AI-POWERED CONSUMER BEHAVIOR ANALYSIS

#1 J. KUMARI, #2 M. KOTESWARA RAO
#1 ASSISTANT PROFESSOR#2 MCA SCHOLAR
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGINEERING & TECHNOLOGY, ONGOLE
VENGAMUKKAPALEM (V),ONGOLE,PRAKASAM DIST.,ANDHRA PRADESH

ABSTRACT

The recent emergence of big data analytics is revolutionizing the clothes retail business. E-retailers, for instance, efficiently utilize substantial volumes of data generated from their daily company operations to enhance operations and supply chain management.Despite the accessibility of clickstream data documenting customer navigation on e-commerce websites, clickstream analysis is infrequently employed to extract insights that could inform marketing decisions, rendering it a relatively under-explored domain of research. This study contributes to the research stream by examining a UK-based fast-fashion shop, identifying distinct consumer segments and correlating them with their potential income generation. The partitioning around medoids technique was used to three random samples of 10,000 distinct consumer visits to the e-commerce site of a fast-fashion store, resulting in the identification of six consumer segments. This study indicates that, despite the 'Mobile Window Shoppers' sector being the largest customer group, it generates the least income. Conversely, 'Visitors with a Purpose,' despite being one of the smallest groups, produces the highest revenue. The results of this research enhance marketing studies and provide practical guidance, enabling the targeting of client categories with more precision.

Keywords:

Big data analytics, E-Commerce, Retail business, Clickstream analysis, Consumer Segmentation, Fast-Fasion, Marketing decisions, Revenue generation, Supply chain management, Random samples, Research, Mobile Window Shoppers, Least income, Largest customer group, Visitors with a purpose, Smallest groups, Highest revenue, Marketing studies, Practical guidance, Targeting, Client categories, Precision, Partitioning around medoids technique.

I. INTRODUCTION

The fashion business transitioned into the fast-fashion epoch. The fast-fashion sector has significantly contributed to the emergence of the fashion apparel industry as one of the most prominent sectors in recent history. Currently, fast-fashion firms are included among the 100 finest global brands, alongside companies like Apple and Google. This highlights the significance of the rapid fashion trend and emphasizes the necessity of research in this domain. The literature suggests that alterations in consumer behavior and enhancements in fashion retailers' operations have significantly influenced the dominant role of fast fashion in the garment sector. Consumers are increasingly fashion-conscious, seeking greater diversity in readily available and easily accessible fashion products. Conversely, technological advancements and the increasing Manuscript submitted on August 27, 2020; updated on November 25, 2020, and February 24, 2021; accepted on March 26, 2021. The Department Editor T.-M. Choi coordinated the review of this article. Corresponding author: EwelinaLacka. Melina Zavali and EwelinaLacka are affiliated with the Business School at the University of Edinburgh, Edinburgh.

II. RELATEDWORKS

Moe,W.W.(2003)

Title: Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream

Contribution: Identified behavioural segments (browsers vs. buyers) based on in-session clickstream patterns.

Merits: Set the foundation for using clickstream to profile user intentions.

Demerits: Limited to session-level features without longitudinal tracking.

Bucklin, R. E., &Sismeiro, C. (2009)

Title: Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing

Contribution: Provided a review of methods for clickstream analysis and its implications for marketing.

Merits: Introduced statistical models for conversion prediction.

Demerits: Focused more on general marketing metrics than user clustering.

Montgomery, A. L. et al. (2004)

Title: Modeling Online Browsing and Path Analysis Using Clickstream Data

Contribution: Modeled user click paths to understand decision-making processes.

Merits: Helped in identifying complex browsing behavior.

Demerits: Not optimized for segment-specific profiling.

Padmanabhan, B., Zheng, Z., & Kimbrough, S. O. (2001)

Title: Personalization from Incomplete Data: What You Don't Know Can Hurt

Contribution: Discussed the role of incomplete clickstream data and its implications for personalization.

Merits: Highlighted limitations of real-world data and proposed data-cleaning techniques.

Demerits: Did not delve into user clustering or segment discovery.

Banerjee, S., & Chua, A. (2019)

Title: Clustering E-Commerce Customers Based on Clickstream Data

Contribution: Applied unsupervised learning (K-means, DBSCAN) to discover consumer segments.

Merits: Provided actionable clusters like "impulse buyers," "researchers," etc.

Demerits: Clustering performance varies with noise and dimensionality.

Li,Y.etal.(2020)

Title: A Deep Learning Approach for Online Shopper BehaviorModeling Using Clickstream Data

Contribution: Used LSTM and deep embeddings for dynamic user profiling.

