Shopping Intensity: Unveiling Consumer Segments Through Clickstream Data

Mr. K. Kishore
Asst. Professor, Dept. of AIML
QIS College of Engineering
and Technology, A.P, India
kishore143babu@gmail.com

G.L.N. Sai Anvitha
Student, Dept. of CSE
Qis College of Engineering
and Technology, A.P, India
anvithasai48@gmail.com

K. Khadarpeersaheb
Student, Dept. of CSE
Qis College of Engineering
and Technology, A.P, India
kkhadarpeersaheb@gmail.com

G. Charitha Sri
Student, Dept. of CSE
QIS College Of Engineering
and Technology, A.P, India
charithasri1900@gmail.com

K. Shanmukha
Student, Dept. of CSE
Qis College of Engineering
and Technology, A.P, India
kshanmukha040@gmail.com

A. Surya Srinivas
Student, Dept. of CSE
Qis College of Engineering
and Technology, A.P, India
suryasrinivasalladi@gmail.com

K. Subbarao
Student, Dept. of CSE
Qis College of Engineering
and Technology, A.P, India
nanikavuri05@gmail.com

Sk. Imran
Student, Dept. of CSE
Qis College of Engineering
and Technology, A.P, India
skimranmbdhfm@gmail.com

Abstract:

Understanding consumers is a top priority for businesses that want to maximize and generate sales. we analyze purchase intensity through clickstream analysis to determine clear segments of consumers based on browser behavior. Apply k-means to classify users using wise attributes such as browsing, session length, and purchasing courses. This study provides insights on many types of buyers, ranging from rare browsers to high quality buyers, allowing marketing to adapt and improve customer loyalty. Through data-controlled segmentation, research shows the underlying dynamics of online shopping, providing insights that can be implemented into e-commerce websites. The results are based on data to improve target marketing and customer maintenance by contributing to the deeper understanding of interaction with consumers in Internet transactions. The visualization of the consumer segments can be very crucial for businesses and to increase their revenue through certain users.

Keyword: Clickstream analysis, AI, ML, Consumer segmentation, E-commerce research, Online shopping personalization techniques.

I. Introduction

The increase in development of e-commerce had changed the way consumers interact with online shopping platforms, generating a massive digital footprint known as clickstream data. This data provides valuable insight into browser behavior, product preferences, and purchase decisions, where every action a user records during an online session. Understanding this behavior is important for businesses to optimize customer experience, improve marketing strategies, and improve sales generation. However, the mere volume and complexity of clickstream data makes it difficult to extract wise patterns without advanced data analysis techniques [1]. In contrast to traditional clustering methods based solely on numerical data, Gower distance allows you to handle both numerical and categorical attributes, making it ideal for analyzing customer interactions on e-commerce websites. Grouping users across different clusters allows businesses to identify different shopping personnel. Pulse buyers, bargain seekers, frequent buyers, occasional browsers. Each segment offers unique opportunities for personalized marketing campaigns, product recommendations and improved user experience. Additionally, using silhouette score analysis will help an optimal number of clusters and ensure accurate segmentation [3]. This project improves customer group visualization and facilitates interpretation of complex behavioral trends. Given the increasing reliance on artificial intelligence and data analytics in retail, understanding consumer segmentation is no longer an option, but a need. The results of this study will be used as the basis for further research into customer behavior analysis, predictive modeling, and AI-controlled personalization, shaping the future of digital trade.

The derived insights from this project can grant businesses to make data-driven decisions, and leading to the higher conversion rates and improves customer satisfaction. With the growing reliance on artificial intelligence and data analytics in the retail sector, understanding consumer segmentation is no longer optional but a necessity. The research from this study will serve as a foundation for further research in customer behavior analysis, predictive modeling, and AI-driven personalization, shaping the future of digital commerce [4].

II. Literature Survey

[1] This paper explores methodologies for predicting costs associated with specific behavior's by analyzing peak stream data. It emphasizes the importance of real-time data analytics in identifying patterns that lead to increased expenses, enabling organizations to take proactive cost-management measures.
By monitoring peak periods in data streams, the study demonstrates how organizations can optimize operations and reduce unnecessary expenditures. It highlights the potential of leveraging peak stream data analysis as a strategic tool to enhance efficiency and profitability.

