advantages, Limitation, and possible improvements. The study shows that KNN-based system can improve personalization, reduce bounce rate, and enhance website organization when used with suitable preprocessing and hybrid approaches.
Existing work: KNN-based ML techniques analyze user browsing behavior and to predict website visits, recommend pages, improves personalization, reduce bounce rate, & increase engagement.
Keywords**: *KNN, Web Usage Mining, Machine Learning, Website Hit Prediction, Click stream Analysis, Recommendation.***

I. INTRODUCTION

With the rapid growth of online platforms, understanding how users interact with websites has become increasingly important for business, educational portals, and digital services. Every website visit generates behavioral data such as visited pages, session duration, navigation order, search queries, & user interactions.

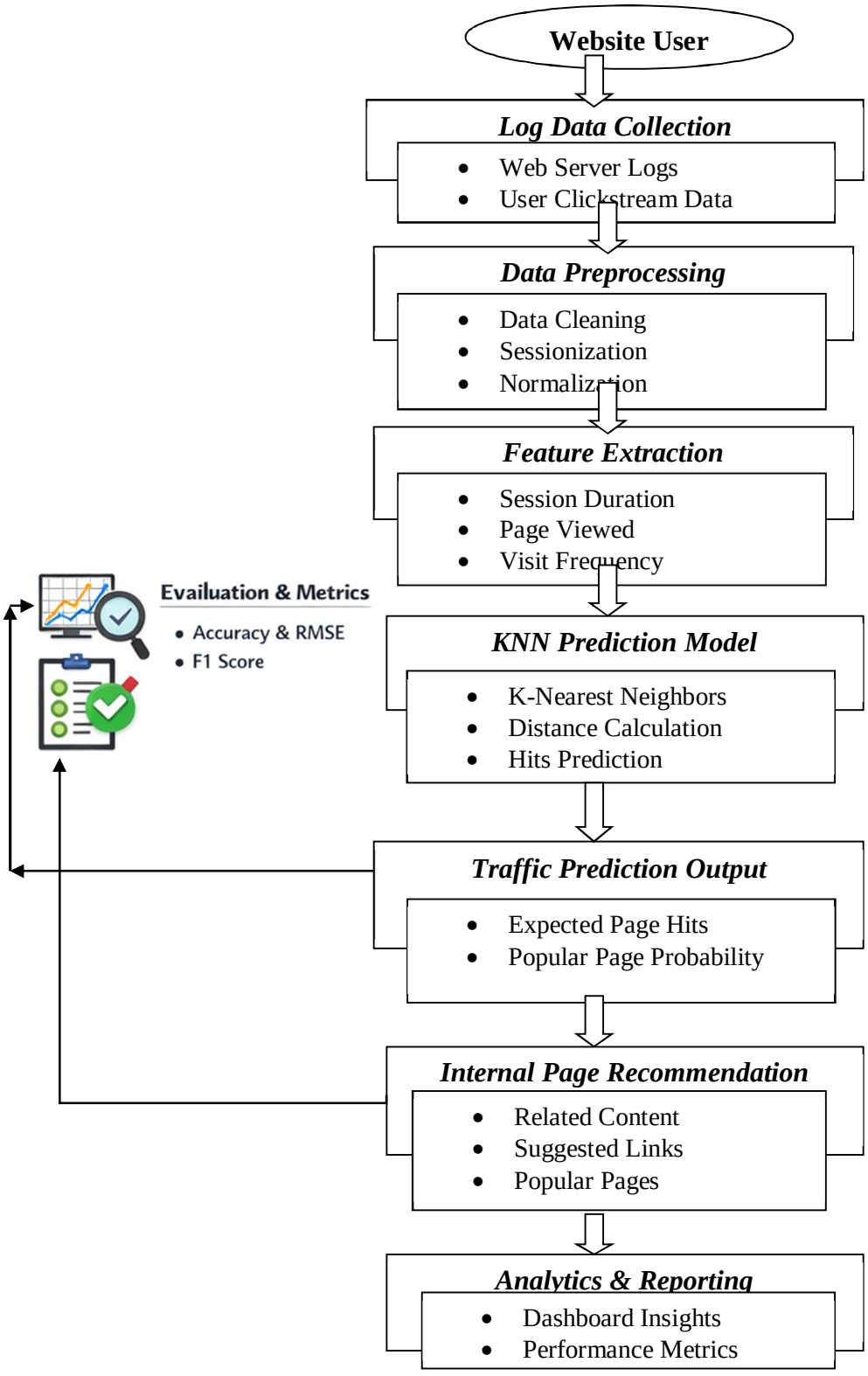
Traditional analytics tools mainly describe past activities and trends but cannot accurately predict future traffic patterns or guide users dynamically. ML techniques overcome this limitation by learning patterns from historical user data and using them to forecast future behavior.

