An algorithm-based approach for Mapping customer journeys by identifying customer browsing behaviors on E-commerce Clickstream data

Shiuann-Shuoh Chen
Dept.of Business Administrattion
National Central University
Taoyuan, Taiwan
kenchen@mgt.ncu.edu.tw

Tzu-Ling Li
Dept.of Business Administrattion
National Central University
Taoyuan, Taiwan
nicki831028@gmail.com

Yu-Chen Wu
Dept.of Business Administrattion
National Central University
Taoyuan, Taiwan
chloewu199609@gmail.com

Vinay singh
Dept. of Electronics und Information
Universität Siegen
Siegen, Germany
vinaysingh.bvp@gmail.com

2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) | 979-8-3503-3752-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/HORA58378.2023.10156808

Abstract - ***We propose a novel method to analyze clickstream data that can assist E-commerce retailers with marketing precision based on a better understanding of customer intentions. Design/Methodology/approach – Multiple approaches have been used with real word data from an electronic online retailer's clickstream data. This database contains 1,921,451 data from July 2019 to March 2020 from an electronic e-commerce retailer in Taiwan. In our approach, we cluster clickstream data with the K-means model and then apply the Decision tree model to identify a pattern in each cluster. On Mapping these clusters, we further analyzed that each cluster could map to the pre-purchase and purchase of customer journey model. Findings – We find a novel approach of using each session of browsing behavior as one unit to prevent a situation in which a customer may have different intentions during different periods. Furthermore, this study mainly focuses on browsing time in each category to analyze customers' preferences. Nevertheless, the results show it is a feasible approach, and one can explore customers' intentions from a new perspective. This study provides a new customer browsing intention's analyzing mode of integrating theory and practice by clickstream data, which can be applied to business. Research Limitations/Implications – Our research is limited to the browsing time in each category. One way to improve precision and accuracy would be to add other attributes to the model, such as browsing paths. Another is clickstream data cannot map retention and advocacy of customer journey. Originality/Value – We overcame the limitation of previous research, which focused on the customer journey, by taking one customer's all consumer behavior as one unit. In our study, we use each session of browsing behavior as one unit to prevent a situation in which a customer may have different intentions during different periods.***

Keywords - Consumer browsing behavior, Customer journey, Clickstream data, E-commerce, Machine learning

I. INTRODUCTION

With the advancement of technology, more and more consumers will shop online. The pandemic has accelerated the shift towards a more digital world and triggered changes in online shopping behaviors that will likely have lasting effects. As a result, Clickstream data has gathered much attraction and is gaining an increasingly important role in companies. Clickstream data represent granular information on every online touchpoint consumer purchase [1]. Clickstream data are ubiquitous now, and companies can use these data to support a wide range of strategic decisions, including website design, advertising, and post-purchase relationship management [2]. In addition, clickstream data also have provided insights into consumer intention or motivation [3].

Understanding consumer intention or motivation is an extensively studied research topic in e-commerce, which has helped enterprises gain large profits and promote applications. So far, the marketing literature distinguishes two broad categories of shopping motivation: goal-directed search and exploratory search. Wolfinbarger and Gilly use the terms "for fun" (exploratory) and "for efficiency" (goal-directed) to describe and distinguish consumers' different orientations. For those exploratory search consumers, how to give them a better consumer experience and promote them properly to increase purchase rate is an important issue for companies. Therefore, multiple companies, such as Amazon and Google, now have chief customer experience officers, customer experience vice presidents, or customer experience managers responsible for creating and managing their customers' experience. Some studies focus on exploratory attempts to conceptualize and measure customer experience [4, 5]. Lemon and Verhoef summarized the conceptual background of customer experience as a dynamic process and mapped the customer journey model for customer experience [6]. Customer journey analysis helps enterprises understand customer views and apply this model in marketing.

During our literature review, we could only find limited studies focusing on the application of clickstream data for customer behavior modeling [3, 9,10]. Thus, this study aims to use clickstream data to map the customer journey model by machine learning. We use a nine-month sample of 1,921,451 clickstream data from an online electronics retailer. We employ the K-means algorithm to cluster clickstream data and use the Decision tree algorithm to analyze the pattern for each cluster to map the customer journey model. To our best knowledge, none of these existing studies has focused on the effective use of clickstream data to identify customer intention with the customer journey model by machine learning. In addition, previous studies about the customer journey model were from the customer's perspective. However, we challenge this framework and use the session as a unit of analysis. The benefit of our approach is it can prevent a situation in which a customer may have different intentions during different periods.

The main contribution of our research can be summarized as follows :

- Our work proposes a novel approach, using each browsing behavior session as one unit to prevent a situation where a customer may have different intentions during different periods.

- This study mainly focuses on browsing time in each category to analyze customers' preferences. The results of our study show it is a feasible approach, and one can explore customers' intentions from a new perspective.

