Journal of Inventive and Scientific Research Studies (JISRS)
www.jisrs.com Vol: II, Issue: 2 December 2024 ISSN: 2584-0630 (Online)

ADVANCED CLUSTERING TECHNIQUES IN WEB USAGE MINING FOR BUSINESS ANALYTICS: A COMPREHENSIVE REVIEW

P. GANGALAKSHMI[*] and B. MONISHA[2]

[*]Assistant Professor, Department of Computer Science, G.Venkataswamy Naidu College (Autonomous), Kovilpatti

[2]Assistant Professor, Department of Computer Science, G.Venkataswamy Naidu College (Autonomous), Kovilpatti

Received: October 06, 2024, **Accepted**: November 08, 2024, **Online Published**: December 15, 2024

ABSTRACT

In current times, business analytics heavily relies on web usage mining. Various companies analyze the vast amount of data generated by website usage to gain valuable insights. Web Usage Mining (WUM), a component of web mining, is crucial for examining patterns and data from web usage logs, including clickstreams and user behavior. This article thoroughly analyzes advanced clustering techniques applied in web usage mining to tackle challenges stemming from the diverse, vast, and complex characteristics of web usage data. It also discusses hybrid clustering models and their applications in business analytics, including customer segmentation, recommendation systems, fraud detection, and website optimization. Lastly, the article addresses difficulties in existing clustering methods and outlines possible future directions, including real-time clustering, distributed systems, and explainable AI.

Keywords: Web Usage Mining, Clustering Techniques, Business Analytics, Data Mining, Customer Behavior, and E-Commerce.

1. Introduction

The rapid increase in internet usage has led to a large amount of user data, offering both advantages and obstacles for companies. Web Usage Mining (WUM) allows companies to derive practical insights through the analysis of user activities on web pages. Yet, conventional techniques frequently struggle to effectively handle the complex, noisy, and ever-changing characteristics of web usage data. Advanced clustering techniques are being used more often to address these issues by detecting patterns, grouping users, and enhancing services [1].

This paper examines how clustering is utilized in WUM, emphasizing its importance, sophisticated techniques, and use in business analytics. Important contributions of this analysis are: Prioritizing self-care and dedicating time for activities that enhance relaxation and well-being is crucial. An exploration of clustering in WUM as a tool for discovering patterns [23].An in-depth exploration of sophisticated clustering methods and their consequences. Practical business scenarios where clustering is applied for proactive decision-making [28].

Web usage mining (WUM) is becoming increasingly important for businesses to extract valuable information from web logs [4,26], ultimately enhancing strategies and improving customer experiences. With the growing complexity of web data, traditional approaches are struggling to uncover useful insights, leading to the utilization of more advanced clustering methods. Clustering is essential for analyzing user behaviors, playing a critical role in identifying patterns such as popular navigation routes and user preferences within WUM [17]. Contemporary clustering methods such as DBSCAN, OPTICS, hierarchical clustering, and Gaussian Mixture Models (GMM) are beneficial for managing complex, multi-dimensional data [23]. Additionally, autoencoders and hybrid models are also employed to address the evolving challenges of WUM. These methods offer businesses important information for segmenting customers, providing personalized suggestions, detecting fraud, and optimizing websites. Although effective, existing clustering techniques face difficulties in scalability, interpretability, and computational complexity, which new developments like distributed clustering, explainable AI, and real-time systems aim to resolve.

1.1 Overview of Web Usage Mining

Web usage mining involves identifying patterns in web logs automatically, recording user actions on websites. By analyzing user behavior, decisions related to website design and personalization can be influenced [16].

