IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 22 May 2024, accepted 21 June 2024, date of publication 8 July 2024, date of current version 20 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424833

RESEARCH ARTICLE

AIDA-Based Customer Segmentation With User Journey Analysis for Wi-Fi Advertising System

SHI-YEN WONG, LEE-YENG ONG, (Senior Member, IEEE), AND MENG-CHEW LEOW, (Senior Member, IEEE)

Faculty of Information Science and Technology, Multimedia University, Bukit Beruang, Melaka 75450, Malaysia

Corresponding author: Lee-Yeng Ong (lyong@mmu.edu.my)

This work was supported by Telekom Malaysia Research and Development under Grant RDTC/221073 (MMUE/230002).

ABSTRACT Customer segmentation is an important aspect in aiding businesses to comprehensively understand their customer base and tailor their marketing strategies for optimal effectiveness. Traditional approaches to segmentation have predominantly concentrated on demographic factors and observable characteristics. However, these approaches have limitations that prevent them from capturing the intricate user journeys of each identified segment. Hence, this paper proposes an approach to customer segmentation using clustering algorithms, specifically the K-Means, BIRCH, and Gaussian Mixture Model on the dataset derived from the Wi-Fi advertising system, with a focus on tracking the user progression through the stages of the AIDA (Attention, Interest, Desire, Action) Model. This paper not only presents an AIDA-based metric designed for Wi-Fi advertising data, it also strives to measure the different stages in the user journey analysis. Through the combination of the AIDA Model and the clustering algorithms, the main objective is to gain a nuanced understanding of the distinct stages characterizing the user journey within each identified segment. This approach further incorporates a dynamic-characteristics range table to delineate the weak and strongly engaged behavioral traits, thereby demonstrating the efficacy of combining the AIDA Model with the clustering algorithms in unraveling nuanced insights into customer behavior across diverse stages of the user journey for each segmented group. Based on the detailed AIDA levels of each user segment, it suggests actionable insights for businesses to enhance marketing strategies by identifying which stages to emphasize, ultimately leading to improved campaign effectiveness and user satisfaction.

INDEX TERMS AIDA model, customer segmentation, user journey analysis, Wi-Fi advertising system.

I. INTRODUCTION

In the world of business, understanding customers and their needs is paramount for achieving success. Customer segmentation is a valuable approach that helps businesses gain insights into their customers and enhances profitability by tailoring appropriate marketing strategies [1]. However, many traditional segmentation approaches tend to narrow their focus solely on demographic factors such as age, gender, and income levels, as well as observable characteristics like purchasing history or geographic location. These methods often overlook the nuanced complexities of user journeys within each identified segment, which represent the path or steps that a person takes as they interact with a product, service, or system. Moreover, certain personal information may be inaccurately provided by the customers, which can result in unreliable data. The inherent limitations of relying solely on these factors impede their efficacy in comprehensively capturing the diverse and intricate paths that the customers undertake in their decision-making processes.

User journey analysis enables businesses to understand customers' experiences throughout the decision-making process. User journey analysis is the process of examining and understanding the various stages and touchpoints that a user goes through, allowing for the identification of areas of improvement, optimization of the customer experience, and development of effective marketing strategies [2]. The AIDA Model, introduced in 1898 and widely adopted in

The associate editor coordinating the review of this manuscript and approving it for publication was Shashikant Patil.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

the areas of traditional and online marketing, evaluates advertisement communication effectiveness and informs marketing communication strategies [3]. It can be broken down to multiple stages that customers traversed before taking action, shedding light on the factors influencing their decisions [4].

User journey analysis and the AIDA Model are closely connected, as the AIDA Model provides a framework for understanding the customer decision-making process across its four stages (Attention, Interest, Desire, and Action), while the user journey analysis enables businesses to analyze user behaviors at each of these stages. By investigating the user journey through the AIDA Model and employing clustering algorithms, including K-Means, BIRCH, and Gaussian Mixture Model, this study aims to demonstrate the combination of these approaches to gain deeper insights into customer behaviors and preferences for Wi-Fi advertising campaigns.

While existing research has demonstrated the importance of customer segmentation and user journey analysis in different domains, there remains a significant gap in applying these techniques specifically to Wi-Fi advertising domain. Moreover, the integration of the AIDA Model with clustering algorithms for in-depth analysis of user journeys in this specific context remains largely unexplored. This study addresses these gaps by proposing an AIDA-based metric utilising the Wi-Fi advertising system to facilitate a more comprehensive understanding of the user behavior. Since traditional segmentation methods often lack the ability to gauge the engagement levels of the customers throughout their journey, the application of the AIDA Model serves to understand the engagement of customers at various stages of their journey, providing a more holistic perspective for businesses to tailor their marketing strategies effectively.

This paper is organized into several sections. It begins with an introduction, followed by a literature review in Section II, which examines existing research and theories relevant to the topic under investigation. Section III outlines the proposed framework, while Section IV presents the obtained results. In Section V, the discussion interprets the findings, and the last section explains the key findings, conclusion and suggestions for future research.

II. LITERATURE REVIEW

A. WI-FI ADVERTISING SYSTEM

With the evolution and advancement of the internet and smart devices such as smartphones, tablets, and laptops, internet advertising has become one of the crucial ways for businesses to deliver marketing messages to the customers. One example of internet advertising is the Wi-Fi advertising on free public networks, where users view advertisements on their devices in exchange for wireless internet access [5].

The Wi-Fi advertising system is an advertising platform that enables businesses to showcase their product, services, or brand to the public. Typically, the public Wi-Fi advertising system follows a similar procedure outlined in Fig. 1. Initially, users log into the system, and they encounter a non-skippable advertisement. Following the mandatory viewing time, users have the option to either continue watching the advertisement or to skip it. Subsequently, users can proceed to connect to the free Wi-Fi or visit the advertisement website.