- It explores a new customer browsing intention's analyzing mode of integrating theory and practice by clickstream data, which can be applied to business.

- We overcame the limitation of previous research, which focused on the customer journey, by taking one customer's all consumer behavior as one unit.

- We also set future research agenda and how more theory can be built on the topic, stating some of the key areas where researchers can focus and further build upon.

The remaining manuscript is organized as follows. In section 2 we review existing related studies and the theoretical background. Section 3 introduces methodology used related work. In Section 4, we describe the methods for explain ability followed by fact sheet analysis in Section 5. Finally, Section 6 summarizes conclusions followed by limitations and future research directions.

II. LITERATURE REVIEW AND THEORETICAL BACKGROUND

A. Related studies for clickstream data

Clickstream data records all search paths of users and is a powerful source of information for consumer behaviors online [8]. However, using clickstream data involves some challenges because data is so disorganized that it is not easy to preprocess and analyze. Thus, more research is needed. The first article is that Moe [3] used clickstream data from e-commerce and took a helpful step toward integrating content variables into the browsing analysis, finally applying a post hoc clustering to identify the consumer search motivation. He further distinguished two goal-directed (directed buying and deliberation search) and two exploratory (hedonic search and knowledge building) search motivations. Deliberation search and knowledge building are beneficial because e-commerce recommended systems can help consumers find products to promote their search goals or support the knowledge-building process. Our study builds on the concept of Moe [3] by integrating content variables into the analysis of browsing and clustering clickstream data to identify consumer intention. According to Bucklin & Sismeiro [8], their results suggest that visitors' browsing behavior indicates their interest level in online buying by revealing their time investment at the site. Whereas Moe [3] used the percentage of pages per session as an analysis unit, our study uses a percentage of page time spent per session as an analysis unit. Furthermore, it employs the Decision tree model to identify each cluster pattern to map the customer journey model.

In addition to modeling the purchase/browsing behavior at an e-commerce website [3, 8, 9, 11], there are other applications for clickstream data. Basu [9] investigated the role of information search in online purchase decisions based on the type of good (experience and search goods). The result showed that search intensities for experience goods are at least three times lower than search goods. He also found that personalized recommendations affect online search and purchase decisions significantly. Furthermore, how recommendations affect search intensities largely depends on shoppers being able to determine match quality for every product they sample. Although a recommendation reduces the search for an experience good, it increases that for a search good. It is observed that recommendation systems play an essential role in consumers' purchase decisions as it considers user intention based on their search histories.

B. Customer journey

Customer experience is the customer's cognitive, emotional, social, and physical responses to a company [12]. Thus, enhancing and managing customer experience and interactions across touchpoints becomes a vital service research issue. Lemon & Verhoef [6] conceptualize customer experience as a customer journey with the firm over time. A customer journey model has been applied in the marketing field. Here, the focus is on consumers' decision processes, from being aware of a company to make a purchase. In this model, customers' behavior and experiences are analyzed according to a predefined onboarding process, structured in steps such as pre-purchase, purchase, and post-purchase [6] or awareness, consideration, preference, purchase, retention, and advocacy [13], often supported by customer relationship management (CRM) systems or web analytics .

According to Kiely [13], the six stages of the customer journey ix stage of customer journey represented below (Fig.1).

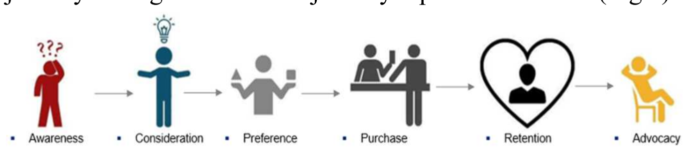

Fig. 1. Six stages of customer journey.

Awareness: It is built in the earliest stage of the customer journey and significantly influences the final choice and holistic experience of customers. The awareness-building activities during the initial contacts are critical touchpoints to increase the spending and satisfaction of customers [14].

Consideration: At this journey stage, the person actively seeks options. The second step is the user's intent to gather all the relevant information to make good decisions. Finally, the consumer processes his options in working memory, adding, or deleting as necessary.

Preference: Customer preferences in this stage are the likes, dislikes, and expectations that influence a customer's decision of whether they should purchase a product or service. Customers will be looking at what you can offer, the price, and more [15]. In addition, they frequently seek unique offers that serve their individual preferences and needs .

Purchase: Here, the customer is looking for a transactional landing page focusing on the product or service they want. It is characterized by behaviors such as choice, ordering, and payment [6]. In retailing and consumer products research, studies focus on the shopping experience [16]. With the myriad touch points and information overload, such as choice overload, purchase confidence, and decision satisfaction might also be considered. These may induce customers to stop searching and complete or defer the purchase.