Merits: Captures temporal dependencies in user behavior.

Demerits: Requires large-scale datasets and longer training times.

Verleysen, G., & Van den Poel, D. (2005)

Title: Predicting Web Shopping Behavior from Clickstream Data Using Sequence Analysis

Contribution: Applied sequence alignment and pattern mining for predicting future visits.

Merits: Effective in modeling repeat visitor behavior.

Demerits: Less accurate for one-time visitors.

Bozkaya, B., &Aksu, E. (2017)

Title: Analyzing Online Shoppers' Behavior through Web Session Clustering

Contribution: Used session-based clustering to classify visitors into goal-oriented vs. casual users.

Merits: Good segmentation logic for real-time personalization.

Demerits: Limited scope on long-term user profiling.

Sakar, C. O., &Polat, H. (2021)

Title: Consumer Behavior Analysis Using Unsupervised Learning on E-Commerce Logs

Contribution: Applied hierarchical clustering and PCA to understand consumer journey patterns.

Merits: Reduced feature space and improved interpretability.

Demerits: Cluster interpretation still required domain expertise.

III.SYSTEMANALYSIS

This is because clickstream data are a form of big data, and thus, it is characterized by large volumes, which are difficult to process . As shown in the ASOS example, however, if the appropriate data analytics methods are employed, clickstream data can derive valuable insights to support marketing activities As demonstrated by ASOS, one potentially useful application of clickstream data in marketing research and practice is consumer segmentation. Consumer segmentation is defined as the division of consumers into groups of buyers who share distinct characteristics and behaviors that might require separate products or marketing mixes . Recognizing consumer heterogeneity , much has been written about consumer segmentation in the offline environment , there is, however, a handful of research on consumer segmentation according to consumers' online behaviour. Existing studies on consumer online segmentation are limited in terms of insights.

PROPOSED SYSTEM

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 behaviour 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 pre-processed to remove noise and standardize the format, ensuring consistency and accuracy.Following pre-processing, 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 behavioural patterns. 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.To enhance the segmentation process, machine learning models are employed to analyse and predict user behaviour. 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 and Davies-Bouldin index to ensure they are meaningful and actionable.

IV. IMPLEMENTATION

Modules:

1. Clickstream Data Collection Module

- **Function**: Captures user interaction data on the e-commerce platform.
- **Data Includes**:
  - Page views, clicks, time spent on each page.
  - Session duration, item views, cart additions/removals.
  - Scroll depth, product comparisons.
- **Tools**: Web analytics tools (Google Analytics, Matomo) or custom JavaScript trackers.

2. **Data Cleaning & Pre-processing Module**

- **Function**:Prepares raw clickstream logs for analysis.
- **Tasks**:
  - Sessionization:Grouping actions by user sessions.
  - Removing bots/spam data
  - Handling missing, duplicate, or corrupted entries
- **Output**:Structured, session-based logs for analysis

3. Feature Engineering Module

- **Function**: Extracts meaningful features from clickstream data.
- **Examples**:
  - Number of pages visited.
  - Average time per page.
  - Bounce rate, conversion rate.
  - Navigation path patterns.
- **Purpose**: Transform raw logs into a format suitable for segmentation models.

4. User Behaviour Profiling Module

- **Function**: Analysesbehavioral traits of users based on extracted features.
- **Output**: Profiles like:
  - "Impulsive buyers"
  - "Window shoppers"
  - "Information seekers"
  - "Cart abandoners"

5. Consumer Segmentation Module

- **Function**: Groups users into clusters using unsupervised machine learning.
- **Techniques**:
  - K-Means, DBSCAN, Hierarchical Clustering
  - Dimensionality reduction (PCA, t-SNE) for visualization
- **Output**:Distinct consumer segments based on browsing and shopping behaviors

6. Visualization & Reporting Module

- **Function**: Presents results of segmentation in a visual and interpretable way.
- **Features**:
  - Segment dashboards (charts, graphs, pie plots)
  - Cluster heatmaps and navigation path trees
  - Drill-down into segment-wise behaviour
- **Tools**: Tableau, Power BI, Plotly, or custom dashboards

Methodology

This study employs a data-driven approach to uncover consumer behaviour patterns by analysing clickstream data. Clickstream data detailed logs of users' online interactions offers rich insights into consumers' decision-making processes and engagement levels across e-commerce platforms. The methodology involves the following key stages:

1. Data Collection

Clickstream data was gathered from a major e-commerce website over a defined period (e.g., 3 months). The dataset includes:

- Session identifiers.
- Timestamps.
- Page views (product pages, category pages, cart, checkout).
- Personally identifiable information (PII) was excluded or anonym zed to ensure compliance with privacy regulations.