[2] This paper presents a machine learning-based model to predict anonymous customer purchase intentions in online using clickstream data. By incorporating multi-behavioural trendiness (MBT) and also product popularity (POP) metrics, the study enhances prediction accuracy, demonstrating effectiveness on a dataset with over 3 million clicks.
The findings highlight that integrating MBT and POP improves predictive performance and reduces the time required for reliable forecasts. This approach offers valuable insights for e-commerce platforms to optimize customer engagement and boost conversion rates.

[3] This paper investigates consumer segmentation by analyzing clickstream data from a UK-based fast-fashion retailer's e-commerce site. Utilizing the partitioning is done using medoids algorithm on samples of 10,000 unique consumer visits, the study identifies six distinct consumer segments. Notably, the largest segment, termed "mobile window shoppers," generates the lowest revenue, while one of the smallest segments, "visitors with a purpose," contributes the highest revenue.
The findings highlight the potential of clickstream analysis in uncovering unique consumer segments and linking them to revenue generation. This approach offers valuable insights for marketing strategies, enabling more tailored targeting of customer segments to enhance profitability. The study underscores the importance of leveraging big data analytics in the apparel retailing industry to inform marketing decisions and optimize operations.

[4] This paper explores the use of machine learning to predict online helping behaviour through clickstream data analysis. By examining user session timestamps, visit durations, and specific interactions, the study identifies patterns that indicate helping tendencies in online platforms.
The research employs classification algorithms to build predictive models, enabling platforms to enhance user experiences and optimize engagement strategies. The findings offer valuable insights for designing more effective online support systems.

[5] The paper discusses how social media platforms are changing the way people shop online. Instead of users searching for products, social media platforms use algorithms to suggest items based on their interests, leading to spontaneous purchases. Social interactions, such as reviews and recommendations, also influence buying decisions.
It also highlights Xiaohongshu, a platform that combines social networking with shopping, attracting users who value lifestyle over just price. The study emphasizes how influencer marketing and personalized content help build trust and encourage purchases, shaping the future of online shopping.

[6] This paper explores how e-Shopping experiences influence consumer loyalty, highlighting the role of emotions. It finds that factors like website design, ease of use, and personalization create positive emotions, leading to trust and satisfaction, which strengthen customer loyalty.
The study suggests that businesses can enhance loyalty by improving user experience, ensuring secure transactions, and personalizing interactions. By fostering positive emotions, e-commerce platforms can boost customer retention and long-term engagement.

S.No	Title	Techniques Used
1	Enhancing E-commerce Recommendations with Clickstream Data and Reinforcement Learning	Reinforcement Learning.
2	Using Clickstream Data for Predictive Analytics in E-commerce	Predictive Analytics (Regression Models)
3	Consumer Behavior Analysis in E-commerce Using Clickstream Data and Social Media Integration	Data Integration and Analysis
4	Clickstream Data Analysis for Improving User Experience in E-commerce	User Experience (UX) Analytics.
5	A Hybrid Approach to Consumer Segmentation Using Clickstream Data and Machine Learning	Hybrid Machine Learning

Table 1: Comparison of different papers based on what are the techniques used

The Table 1 Describes the highlights that leveraging clickstream data for consumer segmentation provides significant advantages for e-commerce businesses. Traditional and advanced clustering methods, combined with machine learning and deep learning techniques, serve as powerful tools for analyzing and predicting consumer behavior. Future research should focus on developing more robust and scalable approaches, integrating real-time processing for dynamic user tracking, and enhancing privacy-preserving techniques to maintain data security. Addressing these challenges will ensure that clickstream-based consumer segmentation continues to evolve as an essential tool for personalized marketing, customer retention, and business growth in an increasingly digital marketplace**.**

III. Existing System

The clickstream data is a like a huge amount of data which is characterized in large quantities that are difficult to process. The Clickstream data is used to provide the number of clicks, time spent and navigation and so on. Consumer segmentation is defined as the splitting of consumer buyers into groups [1]. There are many approaches for extracting the consumer segments in which it helps in identifying the loyal customers and buyers. The consumer segmentation is the process of Unveiling business patterns and insights of certain group through Clickstream data which is used for growth of the company and to maximize their revenue through certain users [7]. The segmentation of consumers is done using K-medoids in the system. As part of the heterogeneity of consumer knowledge, a lot of consumer segmentation of autonomous media has been written. Nevertheless, there is a study on consumer segmentation according to the consumer's behavior on the Internet [3].