Retention: This stage aims to build a long relationship with customers. Holding on to a customer cost far less than acquiring a new one [17]. This stage covers aspects of the customer's experience after purchase that relate to the brand or product/service itself. Thus, the product becomes a critical touch point in this stage [6]. Research on this stage has focused on the consumption experience; service recovery ; and decisions to return products or repurchase.

Advocacy: This is the last stage of the customer journey and the hardest to achieve. They give you great reviews, refer your friends or families, and generate the most powerful kind of marketing, word-of-mouth (WOM). The impact of word-of-mouth is more powerful because WOM is a low-cost and reliable way of disseminating information about products or services. Consumers search for information from knowledgeable others, such as friends or families, to make more informed decisions .

C. *Customer persona and Machine learning frameworks*

A meaningful customer journey map reflects customers' experiences [6]. By fusing persona and customer journey mapping data with an appropriate machine learning algorithm, one can predict customer behavior in real-time and rapidly create personalized suggestions or recommendations for every customer.

One can recall that personas are fictional characters that are heavily based on real customers. Businesses can now internalize the personas to focus on their work, refer to users by name, and consider how each product feature or machine learning capability they develop will impact their customers personally [18]. By analyzing vast volumes of data, machine learning creates clusters of personas with names, hobbies, and other lifestyle tidbits to bring them to life. It also narrates their objectives and goals relating to the product, thus building a complete and credible customer persona.

A good machine-learning algorithm blends domain expertise and expert intuition. It means that machine learning models must be designed with team members who deeply understand the business domain, the familiarity of data sets, and a deep intuition of customers' actual needs (reference).

III. METHODOLORY

A. *Data*

We use an online electronic retailer's clickstream data based out of Taiwan. This database contains 1,921,451 data from 2019.07 to 2020.03. Each of the data contains nine attributes as follows (Table 1)[1]. These data are collated into 282,514 sessions.

TABLE I. DATA ATTRIBUTES

NO			
A1	Session ID	**A6**	Time On Page
A2	Member NO	**A7**	Category
A3	Product ID	**A8**	Source Medium
A4	Browse Url	**A9**	Session ID_count
A5	Browse Date Time		

Among this data attributes, category is an important attribute for this study. Category contains 22 attributes as follows (Table2)[2].

TABLE II. CATEGORY ATTRIBUTES

NO			
B1	BUCKET	**B12**	NG
B2	CALLSERVICE	**B13**	OTHER
B3	CHECKACCOUNT	**B14**	PURCHASE
B4	CMCM	**B15**	SEARCH
B5	CRM	**B16**	SERVICE
B6	EDM	**B17**	SINGLEPROD
B7	ERRORMSG	**B18**	SMARTSELL
B8	HOME	**B19**	SMARTSELL BACKSTAGE
B9	LS3C	**B20**	TOPIC
B10	MONTHEVENT	**B21**	WISEMAN
B11	MULTIPRODUCT	**B22**	THANK

B. *Data preprocessing*

1) *Outliers*

We delete the data which is not the target of this experiment. For example, typically, the least purchasing process needs to browse through the product page, bucket page, purchase page, and thanks page, which is four pages. Hence, if this data's Session ID_count is less than four pages and Category does not contain Purchase or Thanks, we delete it. Furthermore , we use a box-and-whisker plot approach to remove the Session ID outlier. The data will be deleted if the data's Session ID_count value is defined as an outlier and the Category does not contain Purchase or Thanks.

2) *Category's Thanks and Purchase integration*

In this step, we integrated the attributes whose meaning is similar. Therefore, the Category's Thanks is replaced with Purchase because customers must not reach the Thanks page until they pass through the Purchase page. After emerging these two attributes, the count of Category's attributes became twenty-one.

3) *Session ID integration*

We integrated the related browsing behaviors by the following approach: If a Session ID's Category starts with Purchase, it means this Session ID is highly connected to the last one. To verify the connection, we checked these Session IDs, which were merged, and found that the product customer purchased was what they browsed during the last session. Finally, the total number of sessions analyzed is 111,941.

4) *Data Transpositionn*

[1] A detailed explanation of each information is in Appendix I

[2] A detailed explanation of category's attributes is in Appendix II.

Since, with our study, we want to explore how customers distribute their time in each browsing session by every Category attribute's time proportions, we summed up each Category attribute's browsing time in every Session_ID and transposed them into proportion. Then, the preprocessed data is presented as follows (Fig. 2). The horizontal axis represents each Category attribute's browsing time proportion, and the vertical axis represents every Session_ID.