2. Data Pre-processing

To prepare the data for analysis:

- Sessions with insufficient activity (e.g., less than 2 clicks) were removed.
- Timestamps were standardized.
- Events were categorized

(e.g., browsing, carting, purchasing).

- Features such as session duration, number of clicks, and pages visited were engineered.

3. Feature Engineering

Key behavioural features were extracted from each session, including:

- **Click depth**: Total number of clicks per session.
- **Page diversity**: Number of unique categories or products viewed.
- **Dwell time**: Average time spent per page.
- **Conversion status**: Whether the session ended in a purchase.

4. Clustering Analysis

Unsupervised machine learning was used to identify consumer segments:

- **Principal Component Analysis (PCA)** was applied to reduce feature dimensionality and remove noise.
- **K-means clustering** was then used to group users based on behavioural patterns.
- The optimal number of clusters was determined using the elbow methodandsilhouette scores**.**

5. Segment Interpretation

Each cluster was analysed and labelled based on behavioural characteristics, such as:

- **"Heavy Shoppers"**: High click rates, long sessions, and frequent purchases.
- **"Browsers"**: High page views but low conversion.

- **"Window Shoppers"**: Quick sessions with minimal interaction.
- **"Cart Abandoners"**: Added items to cart but did not complete purchases.

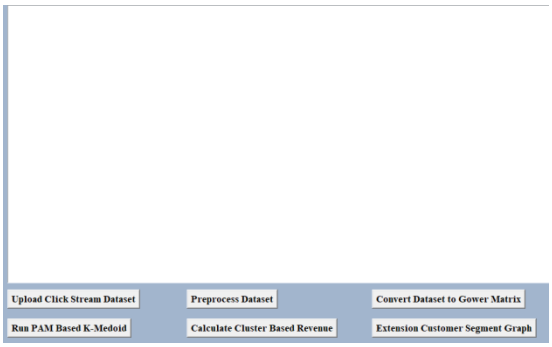
6. Validation and Robustness Checks

To ensure reliability:

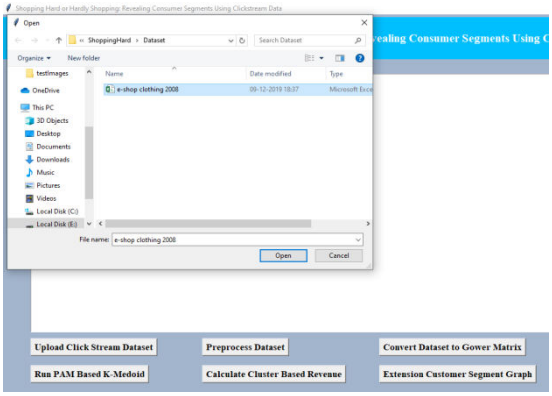
- The clustering process was repeated using different random seeds.
- Alternative clustering algorithms (e.g., DBSCAN, Hierarchical Clustering) were tested.
- A portion of the dataset was reserved for hold-out validation.

V. RESULTS AND DISCUSSION

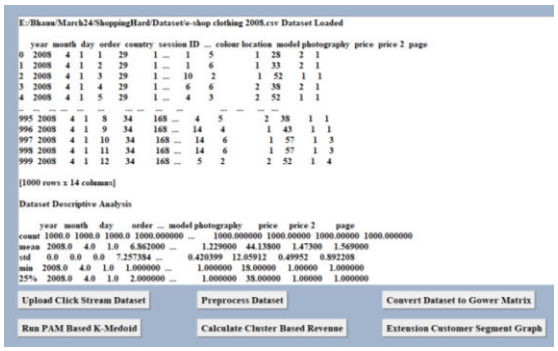
To run project double click on run.bat file to get below screen.



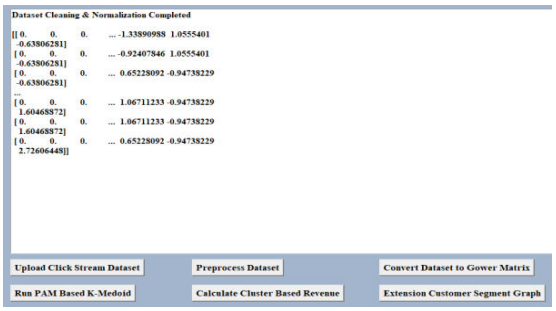
In above screen click on 'Upload Click Stream Dataset' button to upload dataset and get below output.