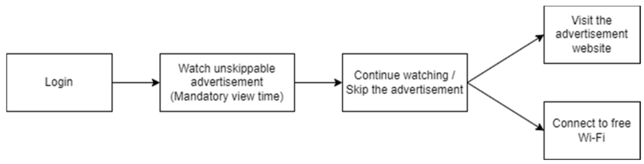

FIGURE 1. The general procedure of the Wi-Fi advertising system.

One of the most common Wi-Fi advertising systems in Malaysia is provided by Telekom Malaysia (TM) under the SSID "wifi@unifi". This system offers free Wi-Fi access for over 7,000 hotspots distributed across public locations, including shopping malls, business centers, universities, and tourist spots [6]. When connecting to the public hotspots for free Wi-Fi access, users are required to log in using their personal information. Subsequently, a non-skippable 10-second advertisement is displayed to users before they gain access to the Wi-Fi network. Another example is the "Wimedia" service in Yerevan, Armenia. It is a Wi-Fi advertising platform that provides free internet access across more than 600 locations. It is now operating in five countries, connecting over 2000 Wi-Fi routers and access points to its platform. Presently, the service encompasses the majority of Wi-Fi hotspots in Yerevan, strategically located in diverse public areas and businesses such as shopping malls, airports, opera squares, universities, buses, cafes and restaurants [7]. Before users can access the internet via the Wimedia hotspot, they must watch a short advertising video.

Wi-Fi advertising leverages the Wi-Fi networks to reach the public, displaying advertisements on various devices such as smartphones, tablets, and laptops in exchange for free Wi-Fi internet access. Since this medium effectively puts advertisements to a wide audience, understanding user behavior towards Wi-Fi advertisements becomes pivotal in understanding its efficacy as a marketing medium. To enhance the effectiveness of targeted advertising within the Wi-Fi advertising systems, customer segmentation is one of the approaches for refining or delivering relevant advertisements based on user preferences and behaviors.

B. CUSTOMER SEGMENTATION

Customer segmentation is a crucial procedure involving the classification of a customer base into distinct groups based on their various characteristics such as demographics, behaviors, and needs [8]. It is a common practice across various domains such as banking, e-commerce, marketing, and advertising, etc. This procedure allows businesses to gain a deeper understanding of their customers and develops tailored marketing strategies to enhance their customer satisfaction, retention, and hence increasing revenue.

An existing work in [9] presents an unsupervised deep learning approach for efficient customer segmentation in digital marketing through the application of AI-based modelling techniques. It proposes a combination of Self-Organizing Map with an Improved Social Spider Optimization Approach to effectively segment customers based on their behavioral patterns. A Modified Social Spider Optimization Model is used in the feature engineering process to select relevant customer behavioral features. Customers are then clustered using a Self-Organizing Neural Network (SONN) based on these behavioral patterns. Deep Neural Networks (DNN) further classify customers into segments based on the SONN clusters. The experimental results demonstrate high clustering and segmentation capability of the proposed model. It shows the effectiveness of the proposed approach in customer segmentation, which improves business profits in the marketing domain by enabling companies to offer the right products to the right customers based on their behavioral patterns and purchasing preferences.

In the domain of social media advertising, a study in [10] investigates the effectiveness of customer segmentation for identifying different customer segments based on their perceptions of social media advertising features. Through cluster analysis, the study identifies three distinct customer segments in terms of individual traits and purchase intentions. Based on the results, the study found that the factors influencing purchase intentions for products presented in social media advertising are different across each segment. This finding underscores the usefulness of customer segmentation in personalizing social media advertising campaigns to target audiences, helping businesses to gain competitive advantage in dynamic markets by delivering targeted messages to different segments more effectively.

In [11], the study proposes a new RFI (Recency, Frequency, and Interest) Model for analyzing audience behavior towards advertisements in the Wi-Fi advertising. Once the recency, frequency, and interest factors of the RFI Model are calculated, different clustering algorithms are employed to perform behavioral segmentation of the audience. The work applies the RFI Model and customer segmentation techniques to two different Wi-Fi advertising attributes. Through experimentation with Wi-Fi advertising attributes, the efficacy of the RFI Model and audience behavioral segmentation is demonstrated. The work successfully interprets various segmented behavioral patterns such as one-time audience behavior, weakly engaged audience behavior, and strongly engaged audience behavior.

The studies presented in [9], [10], and [11] employ different methodologies and model focus on distinct aspects of customer behavior to address customer segmentation. The work in [9] used a sophisticated unsupervised deep learning approach to segment customers based on behavioral data. This study stands out for its use of advanced machine learning techniques in handling complex behavioural patterns for digital marketing applications. On the other hand, the study in [10] applies cluster analysis to segment customers based on their perceptions and individual traits in the context of social media advertising. The method utilized in this study is more straightforward but provides valuable insights into the factors influencing purchase intentions across different segments. Meanwhile, the study in [11] introduces a new RFI Model and employs various clustering algorithms to segment the audience specifically within the Wi-Fi advertising domain. Although this approach is less generalizable due to its focus on a niche area, it effectively interprets different behavioral patterns, demonstrating the efficacy of the RFI Model in understanding audience engagement levels.

While many existing studies show the advantages of customer segmentation in various domains, the majority tend to solely analyze segmented groups without delving into further analysis. Therefore, this paper goes beyond the segmentation by conducting user journey analysis, providing a more comprehensive understanding of customer behaviors.

C. USER JOURNEY ANALYSIS

The user journey analysis is a methodology employed by businesses to comprehend customer interactions with their products or services. This process entails tracking entire customer experience, from initial contact to the final interaction [12]. In product design, this is referred to as a "User Journey Map", documenting user interactions and touchpoints with a specific product. In marketing and service design, it is termed a "Customer Journey Map", presenting a visual narrative of the customer interactions and touchpoints with a brand. Despite nomenclature differences, both approaches are fundamentally similar, emphasizing stages and touchpoints in the customer's interaction with products or services and illustrating the journey from the beginning to the end [2]. Over the years, many studies have been conducted to analyze and map out the user journey in various fields.