Existing systems for monitoring and analyzing clickstream data essentially the digital footprints users leave as they navigate websites or app come with several disadvantages. These systems are widely used in fields like e-commerce, marketing, and user experience design, but they're not without flaws. The major disadvantages are:

3.1 Data Overload and Noise:

Clickstream data can generate massive volumes of information, often capturing every mouse movement, click, or scroll. Many systems struggle to filter out irrelevant actions like accidental clicks or bot traffic leading to cluttered datasets that obscure meaningful patterns. Processing this firehose of data requires significant computational resources, and without proper refinement, insights can get buried in the noise.

3.2 Privacy and Compliance Challenge:

These systems often collect detailed user behavior, raising red flags under regulations like GDPR or CCPA. Tracking cookies, IP addresses, or session IDs can inadvertently scoop up personal data, and many older systems weren't built with consent management or anonymization in mind. Non-compliance risks fines, while overzealous data collection can erode user trust.

3.3 Limited Real-Time Analysis:

Traditional clickstream tools like Google Analytics or legacy enterprise software often rely on batch processing, meaning data is analyzed after the fact rather than live. This delay can be a problem for businesses needing instant insights, like spotting a sudden drop in conversions or reacting to user frustration mid-session.

IV. Proposed Methodology

The proposed system for revealing consumer segments using clickstream data involves a multi-step process that leverages machine learning techniques and advanced data analytics to understand and categorize consumer behavior on e-commerce platforms. The system begins with the collection of clickstream data, which includes detailed logs of user interactions, such as page views, clicks, product searches, and purchase actions. This raw data is then preprocessed to remove noise and standardize the format, ensuring consistency and accuracy [12].

Following preprocessing, the data undergoes feature extraction where meaningful attributes are identified. These features might include session duration, frequency of visits, types of products viewed, and patterns of navigation through the website. Next, clustering algorithms such as K-means, DBSCAN, or hierarchical clustering are applied to group users into distinct segments based on their behavioral patterns [2]. These segments can range from 'browsers' who frequently visit but rarely purchase, to 'bargain hunters' who extensively compare prices before making a purchase, to 'loyal customers' who regularly buy specific brands or types of products.

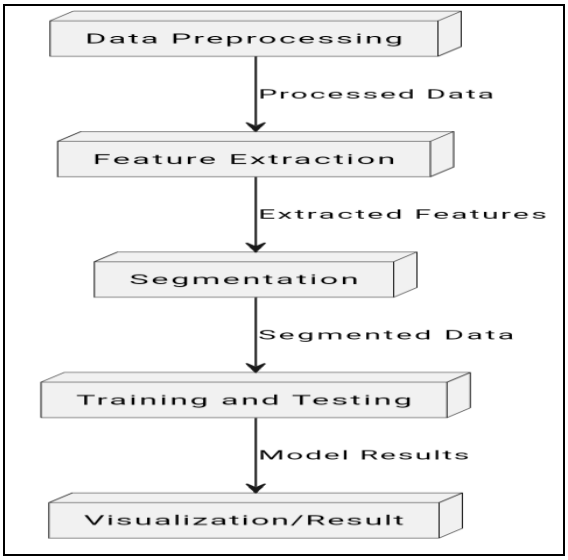

Fig1: Architecture

Fig 1 describes the whole architecture of the Model

4.1 Data Set

The dataset consists of 100,000 rows, with each row representing a unique transaction or event. It contains 11 columns that provide detailed information about user interactions. These columns include User_ID, Session_ID, Timestamp, Page_Type, Product_ID, Category, Action, Device_Type, platform and Location.