these methods, the K-Nearest Neighbor (KNN) algorithm is a powerful supervised approach that predict outcomes based on similarity between current and previous user sessions. By comparing a new session with historical browsing patterns, KNN can estimate next page visits and expected hits for specific pages. This predictive capability enables websites to optimize internal linking, deliver personalization recommendations, improve navigation structure, increase internal page traffic, reduce bounce rates, and enhance overall user engagement.

©IJRASET: All Rights are Reserved

II. DIAGRAM

KNN-Based Website Hits Prediction & Internal Traffic Optimization



III. WEB USAGE MINING AND USER BEHAVIOR ANALYSIS

Web Usage mining is the process of discovering meaningful patterns from web-based data source including server logs, browser cookies, session records, and click stream information. Its primary goal is to analyze user navigation behavior and predict future browsing actions. User behavior analysis involves collecting raw log data, preprocessing it to remove noise such as bot activity, identifying user session, and converting browsing activity into structure feature representations.

These features may include page sequences, time spent per page, visit frequency, referral information, and interaction event. Such attributes enable ML models to measure similarity among users. When users display similar navigation patterns, their sessions can be clustered, allowing system to forecast future visits and recommend relevant pages. This improves prediction accuracy and supports personalization web experiences.

IV. K-NEAREST NEIGHBOR ALGORITHM IN WEBSITE PREDICTION

The K-Nearest Neighbor (KNN) algorithm is based on the assumption that similar inputs tend to product similar output. Rather than constructing an explicit training model, KNN store historical session data and performance prediction when a new query instance appears. The method computers the distance between the new session and previous session using similarity measures such as Euclidean, Manhattan, or cosine distance. It then selects the K closest sessions and generates predictions based on their observed outcomes.

In website analytics, KNN can estimate page popularity by averaging visit counts from similar session. For navigation prediction, it identifies the next page most frequently accessed among the nearest neighbors. Since browsing behavior often follows repetitive patterns, KNN is well suited for web recommendation and traffic prediction tasks.

V. LITERATURE SURVEY

Table 1: Analysis of Research Papers

Sr. No.	Author's Names	Title	Journal/ Conference details	Methodology	Findings	Limitation
1.	Sonya Zhanga and Neal Cabagneb	Search Engine Optimization: Comparison of Link Building and Social Sharing	Journal of Computer Information System 2017, VOL. 57, NO. 2, 148–159	The study compares the SEO impact of link building and social sharing using website analytics and ranking data.	Link building provides stronger long-term ranking improvement s, while social sharing offers faster short-term traffic gains.	Results may vary by industry, sample size, and changing search or social media algorithms.
2.	Mohamed D. Almadh oun and Nurul Hashimah Ahamed Hassain Malim1	Effects of using multi- category web pages on rank estimation of Google search engine results page	Web Intelligence 2025, Vol. 23(1) 39–55 ©2025– IOS	Collected and compared Google rankings of single-category vs. multi-category web pages using selected keywords.	Multi- category pages generally rank lower due to reduced topic relevance but can perform well if well-structured and internally linked.	Results vary with changing Google algorithms, limited keywords, and narrow sample scope.
3.	Goran Mat oševi´c , Jasminka D obša and Dunja Mladeni´c	Using Machine Learning for Web Page Classification in Search Engine Optimization	Future Internet, Volume 13, Issue 1, in 2021	Applied machine learning models to classify web pages based on content features for improving SEO relevance.	Machine learning- based classification improved the accuracy of identify in g relevant pages for search ranking.	Results depend on dataset size and may vary with changing search engine algorithms.
4.	A.J.W. de Vink1, Natalia Amat-Lefort1, Lifeng Han	Review Graph : A Knowledge Graph Embedding Based Framework for Review Rating Prediction with Sentiment Features	ArXiv 2508.13953 (Year: 2025) August 2025	Used knowledge graph embedding with sentiment features to predict product review ratings using machine learning.	The approach improved rating prediction accuracy compared to traditional text-based models.	Performance depends on data quality and may not generalize well across domains with sparse or incomplete review data.