Existing work in [13] introduces a new approach to analyze customer journeys using process mining techniques. The finding demonstrates the usefulness of user journey analysis by showing how process mining techniques can effectively analyze customer journeys, preprocess web logs, and optimize personalized recommendations. It highlights how user journey analysis can lead to actionable insights for improving user experience and customizing recommendations.

In [14], the study assesses how customer co-creation behaviors, customer responses, and customer experiential values influence the degree of customer journey satisfaction. The findings of this study underscore the importance of understanding factors affecting customer journey satisfaction and how user journey analysis can help in segmenting customers to analyze these factors effectively. It shows that user journey analysis can provide insights into the varying impact of factors across different customer segments.

A study in [15] addresses the complexity of customer behaviors by segmenting customers according to their utilization of touchpoints in the customer journey, analyzing covariates, incorporating the influence of mobile devices, and examining customer behavior across these segmented groups. The study identified five consistent customer journey segments across various product categories. It emphasizes the usefulness of understanding customer journey patterns and behaviors, which can inform strategies for enhancing satisfaction, loyalty, and inspiration across these segments.

According to [16], the study investigates the impact of smart technologies on the customer journey within tourism attractions, focusing on three-phase visit cycle from the customer behavior perspective based on the customer journey process model. The findings of this study showcases how user journey analysis can interpret the significant influence of smart technologies on different phases of the customer journey. It emphasizes the role of user journey analysis in understanding how technological advancements shape user experiences and interactions, providing insights into the impact of smart technologies.

Based on the results of these existing works, the findings from each work demonstrate the usefulness of the user journey analysis in gaining insights into customer behaviors, preferences, and satisfaction across various contexts. Therefore, in a similar way, it is valuable to understand the user journey within the Wi-Fi advertising system.

D. AIDA MODEL

Since user journey analysis is a method of examining how customers interact with a product or service throughout their entire experience from initial awareness to the outcome, the AIDA Model is another useful model which helps to understand how customers move through each stage along their journey. The communication theory known as Attention, Interest, Desire, and Action (AIDA) Model explains the psychological and behavioral journey from the initial exposure to marketing communication to eventual consumption behavior as influenced by the media exposure [17]. This model identifies four cognitive stages—Attention, Interest, Desire, and Action—that individuals undergo when they are exposed to a new idea or product [18]. The AIDA Model posits that customers progress through cognitive and affective stages before leading to behavioral stages, as illustrated in Fig. 2.

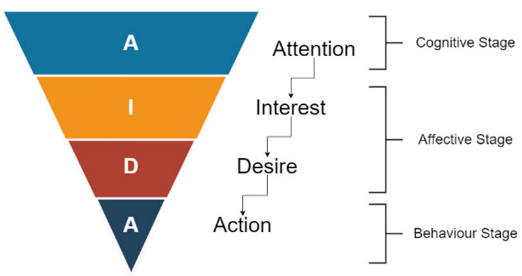

FIGURE 2. The stages of the AIDA Model.

The four stages represent the psychological process a customer goes through when making a purchase decision. The first stage is *attention*, which involves getting the customer's attention and creating awareness about the brand or product in the customer's mind, often through messages on social media or other platforms. Once attention is gained, the focus shifts to generating *interest*. The messages spark curiosity and further observation, leading to initial buying interest. At the *desire* stage, customers become convinced that the product or service will meet their needs, leading to the development of motivation and a desire for the product based on the interest that has been generated. Finally, the *action* stage represents the last action, where the customer exhibits a strong desire to make the actual purchase decision [19], [20], [21]. Typically, the AIDA stages outline a sequential process, with each stage building upon the previous stage. However, a customer does not necessarily move linearly through all the stages. For example, when attention, interest, and desire have been established, the customers may fail to take the ultimate step of making a purchase or taking the final action. To complete the process, a strong motivation is necessary to prompt the customer to convert the desire into action [21].

Despite its long history, the AIDA Model remains relevant in digital marketing, extending its application to social media platforms [18]. Researchers suggest leveraging social media's potential for small businesses and emphasize the importance of a thorough understanding of the AIDA Model in enhancing social media marketing efforts. Additionally, the AIDA theory has found practical use in e-banking services, where the China Construction Bank proposed a marketing model to improve customer experience through practical-teaching innovation and job competency enhancement. This study offers valuable insights into tailored marketing strategies using the AIDA Model to meet specific customer needs and desires [18].

Although the AIDA Model has been widely applied, especially in marketing, there is a lack of its application that is combined with user journey analysis. Furthermore, to the best of the author's knowledge, the AIDA Model has yet to be applied in the Wi-Fi advertising domain. Thus, by leveraging the principles of the AIDA Model with user journey analysis, Wi-Fi advertising platforms can gain deeper insights into the progression of users through the AIDA stages, thereby refining their marketing strategies and optimizing campaign performance.

In addition, the AIDA Model can be complemented by the unsupervised learning techniques, particularly clustering methods, to gain deeper insights into the different user journeys. This uncovers distinct user segments that exhibit comparable characteristics or behaviors across the AIDA stages.

E. UNSUPERVISED LEARNING – CLUSTERING METHODS

Unsupervised learning is a machine learning method that operates with unlabelled data to discern patterns and relationships within datasets. In this approach, algorithms autonomously categorize, label, and group data points without external guidance, eliminating the need for user supervision. In contrast to supervised learning, where algorithms rely on labelled data for analysis, clustering, and predictions, unsupervised learning operates without the need for pre-labelled data, allowing for exploration and pattern recognition in unlabelled datasets [22].