	BUCKET	CALLSERVICE	CHECKACCOUNT	CMCM	CRM	EDM	ERRORMSG	HOME	LS3C	MONTHEVENT	...	NG	OTHER
0	0.177123443	0	0.003624009	0	0	0	0	0.114156285	0	0	...	0.088561721	0
1	0.613545817	0	0	0	0	0	0	0.139442231	0	0	...	0.175298805	0
2	0	0	0.026119403	0	0	0.017257463	0	0.00233209	0	0	...	0.07369403	0
3	0.054237288	0	0	0	0	0	0	0.084745763	0	0	...	0.038983051	0
4	0	0	0	0	0	0	0	0.02764977	0	0	...	0.912442396	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
111936	0.019243986	0	0.147766323	0	0	0.15395189	0	0.142955326	0	0	...	0	0
111937	0.036104114	0	0.345088161	0	0	0	0	0.147774979	0	0	...	0	0
111938	0	0	0.633093525	0	0	0	0	0.061151079	0	0	...	0	0
111939	0	0	0.884735202	0	0	0	0	0.115264798	0	0	...	0	0
111940	0	0	1	0	0	0	0	0	0	0	...	0	0

Fig. 2. Preprocessed data.

C. *Algorithms model*

1) K-means cluster

The clustering algorithm is a process of grouping data into a cluster. It contains data as similar as possible and different from other cluster objects [19]. K-means clustering is one algorithm described in detail by Hartigan [20]. This algorithm can partition data into several clusters in which each data belongs to the cluster with the nearest mean (cluster centroid), but it is sensitive to outliers. Thus, data should be preprocessed to delete outliers before using K-means.

2) Elbow method

K-means clustering requires determining the best count of clusters (K). This study used the Elbow method to determine the best count of clusters to achieve the requirement. The elbow method is one of the approaches to determining the number of clusters in a data set . Using the "elbow" as a cutoff point is a standard method in mathematical optimization to choose a point where diminishing returns are no longer worth the additional cost. In clustering, one should choose several clusters so that adding another cluster does not give much better modeling of the data.

3) Decision tree

The Decision tree model is a commonly used data mining method for classifying data based on multiple variables or developing prediction algorithms for a target variable. It can discover features and extract patterns in large databases important for discrimination and prediction. CART (Classification and Regression Trees) is one of the statistical algorithms for building decision trees. It is a form of binary recursive partitioning, which means each node can only be split into two new nodes .

IV. RESULT

A. *K-means clustering*

We used the elbow method to determine the best count of clusters. According to the result (Fig. 3), we determined that the count of clusters is four. Furthermore, we got each cluster's data counts and attributes' values. We organize each cluster's details, including data count and the average proportion of every attribute's browsing time (Table 3).

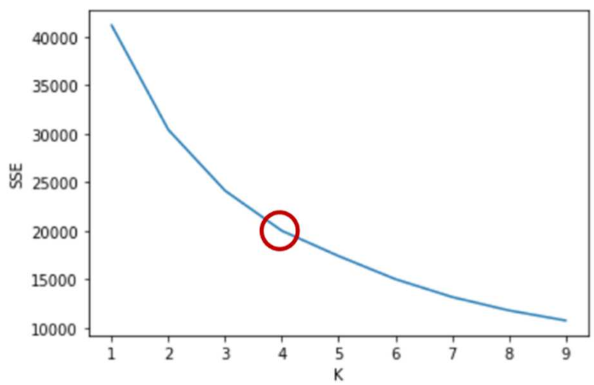

Fig. 3. T*he result of elbow method.*

TABLE III. EACH CLUSTER THE AVERAGE PROPORTION OF EVERY ATTRIBUTE'S BROWSING TIME

Cluster 1 (N = 39836)			
Category's attributes	**The average proportion of browsing time**	**Category's attributes**	**The average proportion of browsing time**
BUCKET	0.063384	**CALLSERVICE**	0.000003
CHECKACCOUNT	0.023230	**CMCM**	0.000047
CRM	0.002026	**EDM**	0.012511
ERRORMSG	0.000478	**HOME**	0.034723
LS3C	0.004031	**MONTHEVENT**	0.001307
MULTIPRODUCT	0.067145	**NG**	0.006538
OTHER	0.000427	**PURCHASE**	0.000285
SEARCH	0.042288	**SERVICE**	0.003683
SINGLEPROD	0.732767	**SMARTSELL**	0.000788
SMARTSELL BACKSTAGE	0.000304	**TOPIC**	0.002209
WISEMAN	0.001824		

Cluster 2 (N = 20684)			
Category's attributes	**The average proportion of browsing time**	**Category's attributes**	**The average proportion of browsing time**
BUCKET	0.022245	**CALLSERVICE**	0.000000
CHECKACCOUNT	0.022490	**CMCM**	0.000016
CRM	0.001819	**EDM**	0.014492
ERRORMSG	0.000546	**HOME**	0.080210
LS3C	0.003303	**MONTHEVENT**	0.001739
MULTIPRODUCT	0.636592	**NG**	0.008697
OTHER	0.001293	**PURCHASE**	0.000033
SEARCH	0.018169	**SERVICE**	0.002374
SINGLEPROD	0.180138	**SMARTSELL**	0.000065
SMARTSELL BACKSTAGE	0.000019	**TOPIC**	0.004087
WISEMAN	0.001674		