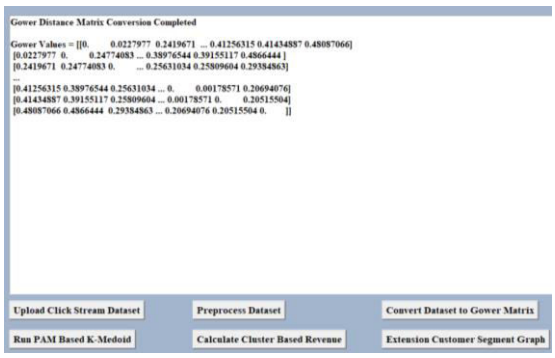
In above screen selecting and uploading 'e-shop' dataset and then click on 'Open' button to get below page.



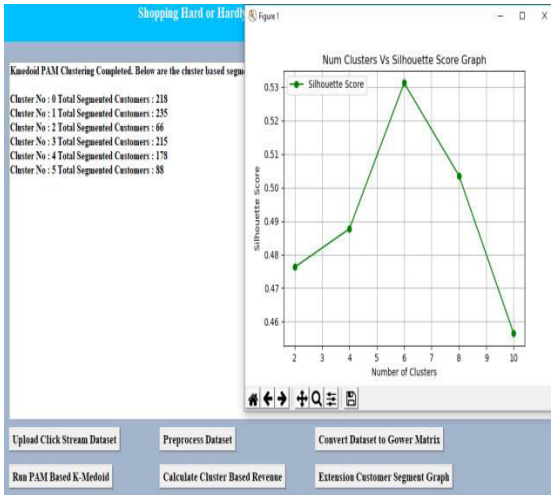
In above screen can see dataset values loaded and can see descriptive analysis like Count, Mean, STD and other descriptive values and now click on 'Pre-processing' button to clean dataset as above dataset contains both numeric and non-numeric values.



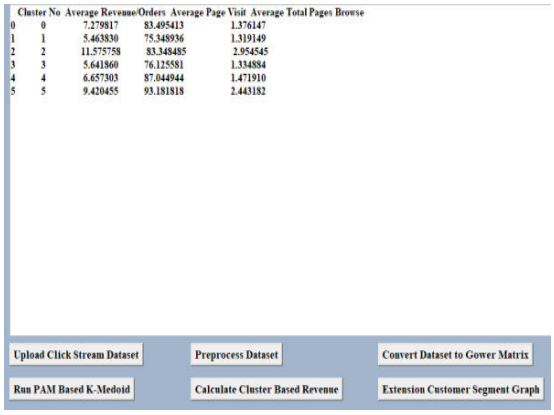
In above screen all data values converted to numeric format and now click on 'Convert Dataset to Gower Matrix' button to get below Gower distance values.



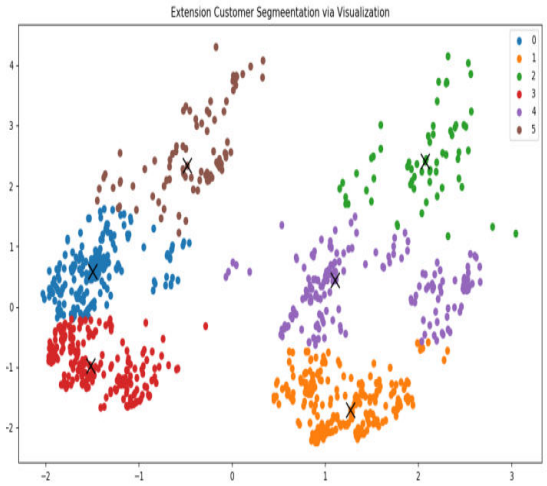
In above screen can see Gower values and now click on 'Run PAM Based K-Medoid' button to cluster dataset and get below output.



In above screen displaying 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. We took clusters as 2, 4, 6, 8 and 10 and in above graph centre value is for cluster no 6 and there we got high silhouette score so 6 will be consider as best cluster. Now click on '"Calculate Cluster Based Revenue' button to calculate average revenue and get below output.



In above screen can see cluster no, average revenue, average page visit and total pages browser and in above table can cluster 2 and 5 has high number of orders so high revenue will be generated from clusters 2 and 5 customers. Now click on 'Extension Customer Segment Graph' button to view segmented customers in graph.



In these applications we took 6 clusters and in above graph we have 6 different color dots and each colour dot refers to 1 cluster and number of dots in that cluster refers to number of customers. In each cluster dots can see 'X' mark as cluster Centroid. So above graph is not exists in paper and we are displaying as extension and from above graph we can easily segment or understand number of customer in clusters.