A common application of unsupervised learning involves grouping datasets into clusters based on similarities among data points. This process, known as clustering, segments data and uncovers hidden patterns or connections that may not be immediately apparent to humans [23]. Over the past decade, numerous clustering algorithms have been introduced and widely utilized across various fields. This paper focuses on three major types of clustering algorithms, which are K-Means (centroid-based), BIRCH (hierarchical-based), and Gaussian Mixture Model (model-based). While the algorithms differ, they belong to unsupervised learning as they uncover structures in the data without the need for predefined labels.

K-Means clustering is a popular unsupervised learning algorithm used for partitioning an unlabelled dataset by assigning data points to the nearest neighbour. It groups the data into K clusters based on their similarities, with K being an integer value [24]. BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) is a hierarchical clustering algorithm tailored for large datasets. It constructs a clustering feature (CF) tree incrementally, representing subclusters or a compressed data version [25]. The tree is dynamically reconstructed with incoming data, enhancing compression and facilitating efficient clustering at the leaf nodes [26]. The Gaussian Mixture Model (GMM) uses the Expectation-Maximization (EM) algorithm to create ellipsoidal groups based on probability density estimations. In contrast to K-Means, GMM characterizes each cluster as a Gaussian distribution, incorporating both the mean and covariance in the model [27].

In the context of the Wi-Fi advertising system, unsupervised learning helps identify and understand user behaviors and preferences without using pre-existing labels on the datasets. By using clustering techniques, it groups users effectively based on similarities in their behavioral patterns and interactions across the AIDA stages within the Wi-Fi advertising system.

F. EVALUATION METHODS

Determining the optimal number of clusters and evaluating cluster quality are important in applying clustering algorithms. This paper uses four popular evaluation metrics, which are the Elbow Method, Silhouette Score, Calinski-Harabasz Index (CH Index), and Dunn Index to assess the performance comparison among different clustering algorithms.

The Elbow Method determines the optimal number of clusters in a dataset by plotting within-cluster sum of squares (WCSS) or sum of squared errors (SSE) against the number of clusters. An "elbow" in the plot indicates the point at which adding more clusters does not significantly decrease WCSS/SSE, suggesting the optimal number of clusters has been reached [28].

The Silhouette Score is a metric that measure how similar each data point is to the others in its cluster versus other clusters. It combines cohesion, the closeness of points within a cluster, and separation, the distance between clusters. The score ranges from -1 to 1, with higher values indicating points that are well-matched to their assigned cluster and poorly matched to the neighbouring clusters [29].

The Calinski-Harabasz (CH) Index evaluates clustering performance based on the ratio of between-cluster variance to within-cluster variance [22]. It uses the intra-cluster covariance, W(K), to measure tightness of clusters and inter-cluster covariance, B(K), to measure dispersion between clusters. Higher values indicate clusters that are tightly packed within and well separated between, which is desirable [30].

The Dunn Index evaluates clustering performance by comparing inter-cluster separation to intra-cluster compactness. It calculates the ratio of the minimum distance between clusters to the maximum cluster diameter. Higher values indicate clusters that are compact within and well separated between, indicating better defined clusters [31].

III. PROPOSED FRAMEWORK

In this section, details about the dataset and the research methodology, along with the proposed framework, are explained. The processes in the proposed framework are illustrated in Fig. 3. The experiment pipeline consists of four key elements: data pre-processing, AIDA modelling, clustering, and user journey analysis.

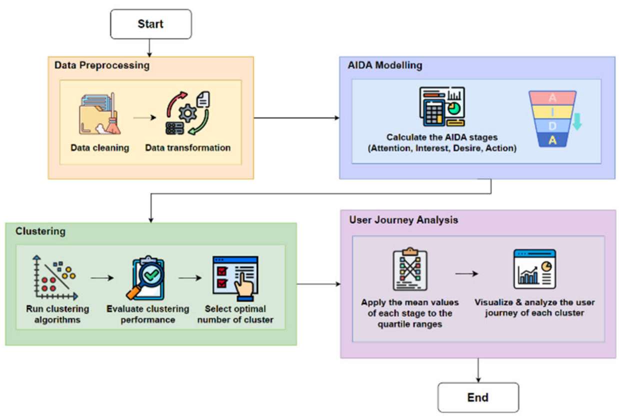

FIGURE 3. **Proposed framework.**

A. DATASET

The dataset was collected from a Wi-Fi advertising system in Malaysia over a three-month period, with information about the Wi-Fi subscribers referred to as the "user" as well as the details on the advertising campaigns. For each user, attributes like *id*, *age*, *email*, *gender*, *location ID*, and *device type* are provided to characterize the users engaging with the campaigns. The campaigns are identified by details like *ad ID*, *package ID*, and *track ID t*o differentiate between advertising campaigns. In addition, the dataset captures behavioral details, such as the time when a user plays or skips the advertisement.

There were a total number of over 667,335 users and a total of 2,772,800 records gathered from July to September 2023. The dataset contains records for two campaigns, which are identified as Campaign 2 and Campaign 3. Both campaigns were used to examine the user journeys based on the AIDA stages. To fairly examine the user journeys and behavioral patterns in relation to the advertising campaigns, the AIDA modelling and clustering process were conducted based on the criteria outlined in the following sections.