Cluster 3 (N = 38198)			
Category's attributes	**The average proportion of browsing time**	**Category's attributes**	**The average proportion of browsing time**
BUCKET	0.154995	**CALLSERVICE**	0.000159

CHECKACCOUNT	0.063646	**CMCM**	0.000295
CRM	0.015160	**EDM**	0.048302
ERRORMSG	0.001712	**HOME**	0.210577
LS3C	0.020239	**MONTHEVENT**	0.005416
MULTIPRODUCT	0.050202	**NG**	0.026906
OTHER	0.001750	**PURCHASE**	0.000163
SEARCH	0.144372	**SERVICE**	0.064193
SINGLEPROD	0.146178	**SMARTSELL**	0.005152
SMARTSELL BACKSTAGE	0.002940	**TOPIC**	0.010292
WISEMAN	0.009184		

B. *Cluster patterns*

To further analyze the four clusters and find each cluster's important attributes, we use the Decision Tree algorithm to define each cluster precisely. This study used the CART model to find each cluster's characteristics and gave them a definition. In other words, we found customer intention for each cluster. As follows, we represent these four decision trees of clusters.[3]

Cluster 4 (N =13223)			
Category's attributes	**The average proportion of browsing time**	**Category's attributes**	**The average proportion of browsing time**
BUCKET	0.031046	**CALLSERVICE**	0.000562
CHECKACCOUNT	0.674867	**CMCM**	0.000008
CRM	0.010629	**EDM**	0.010663
ERRORMSG	0.001005	**HOME**	0.095470
LS3C	0.006074	**MONTHEVENT**	0.000857
MULTIPRODUCT	0.021674	**NG**	0.005775
OTHER	0.000938	**PURCHASE**	0.000016
SEARCH	0.013567	**SERVICE**	0.066732
SINGLEPROD	0.052159	**SMARTSELL**	0.002773
SMARTSELL BACKSTAGE	0.000319	**TOPIC**	0.003834
WISEMAN	0.001032		