Each user is uniquely identified by User_ID, while Session_ID tracks individual sessions. The Timestamp records the exact date and time of interactions. The Page_Type indicates whether the user visited a "Product", "Category", "Home", or "Cart" page. The Action column shows whether the user "Viewed", "Added to Cart", or "Purchased" a product. Additionally, the dataset includes Device_Type (Desktop, Mobile, or Tablet) and Location, which helps in understanding user behavior across different regions.

4.2 Data Preprocessing

Data preprocessing ensures the clickstream data is clean and structured for analysis. It is used to handle missing values, eliminating duplicates, converting timestamps, and encoding categorical features like device type and platform. Numerical values such as session duration and total clicks are normalized and scaled to ensure consistency for clustering algorithms.

4.3 Feature Extraction

Key user behavior metrics are extracted, including session duration, total clicks, unique pages visited, bounce rate, and conversion actions. Other important features include entry/exit pages, navigation patterns, and action sequences, which help in identifying user behavior trends. These structured features make the data ready for segmentation. Key behavioral attributes are extracted, including session duration, total clicks, unique pages visited, bounce rate, and conversion actions [9]. These features provide meaningful insights into user engagement and navigation patterns .

4.4 Segmentation

Users are segmented using K-Means and DBSCAN clustering. K-Means groups users into clusters based on similar browsing patterns, such as frequent buyers, casual visitors, and cart abandoners. The optimal number of clusters and it ensure accurate consumer segmentation.

Number of Clusters (K): Typically set to 3-5 based on user types (e.g., casual browsers, engaged shoppers, loyal customers). Provides well-defined consumer segments that can be used for targeted marketing and user engagement strategies [13].

DBSCAN identifies dense clusters and detects outliers like bots or fraudulent activities, improving the accuracy of segmentation. Detecting outliers and niche consumer behaviors, such as bot traffic, fraudulent activities, or highly specialized buyer groups

4.5 Training and Testing

The K-Means model is trained on extracted features, determining the number of clusters which are optimal using the Silhouette Score. DBSCAN is tested with different distance parameters to detect meaningful clusters. Once trained, the models are applied to new data to predict user segments and assess clustering performance.

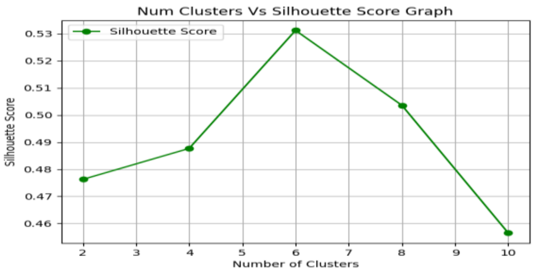

Fig 2: Silhouette score for no. of clusters

The above graph represents each cluster number and count of segmented customer in each cluster and in graph x-axis represents number of cluster and then can see silhouette score for each cluster number and it is done during testing phase. We took clusters as 2, 4, 6, 8 and 10 and in above graph center value is for cluster no 6 which got 0.53 silhouette score and there we got high silhouette score so 6 will be consider as best cluster [3].

4.6 Visualization & Result

The segmented results are visualized using interactive charts and graphs, including bar charts for click sequences, funnel charts for conversion analysis, and scatter plots for cluster representation. Businesses gain insights into user behavior, drop-off points, and high-engagement users, enabling them to optimize marketing strategies and improve retention of customer.