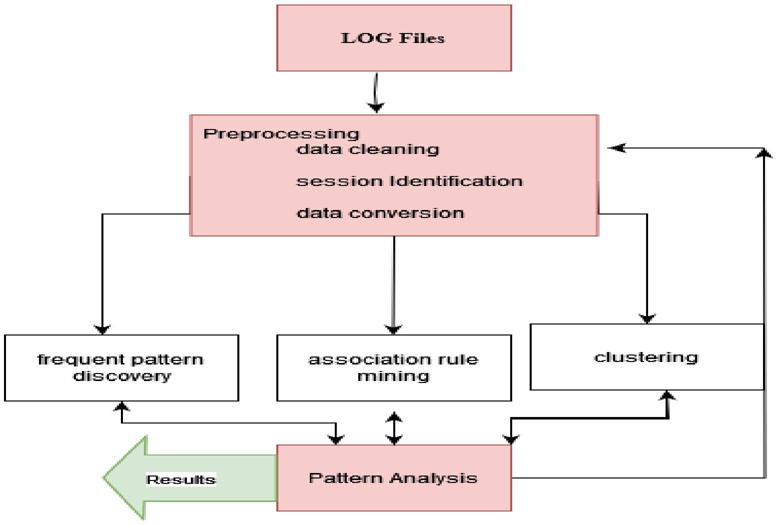

Figure 1: A General Architecture for Web Usage Mining

Advanced techniques are required to manage the growing complexity and sophistication of web data with traditional web usage mining methods. Figure 1 illustrates the three main phases of Web usage mining. Initially, raw data from web logs is collected and scrubbed before being transformed into a format conducive to analysis.

Analyzing the data after formatting it appropriately at this critical stage ensures the quality and relevance of the data used for mining. Data mining techniques such as clustering are used to analyze pre-processed data to uncover hidden patterns. User patterns can indicate user behavior, preferences, and navigation habits [11]. The final phase involves examining the recognized patterns to comprehend them and incorporate them into business choices. Visualizing patterns, summarizing findings, and proposing solutions may be necessary [9,21]. Businesses can utilize this data for segmenting customer base, producing content tailored to user intent, and enhancing marketing campaigns through web usage mining. The large amount of data, along with its lack of organization, presents big obstacles in handling and examining. These challenges can be tackled through the use of clustering methods which help decrease data dimensionality and group similar data points efficiently.

2. Clustering Techniques in Web Usage Mining

Developing a clustering algorithm can be challenging due to factors such as high dimensionality, diverse attributes, determining similarity functions, scalability challenges, and validation processes.

2.1 Traditional Methods for Grouping Data:

K-Means is a clustering technique that classifies data into clusters based on their distance from the cluster 's average [25]. Despite its speed, efficiency declines with more samples and faces challenges detecting non-spherical or differently sized clusters [8,1]. Hierarchical clustering is a valuable method for identifying a hierarchical cluster structure using either an agglomerative or divisive approach, revealing nested clusters with a distinct stopping point [10,27]. However, it can be computationally demanding for large data sets. DBSCAN clustering is efficient with data density, proficient at handling different shapes and noise effectively; however, fine-tuning epsilon and minimum points is essential, posing difficulties for many business applications [12].

2.2 Advanced Techniques for Categorizing Data:

Spectral clustering uses eigenvalues from similarity matrices to reduce dimensions and group web behavioral data, accurately pinpointing remote and nonlinear clusters. SOM transforms complex data into simpler forms, making it easier to uncover hidden patterns. Ensemble clustering combines various algorithms to enhance accuracy and reliability in specific business applications [6]. Additionally, advanced clustering is achievable with deep learning methods using autoencoders, learning feature representations and effectively managing complex, structured data on a large scale. Table 1 demonstrates various advanced clustering techniques and their application in diverse business scenarios.

Clustering Technique	Key Features	Strengths	Weaknesses	Use Cases
Spectral Clustering	Uses eigenvalues of similarity matrix for clustering	Handles non-spherical clusters, flexible in cluster shapes	Computationally expensive, sensitive to scaling of the data	Image segmentation, social network analysis
Gaussian Mixture Models (GMM)	Probabilistic model using Gaussian distributions	Soft clustering, handles varying cluster sizes and shapes	Sensitive to initialization, assumes Gaussian distribution	Density estimation, anomaly detection
DBSCAN Variants (HDBSCAN, OPTICS)	Density-based clustering with noise handling	Identifies clusters with varying densities,	Struggles with varying density within clusters	Geographic data, spatial analysis