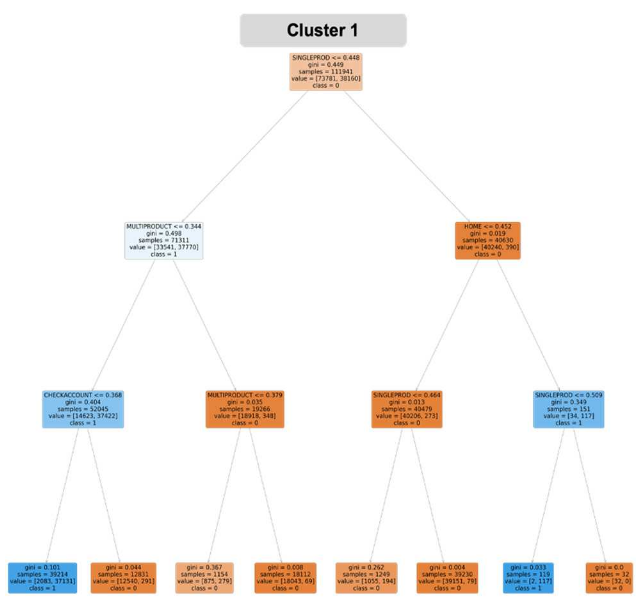

Fig. 4. *Decision Tree of Cluster 1.*

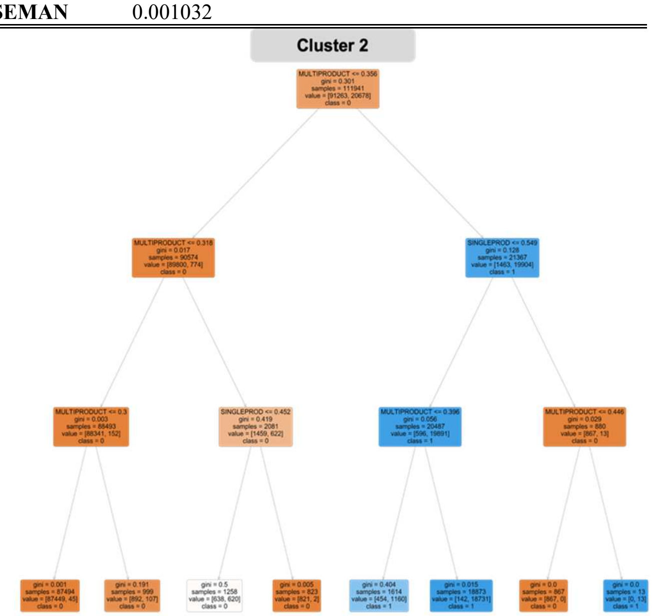

Fig. 5. *Decision Tree of Cluster 2.*

[3] Noted that Certain class = 1; other clusters' class = 0

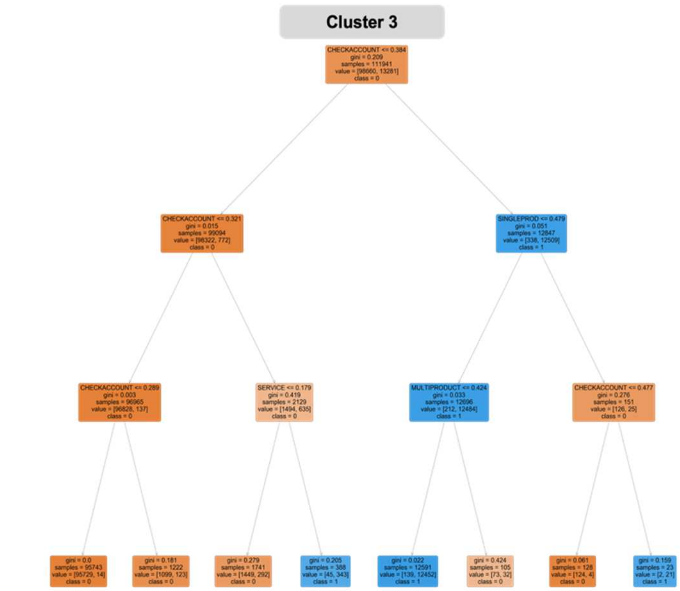

Fig. 6. *Decision Tree of Cluster 3*.

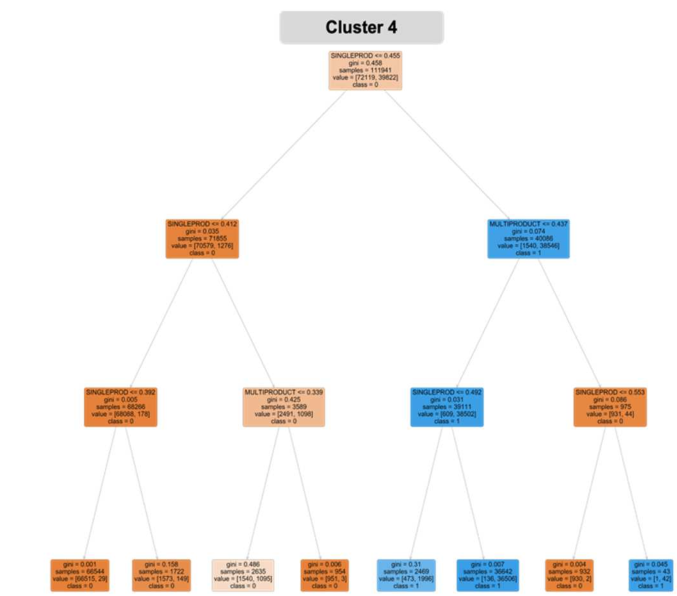

Fig. 7. *Decision Tree of Cluster 4*.

We generalize a pattern for each cluster and show as follows (Table 4).

TABLE IV. PATTERN OF EACH CLUSTER

Cluster 2		
Significant attributes of cluster 2 from Decision Tree	The average value of cluster 2's attribute	The average value of the other clusters' attribute
Multiproduct > 0.356	0.637	0.053
Singleproduct <= 0.549	0.180	0.389

Customers in Cluster 2 browse Multiproduct page on this e-commerce casually. They start to know about this e-commerce platform.

Cluster 3		
Significant attributes of cluster 3 from Decision Tree	The average value of cluster 3's attribute	The average value of the other clusters' attribute
Singleproduct <= 0.448	0.146	0.456
Multiproduct <= 0.344	0.050	0.219
Checkaccount <= 0.368	0.064	0.140

Besides these essential attributes, we also found that Cluster 3's BUCKET (0.154995), HOME (0.210577), SEARCH (0.144372), and EDM (0.048302) average proportion of browsing time is much higher than the other clusters from Table 3. Thus, we think that customers in Cluster 3 check the final discounts and prepare to drop an order.

Cluster 4		
Significant attributes of cluster 4 from Decision Tree	The average value of cluster 4's attribute	The average value of the other clusters' attribute
Checkaccount > 0.385	0.675	0.039
Singleproduct <= 0.478	0.052	0.390
Multiproduct <= 0.424	0.022	0.180

Customers in Cluster 4 already have a shopping list, and they prefer this e-commerce platform. So, they checked their member benefits and compared product prices to other stores.