The dataset contains several key attributes as shown in the snippet of Fig. 4. *campaign_id* refers to the unique identity of the advertisement, *play_time* refers to the time when the user plays the advertisement, *click_time* refers to the time when the user visits the advertisement, *skip_time* refers to the time when the user skips the advertisement, and *proceed_time* refers to the time when the users directly access the Wi-Fi network.

	campaign_id	play_time	click_time	skip_time	proceed_time	audience_gender	audience_age	audience_email
2156879	3	2023-09-14 18:38:25	2023-09-14 18:39:15	2023-09-14 18:39:15	NaT	F	3	***********@*********
1218525	3	2023-08-19 21:43:13	NaT	2023-08-19 21:43:40	2023-08-19 21:43:41	M	4	************@********
195272	2	2023-07-13 12:25:43	NaT	NaT	NaT	F	5	****************@*********
938068	3	2023-08-12 16:23:17	NaT	NaT	2023-08-12 16:23:46	M	3	************@*********
467458	2	2023-07-30 19:26:16	NaT	NaT	2023-07-30 19:26:37	M	4	***************@*********

FIGURE 4. **Snippet of the dataset.**

Besides, the user's interaction in Wi-Fi advertising system in the collected data are similar to that as illustrated in Fig. 1. The process begins with the user entering their personal information. Subsequently, users are required to watch a 10-second mandatory non-skippable advertisement. Following this period, users can decide whether to continue watching or to skip the advertisement. Once the user completes or skips the advertisement, they are presented with two choices of actions. They can click on the ''Visit'' button, directing them to the advertisement website or the users can choose the ''Access'' button, providing them immediate access to the Wi-Fi network.

B. DATA PRE-PROCESSING

The first step in analyzing the Wi-Fi advertising dataset was data pre-processing, with two primary tasks: data cleaning and data transformation. Data cleaning includes the removal of redundant columns, handling missing values, and renaming columns. Meanwhile, data transformation involves converting the data types. For example, *play_time*, *click_time*, *skip_time*, and *proceed_time* were initially in string format. They were converted into date-time format to facilitate diverse calculations, such as determining the *view duration* and *extra view duration* beyond the mandatory time of the advertisements. Additionally, to ensure data privacy, confidential information in the original dataset such as *audience_email* has been masked through data anonymization techniques. This process safeguards sensitive information while allowing for meaningful analysis and insights to be derived from the dataset.

C. THE AIDA MODELLING

After the data pre-processing, the next step is to apply the AIDA Model to the Wi-Fi advertising dataset. This process involves calculating metrics specific to each stage of the model. The calculations for AIDA stages (attention, interest, desire, and action) are performed by assessing the user behaviors in relation to each advertising campaign.

The AIDA Model begins with the *attention* stage, where users first encounter the campaign while attempting to connect to the Wi-Fi network. The *attention* metric is proposed in (1), which calculates the frequency of views for each user for a campaign (*play_time*). A higher number of views indicates a greater level of attention. This is because when users try to connect to the Wi-Fi network, they are first presented with the campaign, which immediately grabs their attention and creates awareness about the brand or product being promoted. By displaying the campaign at this initial touchpoint, the user's focus is captured. This exposure helps establish the brand or product in the users' consciousness, fulfilling the crucial attention phase of the AIDA Model.

$$\text{Attention} = n(p) \tag{1}$$

where $n(p)$ is the frequency of which a user plays the advertisement (*play_time*) for each campaign.

Next comes the *interest* stage, where users show their curiosity to watch the advertisement further beyond the mandatory watch time. The *interest* metric is calculated as in (2), which is evaluated based on the *extra view duration*spent on a specific campaign. It is calculated by subtracting the *total mandatory view time* from the *total view duration*. The *mandatory view time* refers to the non-skippable view duration, which is pre-configured to 10 seconds. The *view duration* encompasses the time when the advertisement begins playing (*play_time*) until the user either skips the advertisement (*skip_time*), visits the advertisement website (*click_time*), or proceeds to connect to the Wi-Fi network (*proceed_time*). A longer *extra view duration* suggests a higher level of interest. This *interest* metric describes users who were initially drawn to the campaign, exhibiting a heightened curiosity, prompting them to invest additional time beyond the mandatory viewing period. By extending their view duration beyond the mandatory time, users showed their willingness to explore deeper into the content, suggesting a growing interest in what the campaign has to offer.

$$Interest = \sum (vd - mt) \tag{2}$$

where

- *vd* refers to the view duration (seconds) by subtracting *skip_time* or *click_time* or *proceed_time* with *play_time*
- *mt* is the mandatory view time of 10 seconds.

Once the user becomes interested, the *desire* stage comes into play, where users are convinced that the product or service will fulfill their needs, leading to a strong motivation and desire to acquire the product or service based on the interest that has been generated. The *desire* metric is derived in (3), which captures the transition by calculating the frequency ratio of the user who watched beyond the mandatory view time to the frequency of their views on the campaign (*attention*). This metric reflects the relationship between the AIDA stages that has been stated in prior studies [19], [20], [21], [32]. *Desire* arises from *interest*, which is established when *attention* is captured and sustained. This stage is intended to transform the *interest* into *desire* [33]. As users may watch the campaign repeatedly, if they repeatedly watch the campaign beyond the mandatory view time, it indicates a stronger motivation and desire to acquire the product or service being advertised. A higher ratio signifies a greater level of motivation and desire. By employing this metric, the *desire* stage aims to reveal how effectively the campaign has converted user interest into a strong desire for the product or service.

$$Desire(Ratio) = \frac{n(vd > 10)}{n(p)} \tag{3}$$

where

- $n(vd > 10)$ represents the frequency when the view duration (vd) is greater than 10 seconds for a user in the campaign
- $n(p)$ is the frequency when a user plays the advertisement (*play_time*) for each campaign.

Finally, the *action* stage is the last phase, where the user takes the concrete steps towards acquiring the product or service, demonstrating their strong desire to take the actual action. This action manifests as users are visiting the advertisement website, referring to the frequency of user website visits on each campaign, as shown in (4). This frequency is indicated by *click_time*, with each visit corresponding to a click on the "Visit" button. This is recognized as the *action* stage because visiting the advertisement website represents the final step within this dataset, aligning with the objective of advertising in the campaigns. A higher count of visits suggests a more substantial level of *action* made by the user, showing the effectiveness of the campaign in motivating users to take the final step.

$$\text{Action} = n(v) \tag{4}$$

where $n\,(v)$ is the frequency when a user visits the advertisement website (*click_time*).