		handles noise		
Self-Organizing Maps (SOM)	Neural network that maps high-dimensional data to 2D grid	Effective for data visualization, handles high-dimensional data	Requires careful tuning, may not perform well with large data	Pattern recognition, data visualization
BIRCH	Incremental clustering, builds CF tree	Scalable to large datasets, efficient with one scan	Less effective with very small clusters, sensitive to order of data	Large-scale data clustering, real-time clustering
Affinity Propagation	Clusters based on exemplars without pre-defining cluster number	Does not require number of clusters beforehand	Computationally intensive, sensitive to parameter settings	Image processing, recommendation systems
Advanced Hierarchical Clustering	Builds dendrogram using advanced linkage methods	Reveals hierarchical relationships no need to specify clusters	Computationally expensive for large datasets	Biological taxonomy, genealogy
Deep Learning-based Clustering	Uses autoencoders for dimensionality reduction	Handles high-dimensional data, integrates with deep learning	Complex, requires large amounts of data and computation	Image clustering, text clustering
Fuzzy Clustering (Fuzzy C-Means)	Allows partial membership of data points in clusters	Useful for ambiguous boundaries, flexible cluster membership	Computationally intensive, sensitive to initial parameters	Medical imaging, market segmentation
Constrained Clustering	Incorporates prior knowledge	Incorporates domain	Dependent on quality of	Document clustering,

	with must-link/cannot-link constraints	knowledge, guides clustering process	constraints, may overfit to constraints	customer segmentation

Table 1: Overview of Advanced Clustering Techniques

2.3 The Working Principle of Clustering Algorithms

Clustering algorithms group objects based on similarity or distance to maximize intra-cluster and minimize inter-cluster similarity [2]. Different algorithms use unique strategies to identify clusters. Here are details on the working principles of common clustering algorithms.

2.3.1. K-Means Clustering

Principle:

Partition data into k clusters based on minimizing squared distances between data points and centroids.

Steps:

K-means clustering: randomly choose centroids, assign points to nearest centroid, update centroids, repeat until convergence. Efficient for large datasets, but sensitive to initial centroids.

2.3.2. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Principle:

DBSCAN and OPTICS are effective in identifying clusters and outliers, particularly for anomalies in web usage patterns[4].

Steps:

1. Define parameters: **ε (epsilon)**, the radius to search for neighbors, and **minPts**, the minimum number of points to form a dense region.
2. Each point is classified as core, border, or noise based on its neighbours. Core points and neighbours are connected to form clusters. It can identify complex shapes and handle noise well. Identifies arbitrary clusters, handles noise well, sensitive to parameters, struggles with density.

2.3.3. Hierarchical Clustering

Principle:

Hierarchical clustering uses dendrograms to show data grouping. It has two modes: agglomerative, which merges clusters, and divisive, which splits them [7].

Steps (Agglomerative):

Compute a distance matrix for data points. Treat each point as a cluster, merging closest clusters based on a linkage criterion iteratively. Stop when all points form one cluster or reach a predefined number. Hierarchical relationships captured, no cluster definition required. Computationally expensive, outlier-sensitive.

2.3.4. Spectral Clustering

Principle:

Spectral clustering leverages graph theory and eigenvalue decomposition to group data. It works well for non-linear or non-convex clusters[11].

Steps:

Data represented as a similarity graph with nodes as data points and edges showing similarity. Laplacian matrix computed by differences between degree and similarity matrices. Eigenvalue decomposition on Laplacian for top k eigenvectors, used in K-Means clustering for clusters. Handles complex clusters, captures global relationships. Computationally intensive, needs parameter tuning for graph construction.

2.3.5. Gaussian Mixture Models (GMM)

Principle:

GMM assumes data is from a mix of Gaussian distributions, with clusters represented by Gaussian distributions with mean and covariance[8].