C. *Managerial and Business Implications*

This study mapped four clusters to the customer journey model to combine the results with business theory and generate practical business marketing suggestions. According to Kiely [21], the customer journey has six stages (awareness, consideration, preference, purchase, retention, and advocacy). In addition, our result can map to pre-purchase and purchase situations (Fig 8).

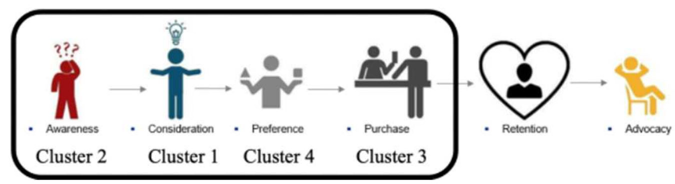

Fig. 8. *The cluster result for mapping customer journey.*

Then, we would represent marketing suggestions for these four stages. Finally, the marketing suggestions would be divided into marketing goals and approaches.

1) *Cutomer Journey-Awarness*

The awareness stage is mapped by cluster 2. Customers in this stage casually browse the Multiproduct page on this e-commerce platform and start to know this e-commerce platform. Therefore, this stage's marketing goal is to enhance customers' trust and give them more helpful information to push them into the next stage. Some studies have indicated that consumer trust is essential for e-commerce companies to drive consumer online transactions . In addition, trust brings another benefit: revisiting intention [21]. Once customers are willing to revisit the e-commerce platform, the possibility they move to the next stage will enhance.

Cluster 1		
Significant attributes of cluster 1 from Decision Tree	The average value of cluster 1's attribute	The average value of the other clusters' attribute
Singleproduct > 0.455	0.733	0.139
Multiproduct <= 0.437	0.067	0.213

Customers in Cluster 1 already have a shopping list, so they browse Singleproduct page on this e-commerce frequently since this company is their consideration.

To reach this goal, companies can put more information about the company's history, brand spirit, and even customer reviews on their website. These can help the new customer become familiar with this brand and build trust quickly.

2) *Cutomer Journey-Consideration*

The consideration stage is mapped by cluster 1. Customers in this stage already have a shopping list, so they browse the Single product page on this e-commerce frequently since this platform is their consideration. Hence, the marketing goal in this stage is to make customers know that this brand is different from others and then prefer this brand.

Introducing this brand's strengths, such as fast shipping and various goods, through Key Opinion Leaders is a viable method to reach the marketing goal. Key opinion leaders play a vital role in marketing communication, and they can provide informal consumption-related advice to others. Moreover, they usually tend to be respected sources of information that possess new and valuable insight while being personable and easy for potential customers to identify with. To summarize, promoting the brand's strengths through Key Opinion Leader is an excellent way to make customers differentiate between this brand and others at this stage, then prefer this brand.

3) *Cutomer Journey-Preference*

The Preference stage is mapped by cluster 4. Customers in this stage already have a shopping list, and they prefer this e-commerce platform. In the meantime, they are checking their member benefits and comparing the product's price to the other stores. This stage aims to give them a thought that purchasing on this platform can save lots of money than placing an order.

Many scholars have indicated price as a customer preference in the past, and most customers are price-conscious. Therefore, we can give customers customized coupons to satisfy customers' needs as this stage's marketing approach. A study has made experience to test a hypothesis that customized coupon campaigns have a positive exposure and redemption effect on customer purchases [22]. The results of this experience approved this hypothesis. Moreover, it also revealed that these coupons are more effective if they provide more discounts, are unexpected, and are positioned as specially selected for and customized to consumer preferences. Based on this research, we can induce giving customers customized coupons is an efficient marketing approach in this stage, and it can push customers to the next stage.

4) *Cutomer Journey-Purchase*

The Purchase stage is mapped by cluster 3. Customers in this stage check the final discounts and drop orders. The priority in this stage is to give customers a good after-sales impression. If we reach the goal, customers will move to the final stage and build long-term relationships with the company. However, the purchase stage is not the end, so customer retention is critical for companies. Some studies also revealed that customers might start a deeper connection with the brand through consumption and share their brand experience with other customers after purchasing.To build customers a good after-sale impression, giving purchasing points when customers make any purchase is an appropriate marketing approach. This approach has been in place for years. Several retailers gift their customers when a threshold spent amount is reached.

V. DISCUSSION

Previous research focused on the customer journey by taking one customer's all consumer behavior as one unit. In this study, we use each session of browsing behavior as one unit to prevent a situation in which a customer may have different intentions during different periods. Furthermore, this study mainly focuses on browsing time in each category to analyze customers' intentions. The results do show it is a feasible approach. We can analyze customers' intentions from a new perspective in the future.