D. CLUSTERING AND PARAMETER SETTING

Rather than analyzing the user journey across all advertisement campaigns, clustering algorithms (K-Means, BIRCH, and Gaussian Mixture Model) were applied to explore the distinct patterns within each specific campaign. In this paper, these clustering algorithms were applied on transformed data rather than the raw data. Specifically, the clustering process was applied after the generation of the AIDA stages. By using the transformed data, it ensured that the data was appropriately prepared and structured for the clustering algorithms to capture the relevant patterns and user behaviors from the Wi-Fi advertising system. This provides more granular insights into how different user segments responded to each campaign.

The three clustering algorithms chosen for this study were selected based on their unique strengths. K-Means is one of the top clustering algorithms which known for its simplicity and efficiency [34]. The BIRCH was selected for its capability to handle large datasets and its efficient memory usage [25], while GMM was included for its effectiveness and efficiency [35]. In configuring K-Means and GMM, parameters such as the number of clusters is considered from 2 to 10, maximum iterations are set to 50 for convergence, and a fixed random state is set to 100 for determining the cluster structure, reliability, and results reproducibility across different runs. Conversely, the parameter setup for BIRCH focuses mainly on the number of clusters, which is also considered from 2 to 10, as it is designed to incrementally builds hierarchical clusters and is not dependent on iterative refinement or random initializations due to its incremental nature.

Once the dataset was clustered, the clustering results were evaluated by using various evaluation methods, including the Elbow Method, Silhouette Score, CH Index and Dunn Index, following established practices from previous research [11]. This evaluation process aids in determining the effectiveness of the respective clustering algorithm and helps to identify the optimal number of clusters and the more suitable clustering algorithm for the different datasets.

IV. RESULTS

Following the completion of clustering and evaluation processes, the optimal number of clusters was determined by comparing and selecting based on results obtained from the four evaluation methods. Table 1 highlighted the optimal number of clusters for each algorithm in Campaign 2 and Campaign 3.

In fact, different clustering algorithms generated different cluster assignments on the same dataset due to different mathematical approaches in dividing the data points into groups. Therefore, it was expected that the different algorithms will lead to dissimilar clustering outcomes, even when analyzing identical input data. Based on the results, the optimal number of clusters for each clustering algorithm was determined based on the frequency of occurrence and their respective scores. Based on the frequency occurrence in Table 1, the optimal number of clusters of Campaign 2 for K-Means, BIRCH, and GMM were 5, 3, and 4 respectively. While in

TABLE 1. The optimal number of clusters in both campaigns.

Campaign	Evaluation Methods	K-Means	BIRCH	GMM
2	Elbow Method	4	**3**	**4**
	Silhouette Score	**5**	**3**	**4**
	CH Index	**5**	**3**	**4**
	Dunn Index	3	**3**	**4**
3	Elbow Method	4	**4**	**4**
	Silhouette Score	**5**	3	**4**
	CH Index	**5**	**4**	**4**
	Dunn Index	**5**	3	**4**

* The blue highlighted value refers to the mode of occurrence that is selected as the optimal number of clusters respectively.

Campaign 3, the optimal number of clusters were 5, 4, and 4 respectively.

Notably, when applying the BIRCH algorithm to Campaign 3, both the Elbow Method and the CH Index indicated that the optimal number of clusters was 4, while the Silhouette Score and Dunn Index suggested 3 clusters. In this scenario, a comparative analysis of the scores for each evaluation method was implemented. By comparing the scores for both number of clusters, the Silhouette Score for 3 clusters and 4 clusters are 0.94 and 0.89 respectively, while the Dunn Index are 0.44 and 0.24. Both scores exhibited only a slight decrement between 3 clusters and 4 clusters. Additionally, it was observed that the Elbow Method showed the most substantial reduction when transitioning from 3 clusters (1,004,617.10) to 4 clusters (647,341.21). The CH Index displayed the most significant increment from 3 clusters (55,140.06) to 4 clusters (118,505.72). Therefore, 4 was selected as the optimal number of clusters for the BIRCH algorithm in Campaign 3 based on the significant gap in the Elbow Method and CH Index.

With the optimal number of clusters selected for each clustering algorithm, the scores of the selected optimal number of clusters were compared in Table 2 to identify the best-performing clustering algorithms. The bold values in blue refer to the best performance among all the clustering algorithms used. Additionally, Fig. 5, Fig. 6, Fig. 7, and Fig. 8 depict various line charts that facilitate comparison of score performance across each evaluation method based on Table 2.

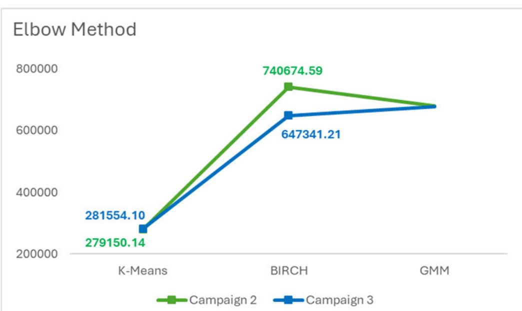

FIGURE 5. Score comparison for Elbow Method.

TABLE 2. Comparison of the clustering performance scores.

Campaign	Evaluation Methods	K-Means (5 Clusters)	BIRCH (3 Clusters)	GMM (4 Clusters)
2	Elbow Method ↓	**279150.14**	740674.59	679429.71
	Silhouette Score ↑	0.79	**0.91**	0.63
	CH Index ↑	**314560.16**	133276.60	106875.47
	Dunn Index ↑	0.06	**0.51**	0.11

Campaign	Evaluation Methods	K-Means (5 Clusters)	BIRCH (4 Clusters)	GMM (4 Clusters)
3	Elbow Method ↓	**281554.10**	647341.21	677378.63
	Silhouette Score ↑	0.79	**0.89**	0.63
	CH Index ↑	**312848.82**	118505.72	108313.00
	Dunn Index ↑	0.06	**0.24**	0.11

* The downward arrows (↓) indicate that the lower the score, the better the performance. Conversely, the upward arrows (↑) indicate that the higher the score, the better the performance. The bold value in blue refers to the best performance among all clustering algorithms respectively.