Steps:

Initialize parameters for each Gaussian component. Use EM algorithm: Expectation computes probabilities of point belonging to each cluster, Maximization updates Gaussian parameters to maximize data likelihood. Repeat steps until convergence. Soft clustering and cluster covariance are strengths. Assumes Gaussian distributions and struggles with high dimensions.

3. Comparative Analysis of Clustering Techniques

The study focuses on clustering methods for web-usage mining, emphasizing precision criteria such as cohesion and separation conditions. Standard methods are limited by data complexity, but advanced techniques like spectral clustering and deep learning handle complex patterns and large datasets effectively [17]. These sophisticated approaches are able to address the challenges posed by the dimensionality and abundance of circuit data.

Clustering Algorithm	Number of Clusters	Cohesion (Intra-Cluster Distance)	Separation (Inter-Cluster Distance)	Computational Time (seconds)	Scalability (Large Dataset Handling)	Suitable For Business Applications
K-Means	5	13.2	35.1	1.5	Poor (Scales poorly with large datasets)	Customer Segmentation, Personalization
Hierarchical Clustering	5	15.6	37.8	5.2	Fair (Can be computationally expensive)	Product Category Discovery, Web Traffic Analysis

Clustering Algorithm	Number of Clusters	Cohesion (Intra-Cluster Distance)	Separation (Inter-Cluster Distance)	Computational Time (seconds)	Scalability (Large Dataset Handling)	Suitable For Business Applications
DBSCAN	4	14.5	32.5	3.1	Good (Can handle noise and varying densities)	Anomaly Detection, Churn Analysis
Spectral Clustering	6	16.2	39.4	7.8	Fair (Computationally expensive for large datasets)	Image Segmentation, Social Network Analysis
Gaussian Mixture Models (GMM)	5	14.0	36.5	6.3	Fair (Depends on Gaussian distribution)	Customer Behavior Analysis, Anomaly Detection
Self-Organizing Maps (SOM)	5	12.3	33.2	4.6	Moderate (Requires dimensionality reduction)	Pattern Recognition, Data Visualization
BIRCH	5	13.5	34.8	2.4	Excellent (Handles large datasets efficiently)	Real-Time Clustering, Large-Scale Data Analysis
Fuzzy C-Means	5	14.1	33.9	5.5	Fair (Computationally intensive)	Market Segmentation, Medical Imaging
Mean Shift	4	15.0	36.3	4.9	Fair (Not ideal for very large datasets)	User Behavior Clustering, Personalization
Affinity Propagation	5	15.8	37.5	6.1	Poor (Computationally	Recommendation Systems, Image

Clustering Algorithm	Number of Clusters	Cohesion (Intra-Cluster Distance)	Separation (Inter-Cluster Distance)	Computational Time (seconds)	Scalability (Large Dataset Handling)	Suitable For Business Applications
					intensive)	Processing

Table 2: Comparative Analysis of Clustering Techniques based on quality and performance

Table 2 explains that understand the performance of traditional vs. advanced clustering techniques based on the selected metrics. This comparison will help to determine which techniques are more effective for web usage mining in business data [25].

4. Business Applications of Clustering in Web Usage Mining

Businesses use clustering to segment customer bases by behavior and preferences. This helps in developing effective marketing strategies and improving customer satisfaction. K-means clustering identifies different customer groups like regular buyers and occasional buyers to target with specific promotions. In web usage mining, clustering helps companies understand user behavior, optimize content, and enhance the overall web experience.

4.1 Key Factors in Clustering:

Clustering algorithms in web usage mining are evaluated based on key metrics like Number of Clusters, Cohesion, Separation, Computational Time, Scalability, and Business Applications [5]. Number of Clusters shows how many distinct groups the algorithm forms, Cohesion measures object similarity within clusters, Separation assesses cluster distinctiveness, and Computational Time reflects efficiency [3]. Scalability is essential for handling large datasets, while Business Applications determine how well the algorithm fits business needs like customer segmentation and personalization. These metrics help in improving user experience and targeted marketing in web usage mining. The line chart displays algorithm comparison through Cohesion, Separation, and Computational Time visualizations for clarity.