The limitation of the research is that we mainly focus on the browsing time in each category. It is possible to make this analysis more precise and complete if we add other attributes to the model, such as browsing paths. Aadditionally clickstream data cannot map retention and advocacy of customer journey. Researchers can use the proposed method in different fields or set research units based on the proposed concept for future research.

ACKNOWLEDGMENT

Thanks to **Liang-xing Electronics** for providing browsing clickstream data.

REFERENCES

[1] H. (Alice) Li and P. K. Kannan, "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," Journal of Marketing Research, vol. 51, no. 1, pp. 40–56, Feb. 2014.

[2] M. Wedel and P. K. Kannan, "Marketing Analytics for Data-Rich Environments," Journal of Marketing, vol. 80, no. 6, pp. 97–121, Nov. 2016.

[3] W. W. Moe, "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream," Journal of Consumer Psychology, vol. 13, no. 1, pp. 29–39, Mar. 2003.

[4] D. Grewal, M. Levy, and V. Kumar, "Customer Experience Management in Retailing: An Organizing Framework," Journal of Retailing, vol. 85, no. 1, pp. 1–14, Mar. 2009.

[5] N. M. Puccinelli, R. C. Goodstein, D. Grewal, R. Price, P. Raghubir, and D. Stewart, "Customer Experience Management in Retailing: Understanding the Buying Process," Journal of Retailing, vol. 85, no. 1, pp. 15–30, Mar. 2009.

[6] K. N. Lemon and P. C. Verhoef, "Understanding Customer Experience Throughout the Customer Journey," Journal of Marketing, vol. 80, no. 6, pp. 69–96, Nov. 2016.

[7] S. Basu, "Information search in the internet markets: Experience versus search goods," Electronic Commerce Research and Applications, vol. 30, pp. 25–37, Jul. 2018.

[8] R. E. Bucklin and C. Sismeiro, "A Model of Web Site Browsing Behavior Estimated on Clickstream Data," Journal of Marketing Research, vol. 40, no. 3, pp. 249–267, Aug. 2003.

[9] A. L. Montgomery, S. Li, K. Srinivasan, and J. C. Liechty, "Modeling Online Browsing and Path Analysis Using Clickstream Data," Marketing Science, vol. 23, no. 4, pp. 579–595, Nov. 2004.

[10] A. A. Tudoran, "A machine learning approach to identifying decision-making styles for managing customer relationships," Electronic Markets, Jan. 2022.

[11] V. Singh, B. Nanavati, A. K. Kar, and A. Gupta, "How to Maximize Clicks for Display Advertisement in Digital Marketing? A Reinforcement Learning Approach," Information Systems Frontiers, Jul. 2022.

[12] P. C. Verhoef, K. N. Lemon, A. Parasuraman, A. Roggeveen, M. Tsiros, and L. A. Schlesinger, "Customer Experience Creation: Determinants,

Dynamics and Management Strategies," Journal of Retailing, vol. 85, no. 1, pp. 31–41, Mar. 2009.

[13] T. J. Kiely, "Why and How to Create a Customer Journey Map," Meltwater, May 20, 2020. https://www.meltwater.com/en/blog/customer-journey.

[14] M. Khanna, I. Jacob, and N. Yadav, "Identifying and analyzing touchpoints for building a higher education brand," Journal of Marketing for Higher Education, vol. 24, no. 1, pp. 122–143, Jan. 2014.

[15] A. Al Adwan, "E-marketing strategy: to improve customer preference for local brand over foreign brand in the era of a developing country," Innovative Marketing, vol. 15, no. 3, pp. 85–98, Sep. 2019.

[16] J. Baker, A. Parasuraman, D. Grewal, and G. B. Voss, "The Influence of Multiple Store Environment Cues on Perceived Merchandise Value and Patronage Intentions," Journal of Marketing, vol. 66, no. 2, pp. 120–141, Apr. 2002.

[17] S. Coyles and T. C. Gokey, "Customer retention is not enough," Journal of Consumer Marketing, vol. 22, no. 2, pp. 101–105, Mar. 2005.

[18] S.-S. Chen, B. Choubey, and V. Singh, "A neural network based price sensitive recommender model to predict customer choices based on price effect," Journal of Retailing and Consumer Services, vol. 61, p. 102573, Jul. 2021.

[19] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," IOP Conference Series: Materials Science and Engineering, vol. 336, p. 012017, Apr. 2018.

[20] J. Hartigan, Clustering Algorithms. 1975.

[21] V. Singh, I. Konovalova, A. Kar," When to choose ranked area integrals versus integrated gradient for explainable artificial intelligence–a comparison of algorithms," Benchmarking: An International Journal, Aug. 2022.

[22] R. Venkatesan and P. W. Farris, "Measuring and Managing Returns from Retailer-Customized Coupon Campaigns," Journal of Marketing, vol. 76, no. 1, pp. 76–94, Jan. 2012.