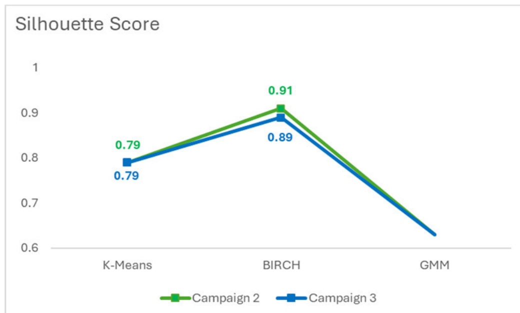

FIGURE 6. Score comparison for Silhouette Score.

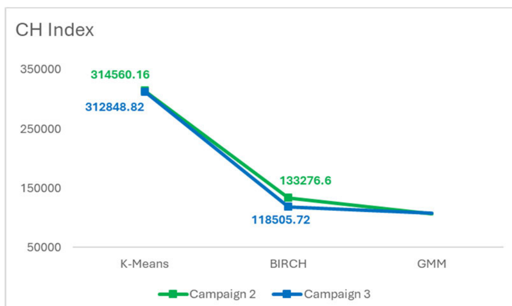

FIGURE 7. Score comparison for CH Index.

From Table 2, Fig. 6, and Fig. 8, BIRCH exhibited slightly better performance by achieving the highest scores according to the Silhouette Score and Dunn Index in both campaigns. However, based on the performance scores in Table 2, Fig. 5, and Fig. 7, K-Means significantly outperformed BIRCH and GMM in terms of the Elbow Method in both campaigns, obtaining the lowest score, which indicates better performance. Additionally, K-Means also significantly outperformed BIRCH and GMM in the CH Index by

TABLE 3. Mean of AIDA stages in each cluster of both campaigns.

Campaign	Cluster	Total Users	Attention	Interest	Desire	Action
2	0	225701	2.3230	48.8032	0.9722	0.0828
	1	6574	55.7320	1167.4322	0.7939	2.6743
	2	99304	2.2791	10.4765	0.1024	0.0201
	3	240	150.6833	3327.0583	0.8129	96.7667
	4	1451	155.5183	3864.8973	0.7970	2.6954
3	0	226608	2.3435	49.2915	0.9718	0.0816
	1	1423	157.1377	3906.8096	0.7969	2.6676
	2	6468	56.6289	1186.9665	0.7934	2.7412
	3	99328	2.2769	10.5061	0.1032	0.0213
	4	238	145.7353	3255.4706	0.8078	97.5126

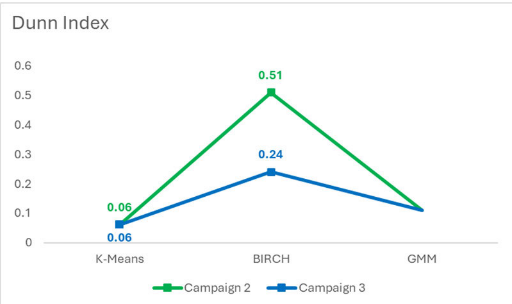

FIGURE 8. Score comparison for Dunn Index.

achieving the highest score in both campaigns. Considering the comprehensive evaluation and the significance of this gap between the Elbow Method and CH Index, K-Means was chosen as the preferred clustering method for both campaigns.

As the result, the overall performance suggested that K-Means clustering was the top-performing algorithm, closely followed by BIRCH. Conversely, GMM exhibited less favourable results in comparison to both clustering methods. Consequently, the user groups segmented by K-Means clustering were further analyzed to interpret the segmented user journeys through the AIDA Model in the following section.

V. DISCUSSION

The AIDA Model was used to understand the user journey of each segment and to identify how the users go through each campaign. Table 3presents the mean values of each AIDA stage within clusters for both campaigns. The decision to use mean values over range and median was because the mean provides a more comprehensive representation of the entire cluster by incorporating all data points; unlike the median, which focuses on the central value, and the range, which considers only minimum and maximum values.

However, mean values are hard to understand and may be confusing to interpret. Thus, a dynamic characteristic quartiles table is created in Table 4 to provide a clearer picture of the user's behavior interpretation. This dynamic characteristic quartiles table divides the AIDA values into four quartiles for each campaign respectively. Five characteristics are defined according to different quartiles, where the upper limit of each quartile is determined by 25%, 50%, and 75%. The initial characteristics for each stage are set at zero, while the subsequent segments are represented as low, moderate, high, and highest respectively.

By utilizing the dynamic quartiles table, the actual range may vary between campaigns due to differences in the data distribution within the AIDA stages across different campaigns. With the defined quartiles ranges, the actual ranges for Campaign 2 and Campaign 3 are also shown in Table 4.

By referring to the mean value in Table 3 and the dynamic quartiles table in Table 4, the characteristics of each user segment in both campaigns are interpreted. To highlight the values of the least favourable cluster and the most favourable cluster, the values are highlighted in red and blue respectively. For Campaign 2, Cluster 0 was the largest proportion, where this user groups shows a low *attention* (2.3230), low *interest* (48.4032), highest *desire* (0.9722), and low *action* (0.0828). Cluster 1 then demonstrated low *attention* (55.7320), moderate *interest* (1167.4233), highest *desire* (0.7939), and low *action* (2.6743). Next, Cluster 2 portrayed the least favourable performance, with the users showing low level across all stages, with the values of **(2.2791)**, **(10.4765)**, **(0.1024)**, and **(0.0201)** respectively. Finally, the most favourable performance was shown in Cluster 3, users in this cluster demonstrated a moderate *attention* **(150.6833)**, highest *interest* **(3327.0583)**, highest *desire* **(0.8129)**, and high *action* **(96.7667)**.