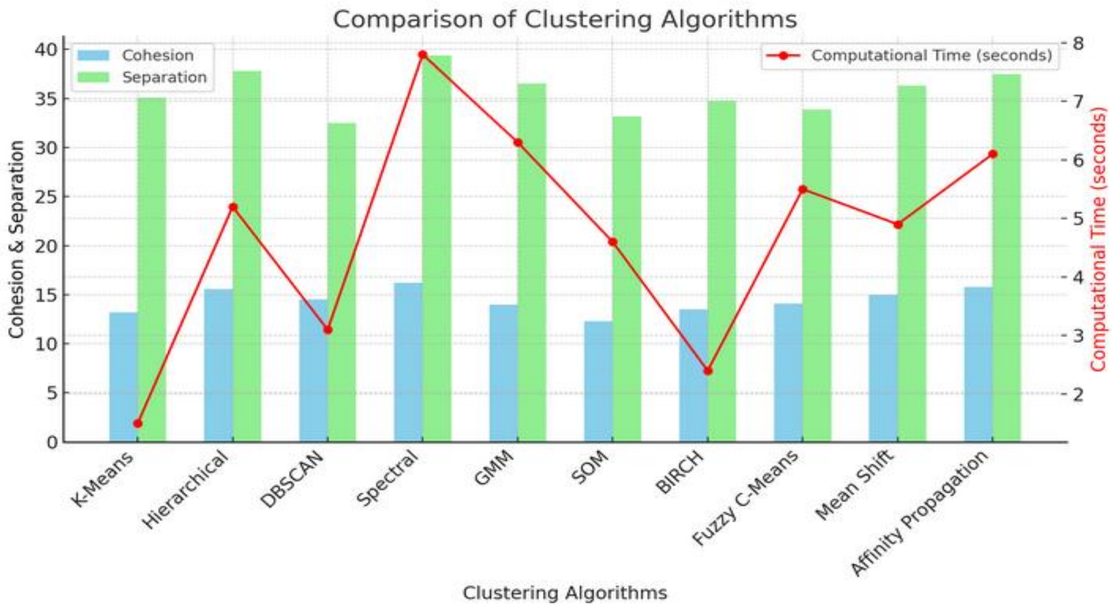

Figure 2: comparison of clustering algorithm

Cohesion measures data similarity in clusters, with lower values showing better cohesion. Separation indicates cluster distinctiveness, with higher values suggesting clearer differences. Computational Time measures algorithm processing time in seconds, with lower times preferred for efficiency.

4.1.1. Key Insights:

Spectral Clustering excels in separation, distinguishing clusters more effectively than other algorithms. K-Means and SOM are more computationally efficient with lower processing times, making them suitable for certain scenarios. Hierarchical Clustering and GMM require more computational resources, making them less efficient for large datasets. DBSCAN offers balanced performance in both cohesion and separation, with a moderate computational time.

4.1.2. Steps for the experiment:

Simulated web usage dataset includes User ID, Session Duration, Pages Viewed, Clicks, Bounce Rate, and Return Rate. Various clustering algorithms used include K-Means, DBSCAN, Spectral Clustering, and Gaussian Mixture Model (GMM). Each method has specific strengths, from traditional partitioning to handling noise and complex cluster shapes. GMM is effective for soft clustering with varying shapes and sizes. Evaluation metrics include cohesion (distance between data points in the same cluster), separation (distance between centroids of different clusters), and computational time (processing time). Low cohesion, high separation, and short processing times indicate better performance.

User ID	Session Duration (min)	Pages Viewed	Clicks per Session	Bounce Rate (%)	Return Rate (%)
1	30	5	20	40	10
2	15	2	8	60	5
3	25	4	18	50	15
4	10	1	3	70	2
5	50	7	30	30	25
6	20	3	12	55	10
7	35	6	25	45	20
8	45	8	35	35	30
9	12	2	7	65	7
10	40	6	28	40	20

Table 3: web usage data set

The clustering results of different algorithms were analyzed based on cohesion, separation, and computational time. K-Means had a cohesion of 0.75, separation of 2.8, and a time of 0.2s. DBSCAN showed cohesion of 0.80, separation of 2.5, and a time of 0.25s. Spectral Clustering had a cohesion of 0.85, separation of 3.2, and took 1.5s. GMM had a cohesion of 0.70, separation of 3.0, and a time of 0.8s. Spectral Clustering performed the best in terms of separation, while DBSCAN was effective at similarity grouping and noise handling. K-Means was fastest, and GMM had a good balance between cohesion and separation for clusters of different shapes and sizes.