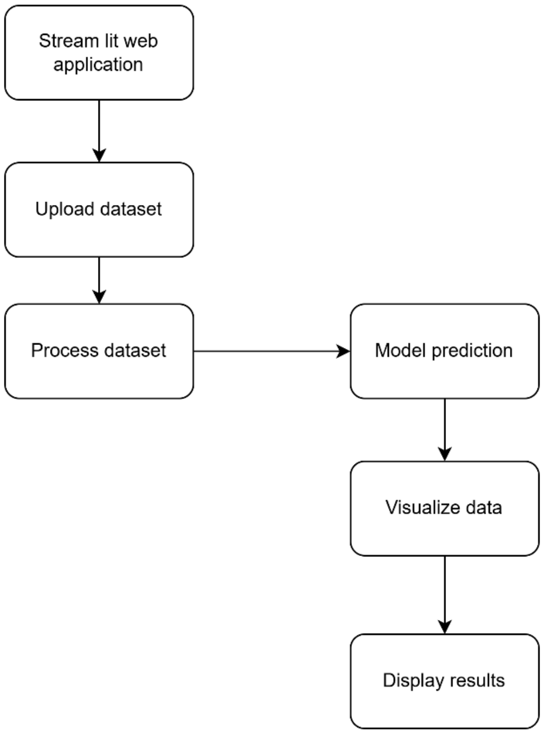

Fig 3: working process

Fig 3 shows the general flow of the website how we represent consumer segments.

The proposed system for revealing consumer segments using clickstream data involves a multi-step process that leverages advanced data analytics and machine learning techniques to understand and categorize consumer behavior on e-commerce platforms. The system begins with the collection of clickstream data, which includes detailed logs of user interactions, such as page views, clicks, product searches, and purchase actions. This raw data is then preprocessed to remove noise and standardize the format, ensuring consistency and accuracy [16].

To enhance the segmentation process, machine learning models are employed to analyze and predict user behavior. For instance, using supervised learning techniques, the system can classify new users into predefined segments based on their initial interactions with the platform. Additionally, advanced models such as neural networks can uncover complex patterns and insights that traditional methods might miss. The resulting consumer segments are then validated and refined using metrics such as silhouette scores to ensure they are meaningful and actionable.

The final results include visualizations of user clusters, insights into behavioral trends, and recommendations for personalized marketing strategies [7]. These insights empower businesses to optimize their website structure, improve targeted advertising, and enhance customer engagement, ultimately leading to higher conversion rates and improved user experience.

V. Results and Discussion

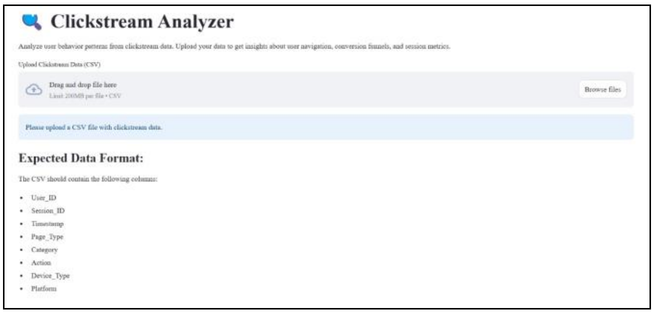

Fig 4: Clickstream Analyzer

This is the input page which takes the data set as an input. The data which is to be processed has to be in the designated format otherwise errors are shown.

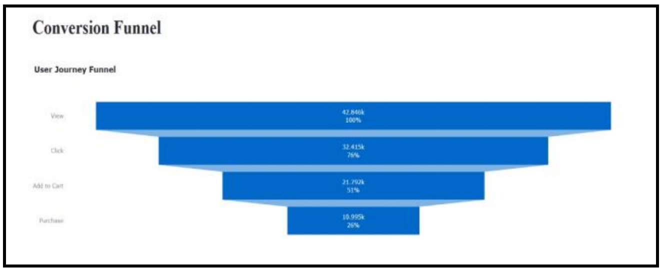

Fig 5: Conversion Funnel

The funnel visualizes the drop-off rates at different stages of the user journey:

View (100%): 42.846k users viewed a page.

Click (76%): 32.415k users interacted by clicking.

Add to Cart (51%): 21.792k users added items to their cart.

Purchase (26%): 10.995k users completed the transaction.

The significant drop-off from "Add to Cart" to "Purchase" indicates potential friction in the checkout process.