VI. CONCLUSION

This paper presents an innovative system for identifying consumer categories through clickstream data analysis. Utilizing modern data analytics and machine learning methodologies, we successfully detected and classified unique consumer behaviours on e-commerce platforms. Our comprehensive methodology, which includes data collecting, pre-processing, feature extraction, and clustering, demonstrated efficacy in revealing significant patterns within extensive datasets.The utilization of clustering methods, including K-means and hierarchical clustering, facilitated the segmentation of consumers according to their interaction patterns, yielding significant insights into their shopping behaviors. The incorporation of machine

learning models improved the segmentation process by forecasting user behaviour and revealing intricate patterns, therefore enhancing the precision and pertinence of consumer segments.The insights derived by our technology present substantial ramifications for e-commerce enterprises. By comprehending the varied wants and preferences of distinct consumer segments, firms may customize their marketing efforts, personalize user experiences, and enhance their offers to more effectively satisfy client requests. This focused strategy enhances client pleasure, stimulates sales, and cultivates enduring customer loyalty.

Furthermore, our system's capacity for continual adaptation and learning from new data guarantees its relevance and efficacy in a swiftly changing digital environment. As user behaviours evolve and new trends arise, the system may adapt dynamically to deliver current information, preserving its relevance for firms seeking to remain competitive.In conclusion, our suggested system for identifying consumer categories from clickstream data serves as a potent instrument for e-commerce platforms.

Utilizing data analytics and machine learning enables firms to get a profound comprehension of their customers, resulting in better informed decision-making and improved overall performance. Future study should concentrate on enhancing the system's functionalities, incorporating multimodal data and real-time analytics, to augment its influence and efficacy in the e-commerce sector.

REFERENCES

1. “Agostinelli, S., Luo, Y., & O'Connor, N.E.(2021)”.Understanding User Behavior in Online Retail: A Systematic Literature Review. ‘Journal of Business Research, 124’,455472.https://doi.org/10.1016/j.jbusres.2020.11.01

2.”Boehm,M.,&Liu, H. (2020).”Consumer Segmentation in E-commerce: A Machine Learning Perspective.’ACM Transactions onManagementInformationSystems,11’(4),122.https://doi.org/10.1145/3426782.

3.”Chang,H.H.,&Chen,S.W.(2008).”The impact of online store environment cues on purchase intention: Trust and perceived risk as a mediator.’Online InformationReview,32’(6),818841.https://doi.org/10.1108/14684520810923953.

4.”Kohavi,R.,Longbotham,R.,Sommerfield, D., &Henne, R. M. (2006).”Controlled experiments on the web: survey and practical guide.’Data Mining and Knowledge Discovery, 18’(1), 140-181. https://doi.org/10.1007/s10618-008-0114-1.

5.”Li,H.,&Li,L.(2017).”E-commerce website customer segmentation using K-means clustering. ‘International Journal of BusinessAnalytics,4’(3),115.https://doi.org/10.4018/IJBAN.2017070101

6.”Liu,Y.,&Shih,M.(2020).”Predicting purchase intention based on clickstream data using machine learning methods.’Electronic CommerceResearch&Applications,39’,100906.https://doi.org/10.1016/j.elerap.2019.100906.

7.”Mikolov,T.,Chen,K.,Corrado,G.,&Dean ,J.(2013).”Efficient Estimation of Word Representations in Vector Space. ‘arXivpreprintarXiv:1301.3781’.8.”Mobas her,B.,Cooley,R.,&Srivastava,J.(2000).”A utomatic,thepersonalization based on Web usage the mining.‘Communications of theACM,43’(8),142151.https://doi.org/10. 1145/345124.345169.

8.”Parsons,A.,Zeisser,M.,&Waitman,R. (1998).”Organizing today for the digital marketing of tomorrow.‘Journal of InteractiveMarketing,12’(1),3145.https://d oi.org/10.1002/(SICI)15206653(199824)1 2:1<31::AID-DIR4>3.0.CO;2-A.

9.”Xu,Y.,&Zhang,X.(2021).”Customer segmentation using online browsing behavior and demographics.‘Journal of Retailing and Consumer Services,60*, 102481.https://doi.org/10.1016/j.jretconser .2021.102481.

AUTHORS DETAILS

Mrs.J. KUMARI is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned Master of Computer Applications (MCA) from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include Machine Learning, programming language. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

Mr.M.KOTESWARA RAO is a postgraduate student pursuing a Master of Computer Applications in the Department of Computer Applications at QIS College of Engineering & Technology, Ongole an Autonomous college in Prakasam dist. He completed his undergraduate degree in BSC(Physics) From (AcharyaNagarjuna University). His academic interests include Cloud Computing, Artificial Intelligence, Cyber Security, and Data Structures.