5.	Joni Salminen, Juan Corporán	Using Machine Learning to Predict Ranking of Web pages in the Gift Industry: Factors for SEO	Conference Paper · March 2019 DOI: 10.1145/3361570.3361578	Used Light GBM and XG Boost o n 733 first-page Google gift- industry web pages with 30 SEO features to predict ranking.	XG Boost achieved ~ 0.86 accuracy, andmost important features were links (internal + external), domain security, and H3 heading length.	The model may not generalize beyond the gift-industry dataset and depends on the selected 30 features.
6.	Mayur Shah & Anmol Valecha	Fake review Detection System using Machine learning	Journal – International Journal of Scientific Research & Engineering DevelopmentVolume-4 Issue5	Machine learning model built using review-based features such as rating, sentiment, and review count, processed and classified using the KNN algorithm.	Higher star ratings, positive sentiment, and frequent recent reviews strongly improve Local SEO ranking prediction accuracy.	The model only uses review-based features and cannot account for other Google ranking factors such as distance, back links, or business profile completeness.
7.	K R Lalitha1, Prof. Dr. T. Tripura Sundari2	Adoption of Local Search Engine Optimization (SEO) among Small-Scale Entrepreneur s in Chennai: Thematic Qualitative Analysis of Expert Insights	Journal of Information Systems Engineering and Management Vol. 10, No. 43s year 2025	Conducted semi- structured interviews with 25 SEO experts in Chennai and analyzed them using Braun & Clarke's six- phase thematic analysis.	Experts identified key adoption barriers like cost and complexity, misconceptions among entrepreneur s, and emphasized local SEO strategies such as GMB optimization and peer influence.	Based only on expert perspectives rather than Direct interviews with the small business owners themselves.
8.	Firas ALMUKHTAR, Nawzad MAHMOODD, Shahab KAREEM	Search Engine Optimization: A Review	Applied Computer Science, vol. 17, Year 2021	Review existing literature and industry sources to analyze common on- page and off- page SEO tech. and ranking factors.	Identified that quality content, back links, keyword optimization, and technical SEO significantly influence search engine rankings.	Review depends on secondary data and may become outdated due to continuously changing search engine algorithms.
9.	Sahana R MBA Student BNMIT, Bangalore	A Study on Effective On- Page and Off-Page SEO Strategies for Building Brand Visibility with Referenceto Hommlie	SSRN working paper, 2025	Used a mixed- method design: surveyed 100 Hommlie users in Bengaluru and analyzed SEO metrics via SEM rush & Ahrefs.	On-page SEO boosts brand discoverability; off-page SEO builds trust, but their synergy and local SEO drive the most impact.	The study is narrowly focused on Hommliein a single city, limiting its broader generalizability
10.	Dina Maher Saad, Hayam Mousa2 , Khalid Amin3	Search Engine Optimization Metric	International Journal of Comp. Trends & Tech.Volume 70 Issue 9, 13-32,September 22	Proposed a new SEO metric called Total- Rank (TR) by combining on-page, off-page, and novel SEO factors.	TR metric outperforms previous SEO metrics in assessing website ranking more accurately.	The metric's ac curacy depends on correct weighting of factors and may not generalize across all website types.

VI. PREVIOUS WORK

1) Web Usage Mining for user Behaviour Analysis: The concept of web usage Mining, which analyzes server logs to understand user browsing patterns. Their work demonstrated hoe user navigation data can be used to improve website structure and recommended relevant pages to user.

2) Predicting web pages Access Using Machine Learning: Machine Learning Model To predict user click and page visits based on browsing history. The study highlighted that user session data can be used to predict future page access, which helps In improving internal page traffic.

3) Website Traffic Prediction using Machine Learning: The application of machine learning algorithms such as KNN, Decision Trees, and Neural Networks for analyzing user activity data. The study conclude that machine learning model can improve website recommendation system and increase internal page visits.