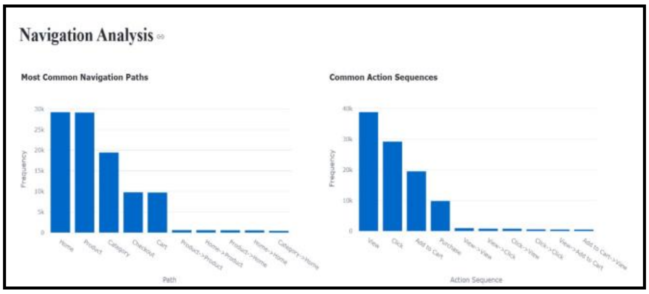

Fig 6: Navigation Analysis

The section highlights how users navigate the site.

Most Common Navigation Paths:

Home and Product pages are the most visited. Category and Checkout pages see fewer visits. Less common paths include transitions between specific pages.

Common Action Sequences:

Users typically start with viewing a page, then clicking, followed by adding items to the cart, and finally purchasing. Some users revisit pages before taking further actions.

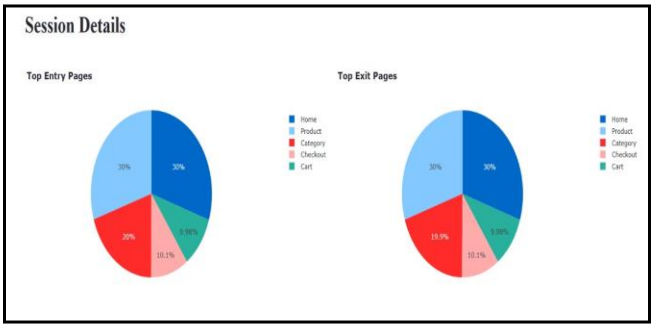

Fig 7: Session Depth Distribution

This provides insights into where users enter and exit the website.

Top Entry Pages:

Home (30%) and Product (30%) pages are the most common entry points.
Category (20%), Checkout (10.1%), and Cart (9.98%) have lower entry rates.

Top Exit Pages:

Home and Product pages also have the highest exit rates.
Category and Checkout pages also see notable user exits, potentially indicating abandonment before completing a purchase.

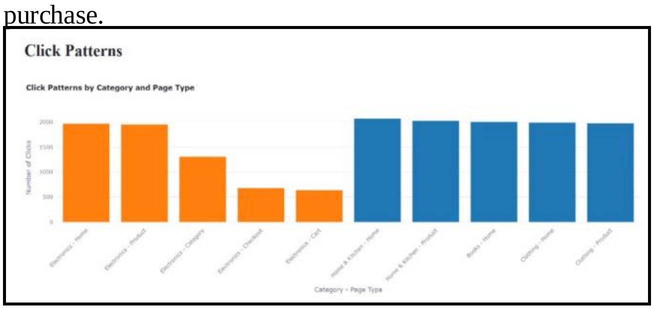

Fig 8: Click Patterns

This visualization shows where users click the most based on category and page type.

1.Electronics (Home & Product) have the highest number of clicks.
2.Electronics (Category, Checkout, and Cart) receive fewer clicks.
3.Other categories, such as Home & Kitchen, Books, and Clothing, also see significant interactions.

VI. Conclusion

This study developed an innovative system for identifying consumer segments using clickstream data and machine learning techniques. By applying K-Means and DBSCAN clustering, we effectively categorized user behaviors, uncovering valuable insights into shopping patterns. These findings enable businesses to personalize user experiences, optimize marketing strategies, and enhance customer engagement. The system's ability to adapt and learn from new data ensures its relevance in the evolving digital landscape, making it a powerful tool for e-commerce platforms seeking data-driven decision-making and long-term customer retention.
The segmented results are visualized using interactive charts and graphs, including bar charts for click sequences, funnel

charts for conversion analysis, and scatter plots for cluster representation. Businesses gain insights into user behavior, drop-off points, and high-engagement users, enabling them to optimize marketing strategies and improve retention of customer.

VII. REFERENCES

1. Gumber, M., Jain, A., & Amutha, A. L. (2021, May). Predicting Customer Behavior by Analyzing Clickstream Data. In 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) (pp. 1-6). IEEE.