4.2 Business Applications:

Businesses utilize web usage mining to categorize customers based on behavior, demographics, preferences, and product engagement[23]. This enables personalized marketing campaigns, improved customer experience, increased conversion rates, and optimal resource allocation to enhance product development and meet customer needs effectively.

Segment	Key Characteristics	Marketing Strategy
Budget Shoppers	Price-sensitive, frequent discount seekers	Focus on discounts, loyalty programs
Premium Customers	High-spending, brand-loyal	Personalized offers, premium services
Occasional Buyers	Low-frequency, opportunistic buyers	Seasonal promotions, targeted ads

Table 4: Example of Customer Segmentation Using Clustering

4.2.1 Personalization and Recommendations:

Clustering methods like Hierarchical Clustering, Mean Shift, and DBSCAN group users based on historical behavior to personalize content or products for better customer satisfaction.

4.2.2 Commercial utility:

Analyzing purchase baskets provides retail businesses with insights for cross-selling and product displays[24]. Market Basket Analysis uses techniques like Hierarchical clustering and Mean shift to identify frequently purchased items.

4.2.3 Clickstream Analytics:

Techniques like DBSCAN and Spectral Clustering detect anomalies and optimize user flow, improving conversion rates through understanding user interests.

4.2.4 Churn analysis:

Classical probability models can help you predict and reduce the churn using Gaussian Mixture Models (GMM), Hierarchical clustering, or K-Means to flag users who are at-risk of churning so that teams can work with them on a suitable retention strategy

4.2.5 Web Traffic Analysis:

Here, techniques such as Spectral Clustering and GMM will be used to analyse web traffic patterns which can then help teams optimize their website design and content placement for capturing user interaction.

4.2.6 Content Optimization:

It enhances content relevance and user interaction analysis with the help of Clustering Feature Popular content is discovered revising and reinforcing your overall approach to content through techniques like K-Means, Hierarchical Clustering Population.

By selecting the appropriate clustering techniques and validating results, businesses can gain valuable insights and enhance their strategies for marketing, customer retention, and overall user experience.

5. Limitations and Future Directions

Existing clustering techniques in web usage mining face challenges such as high dimensionality, scalability, difficulty in handling dynamic and overlapping behaviors, interpretability, and data privacy concerns [22]. Future advancements should focus on scalable, real-time algorithms, hybrid AI models, and privacy-preserving techniques. Emerging trends include integrating AI and machine learning for more accurate predictions and insights. Improving transparency and interpretability of clustering models is essential to make them more accessible for business applications.

6.Conclusion

This review accentuates the importance of clustering in web usage mining (WUM) for business analytics, enabling insights into customer behavior and improved decision-making. Techniques like K-Means (fast but limited with complex data), DBSCAN (handles noise but slower), Spectral Clustering (effective but computationally costly), and GMM (flexible for varying shapes) transform raw web data into actionable intelligence, with ongoing innovation crucial for future needs. Future work in WUM clustering should address scalability, real-time processing, and interpretability through distributed computing, explainable AI (XAI), and hybrid approaches, ensuring adaptability to evolving business needs.