VII. PROPOSED WORK

1) To collect and preprocess website Log data for user behavior analysis.
2) To implement the KNN algorithm for predicting website hits.
3) To analyze user browsing patterns and identify similar user sessions.
4) To predict the next web pages users are like to visit.
5) To increase internal page traffic through predictive recommendations.
6) To evaluate the performance of the proposed model using standard machine learning metrics.

VIII. RESEARCH GAP

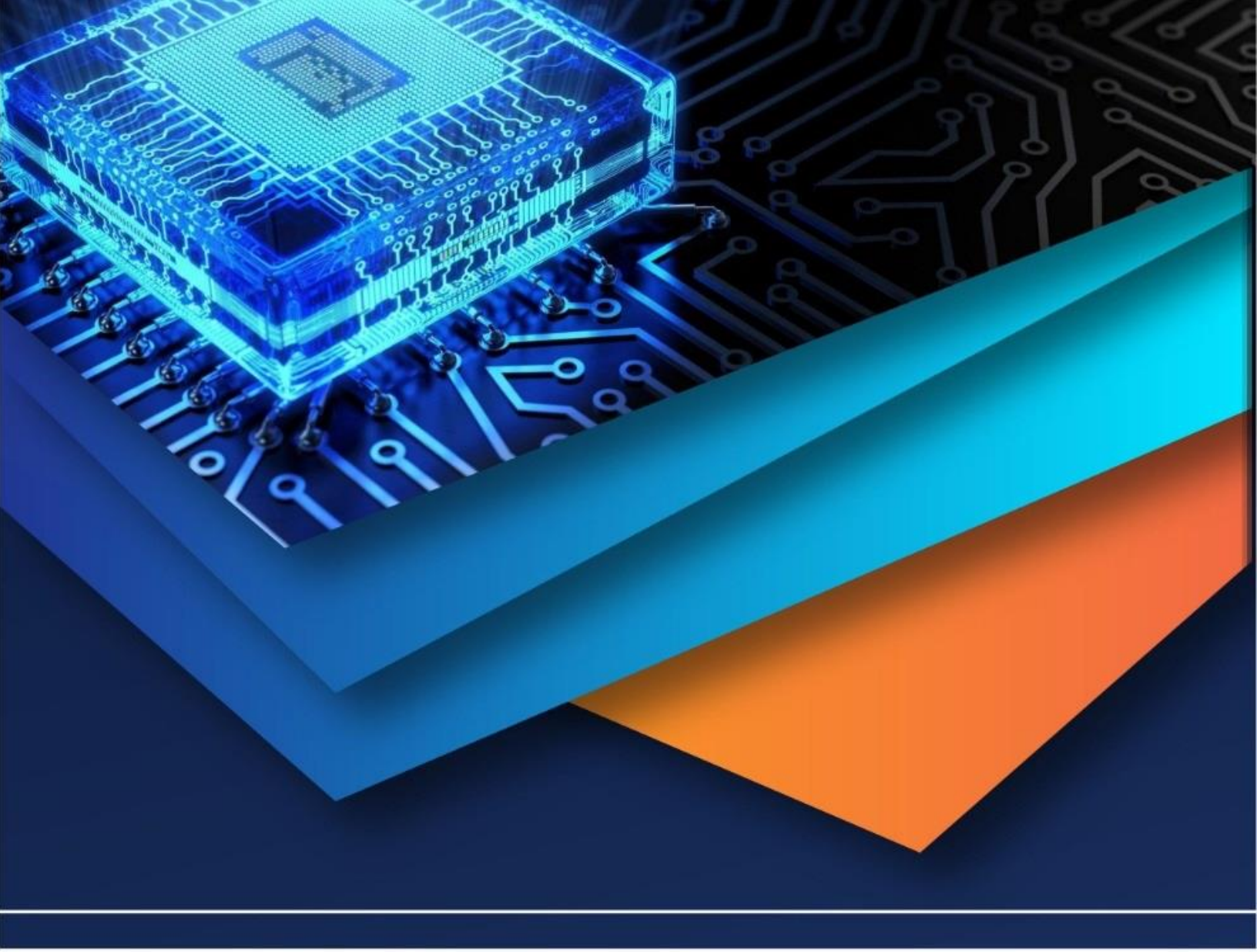
Several studies have examined web usage mining and website traffic analyze past website data instead of predicting which internal pages users may visit in the future. Many traditional methods use statistics models that only describe traffic patterns but do not provide very accurate predictions. In addition, earlier research mostly focused on overall website traffic rather than increasing internal page visits through user behavior analysis. Only a limited number of studies have used machine learning techniques such as the **K-NEAREST NEIGHBOR Algorithm** to predict website hits based on similarities in user sessions. Therefore, there is a develop an effective predictive model the studies user browsing behavior and recommends relevant internal pages to improve website engagements and traffic.