2. Wen, Z., Lin, W., & Liu, H. (2023). Machine-learning-based approach for anonymous online customer purchase intentions using clickstream data. Systems, 11(5), 255. https://doi.org/10.3390/systems11050255

3. Zavali, M., Lacka, E., & De Smedt, J. (2021). Shopping hard or hardly shopping: Revealing consumer segments using clickstream data. IEEE Transactions on Engineering Management, 70(4), 1353-1364.

4. Anitha, M., Ramya, K. B., & Zeenath, MD. (2024). Predicting online shopping behavior through clickstream analysis. SRK Institute of Technology, 20.

5. Lee, H. (2024). Interest-Based E-Commerce and users' purchase intention on social network platforms. IEEE Access, 12, 87451–87466. https://doi.org/10.1109/access.2024.3417440

6. Cachero-Martínez, S., & Vázquez-Casielles, R. (2021). Building consumer loyalty through e-shopping experiences: The mediating role of emotions. Journal of Retailing and Consumer Services, 60, 102481. https://doi.org/10.1016/j.jretconser.2021.102481

7. Koehn, D., Lessmann, S., & Schaal, M. (2020). Predicting online shopping behavior from clickstream data using deep learning. Expert Systems with Applications, 150,113342. https://doi.org/10.1016/j.eswa.2020.113342

8. Somya, R., Winarko, E., & Privanta, S. (2021, September). A novel approach to collect and analyze market customer behavior data on online shop. In 2021 2nd International Conference on Innovative and Creative Information Technology (ICITech) (pp. 151-156). IEEE.

9. Noviantoro, T., & Huang, J. P. (2021). Applying data mining techniques to investigate online shopper purchase intention based on clickstream data. Rev Bus Account Financ, 1(2), 130-159.

10. Santhosh Sv & Dr. Basava Rajappa B. (2022). A Study on Consumer Behaviour Towards Online Shopping. International Journal of Research in Analytical Reviews (IJRAR), 9(1), February 2022.

11. Siddhant Sharma, Akhilesh A. Waoo, "Customer Behaviour Analysis in E-Commerce using Machine Learning Approach: A Survey", International Journal of Scientific Research in Computer Science Engineering and Information Technology, Volume 10, Issue 2, March-April 2023, DOI: 10.32628/CSEIT239028.

12. Mofokeng, T. E. (2021). The impact of online shopping attributes on customer satisfaction and loyalty: Moderating effects of e-commerce experience. Cogent Business & Management, 8(1), 1968206. https://doi.org/10.1080/23311975.2021.1968206

13. Chen, Y., & Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. Management Science, 63(12), 4345-4365.

14. Hyun, H., Thavisay, T., & Lee, S. H. (2021). Enhancing the role of flow experience in social media usage and its impact on shopping. Journal of Retailing and Consumer Services. https://doi.org/10.1016/j.jretconser.2021.102492

15. Chaffey, D., Ellis-Chadwick, F., Johnston, K., & Mayer, R. (2019). Digital marketing: Strategy, implementation and practice.

16. Pal, G., Atkinson, K., & Li, G. (2021). Real-time user clickstream behavior analysis based on apache storm streaming. Electronic Commerce Research, 23, 1829–1859. https://doi.org/10.1007/s10660-021-09518-4

17. Requena, B., Cassani, G., Tagliabue, J., Greco, C., & Lacasa, L. (2020). Shopper intent prediction from clickstream e-commerce data with minimal browsing information. Scientific reports, 10(1), 16983.

18. Kumar, A., Salo, J., & Li, H. (2019). Stages of User Engagement on Social Commerce Platforms: Analysis with the Navigational Clickstream Data. International Journal of Electronic Commerce, 23(2), 179–211. https://doi.org/10.1080/10864415.2018.1564550

19. C. Kaushal and H. Singh, "Comparative study of recent sequential pattern mining algorithms on web clickstream data," in 2015 IEEE Power, Communication and Information Technology Conference (PCITC), Conference Proceedings, pp. 652–656C.

20. Dextras‐Romagnino, K., & Munzner, T. (2019). Segmentifier: interactive refinement of clickstream data. Computer Graphics Forum, 38(3), 623 – 634. https://doi.org/10.1111/cgf.13715