References

1. Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1991). Knowledge discovery in databases: An overview. In *Knowledge Discovery in Databases* (pp. 1–27). Cambridge, MA: AAAI/MIT Press.
2. Klemettinen, M., Mannila, H., & Toivonen, H. (1997). A data mining methodology and its application to semi-automatic knowledge acquisition. In *DEXA Workshop* (pp. 670–677).
3. Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *SIGKDD Explorations*, 2(1), 1–15.
4. Srivastava, J., Cooley, R., Deshpande, M., & Tan, P.-N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 1(2), 12–23.
5. Zaïane, O. R. (n.d.). Principles of knowledge discovery in databases - Chapter 8: Data clustering. Retrieved from http://www.cs.ualberta.ca/~zaiane/ courses /cmput690/ slides/ Chapter8/index.html
6. Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2–3), 107–145. https://doi.org/10.1023/A:1012801612483
7. Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 914–925.
8. Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Burlington, MA: Morgan Kaufmann.
9. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the*

Second International Conference on Knowledge Discovery and Data Mining (KDD-96) (pp. 226–231). Portland, OR, USA.

10. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323.
11. Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2002). Discovery and evaluation of aggregate usage profiles for web personalization. *Data Mining and Knowledge Discovery*, 6(1), 61–82.
12. Kriegel, H.-P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3), 231–240.
13. Pawar, A., Jawale, M., & Kale, C. (2020). Powerful techniques and applications of web mining. *Advances in Pattern Recognition and Artificial Intelligence*, 10(3), 153–167.
14. Ibrahim,K.K.,&Obaid,A. J.(2021).Web mining techniques and technologies: A landscape view.*IOP Conference Series: Materials Science and Engineering*,1879,1–6.
15. Rao, R. S., & Arora, J. (2017). A survey on methods used in web usage mining. *International Research Journal of Engineering and Technology (IRJET)*, 4(5), 2627–2631.
16. Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. In *Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)* (pp. 558–567). Newport Beach, CA, USA.
17. Ivancsy, R., & Kovacs, F. (2006). Clustering techniques utilized in web usage mining. In *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering, and Databases* (pp. 237–242). Madrid, Spain.
18. Kohavi R,&Provost,F. (1998). Glossary of terms. *Machine Learning*, 30(2), 271–274.
19. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). Berkeley, CA, USA.
20. Klemettinen, M., Mannila, H., & Ponkala, P. (2004). A framework for data mining applications in

marketing. *Journal of Marketing Analytics*, 4(1), 1–15.

21. Aznoli, F., & Navimipour, N. J. (2018). A meta-heuristic algorithm for clustering the web pages based on the usage data. *Computers in Human Behavior, 92*, 685–694. https://doi.org/10.1016/j.chb.2018.03.040
22. Tiwari, R., & Prasad, P. W. C. (2018). A framework for clustering web usage mining data using data mining methods. *Journal of King Saud University - Computer and Information Sciences, 34*(5), 2572–2581. https://doi.org/10.1016/j.jksuci.2018.05.009
23. Prasad, A., & Somani, S. (2019). Web usage mining: A survey on pattern extraction from web logs. *Procedia Computer Science, 167*, 2360–2368. https://doi.org/10.1016/j.procs.2020.03.289
24. Ahsan, M., & Siddique, M. A. B. (2020). Machine learning-based clustering algorithms for web user navigation patterns. *Machine Learning with Applications, 1*, 100001. https://doi.org/10.1016/j.mlwa.2020.100001
25. Zhu, J., Zheng, R., & Sun, J. (2020). Clustering user behavior in e-commerce using improved K-means algorithm. *Electronic Commerce Research and Applications, 43*, 101004. https://doi.org/10.1016/j.elerap.2020.101004
26. Raza, M., & Kumar, V. (2021). Web usage mining using hybrid hierarchical agglomerative clustering with optimization techniques. *International Journal of Computational Intelligence Systems, 14*(1), 1277–1293. https://doi.org/10.1002/ijci.1119
27. Wang, H., Zhang, L., Wang, L., & Ma, J. (2022). Customer segmentation based on web behavior and clustering analysis. *Expert Systems with Applications, 193*, 116372. https://doi.org/10.1016/j.eswa.2022.116372
28. Yadav, S., & Chauhan, A. (2023). Web usage mining in personalized recommendations: A clustering perspective. *Journal of Information Science and Engineering, 39*(2), 315–330. https://doi.org/10.6688/JISE.202303_39(2).0005