IX. CONCLUSION

In Conclusion, predicting website hits and increasing internal page traffic has become an important task for improving website performance and user engagement. By analyzing user behavior and browsing patterns, it is possible to understand how users interact with different web pages. Machine learning techniques, especially the K-Nearest Neighbor Algorithm, can effectively identify similarities between user session data and web log analysis to recommended relevant internal pages to users. This helps improve navigation, increases pages visits, and enhance the overall user experience. Therefore, applying KNN-based prediction models can support better website management and contribute to higher internal traffic and user satisfaction.

REFERENCES

[1] Sonya Zhanga and Neal Cabageb , Search Engine Optimization: Comparison of Link Building and Social Sharing, / Journal of Computer Information System 2017, VOL. 57, NO. 2, 148–159 https://doi.org/10.1080/08874417.2016.1183447
[2] Mohamed D. Almadhoun and Nurul Hashimah Ahamed, Hassain Malim, Effects of using multi-category web pages on rank estimation of Google search engine results page, Web Intelligence 2025, Vol. 23(1) 39–55 ©2025– IOS 10.3233/web-230239\
[3] Goran Matoševi´c , Jasminka Dobša and Dunja Mladeni´c , Using Machine Learning for Web Page Classification in Search Engine Optimization, Future Internet, Volume 13, Issue 1, in 2021 https://www.researchgate.net/publication/348201897_Using_Machine_Learning_for_Web_Page_Classification_in_Search_Engine_Optimization
[4] A.J.W. de Vink1, Natalia Amat-Lefort1, Lifeng Han , Review Graph:A Knowledge Graph Embedding Based Framework for Review Rating Prediction with Sentiment Features, arXiv 2508.13953 (Year: 2025) August 2025 https://doi.org/10.48550/arXiv.2508.13953
[5] joni Salminen, Juan Corporan Using Machine Learning to Predict Ranking of Webpages in the Gift Industry: Factors for Search Engine Optimization, Conference Paper · March 2019 DOI: 10.1145/3361570.3361578 10.1145/3361570.3361578
[6] Mayur Shah & Anmol Valecha, Fake review Detection System using Machine learning, Journal –International Journal of Scientific Research & Engineering DevelopmentVolume-4 Issue 5 Year 2021 https://ijsred.com/volume4/issue5/IJSRED- V4I5P81.pdf?utm_source=chatgpt.com
[7] Mayur Shah & Anmol Valecha, Fake review Detection System using Machine learning, Journal –International Journal of Scientific Research & Engineering DevelopmentVolume-4 Issue 5 Year 2021 https://ijsred.com/volume4/issue5/IJSRED- V4I5P81.pdf?utm_source=chatgpt.com
[8] K R Lalitha1, Prof. Dr. T. Tripura Sundari, Adoption of Local Search Engine Optimization (SEO) among Small-Scale Entrepreneurs in Chennai: Thematic Qualitative Analysis of Expert Insights, Journal of Information Systems Engineering and Management Vol. 10, No. 43s year 2025 https://doi.org/10.52783/jisem.v10i43s.8515
[9] Firas ALMUKHTAR, Nawzad MAHMOODD, Shahab KAREEM, Search Engine Optimization: A Review, Applied Computer Science, vol. 17,Year 2021 https://doi.org/10.23743/acs-2021-07
[10] Sahana R MBA Student BNMIT, Bangalore, A Study on Effective On-Page and Off-Page SEO Strategies for Building Brand Visibility with Reference to Homilies, SSRN working paper, year **2025** https://ssrn.com/abstract=5476406
[11] Dina Maher Saad, Hayam Mousa, Khalid Amin, Search Engine Optimization Metric, International Journal of Computer Trends and Technology Volume 70 Issue 9, 13-32, September 2022 https://ijcttjournal.org/2022/Volume-70%20Issue-9/IJCTT- V70I9P103.pdf?utm_source=chatgpt.com



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)