For Campaign 3, Cluster 0 obtained the largest proportion. Users in this cluster portrayed low *attention* (2.3435), low *interest* (49.2915), highest *desire* (0.9718), and low *action* (0.0816). Then, users in Clusters 1 showed a moderate *attention* (157.1377), highest *interest* (3906.8096), highest *desire* (0.7969), and low *action* (2.6676). Cluster 2 demonstrated a user group with low *attention* (56.6289), moderate *interest* (1186.9665), highest *desire* (0.7934), and low *action* (2.7412). The least favourable performance was showed by Cluster 3, where the users had low *attention* **(2.2769)**, low *interest* **(10.5061)**, low *desire* **(0.1032)**, and low *action* **(0.0213)**. In contrast, Cluster 4 exhibits the most favourable

TABLE 4. Dynamic characteristics quartiles table with the actual range of both campaigns.

Stages	Quartile Range	Actual Range		Characteristics
		Campaign 2	Campaign 3	
Attention (A)	A = 0	A = 0	A = 0	Users with no attention
	0% < A ≤ 25%	0 < A ≤ 80	0 < A ≤ 81	Users with low attention
	25% < A ≤ 50%	80 < A ≤ 160	81 < A ≤ 161	Users with moderate attention
	50% < A ≤ 75%	160 < A ≤ 247	161 < A ≤ 243	Users with high attention
	A > 75%	A > 247	A > 243	Users with the highest attention
Interest (I)	I = 0	I = 0	I = 0	Users with no interest in the campaign
	0% < I ≤ 25%	0 < I ≤ 881.0	0 < I ≤ 885.0	Users with low interest in the campaign
	25% < I ≤ 50%	881.0 < I ≤ 1808.0	885.0 < I ≤ 1806.0	Users with moderate interest in the campaign
	50% < I ≤ 75%	1808.0 < 1 ≤ 3106.0	1806.0 < 1 ≤ 3111.0	Users with high interest in the campaign
	I > 75%	1 > 3106.0	1 > 3111.0	Users with the highest interest in the campaign
Desire (D)	D = 0	D = 0	D = 0	Users with no desire in the campaign
	0% < D ≤ 25%	0 < D ≤ 0.25	0 < D ≤ 0.25	Users with low desire in the campaign
	25% < D ≤ 50%	0.25 < D ≤ 0.50	0.25 < D ≤ 0.50	Users with moderate desire in the campaign
	50% < D ≤ 75%	0.50 < D ≤ 0.75	0.50 < D ≤ 0.75	Users with high desire in the campaign
	D > 75%	D > 0.75	D > 0.75	Users with the highest desire in the campaign
Action (A)	A = 0	A = 0	A = 0	Users with no action taken in the campaign
	0% < A ≤ 25%	0 < A ≤ 38	0 < A ≤ 37	Users with low action taken in the campaign
	25% < A ≤ 50%	38 < A ≤ 76	37 < A ≤ 75	Users with moderate action taken in the campaign
	50% < A ≤ 75%	76 < A ≤ 123	75 < A ≤ 123	Users with high action taken in the campaign
	A > 75%	A > 123	A > 123	Users with the highest action taken in the campaign

performance with moderate *attention* **(145.7353)**, highest *desire* **(3255.4706)**, highest *interest* **(0.8078)**, and high *action* **(97.5126)**. To analyze the user journey of each cluster in both campaigns, Table 5 summarizes the quartiles range and the characteristics of each cluster associated with their actual range.

Based on the user behavioral characteristics in each cluster in Table 5, each of the 5 clusters exhibited corresponding characteristics in both campaigns. For example, Cluster 2 in Campaign 2 and Cluster 3 in Campaign 3 were characterized by low *Attention*, low *Interest*, low *Desire*, and low *Action*, sharing the same characteristics. Therefore, the user journey for the two clusters that share the same characteristics are visualized in Fig.5. The y-axis represents levels ranging from 0 to 4, indicating varying degrees of the AIDA stage as shown in Table 5. At 0, there is no observable engagement, while 1 signifies low level, 2 indicates moderate level, 3 denotes high level, and 4 represents the highest degree of *Attention*, *Interest*, *Desire*, or *Action*.

Fig. 9 shows the user journey of user groups with weakly engaged behavior, which are Cluster 2 in Campaign 2 and Cluster 3 in Campaign 3. These clusters represented the least favourable user performance, accounting for 29.80% and 29.73% of the total users in both campaigns, consistently exhibited low level across all stages. Users within these clusters displayed minimal attention to the campaign, leading to the conclusion of a lack of interest. The content seems to be failing to resonate with their needs or preferences. This lack of *Attention* and *Interest* extended to low levels of *Desire*. Consequently, the lack of *Attention*, *Interest*, and *Desire* resulted in limited user *Action*. The user journey of individuals in these clusters painted a picture of disengagement with both campaigns, suggesting that the content failed to capture their interest. These users were not interested in the advertised campaigns.

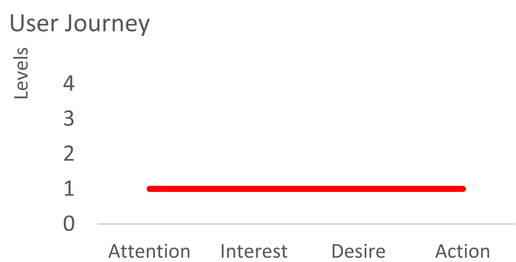

FIGURE 9. User journey of cluster 2 (Campaign 2) & cluster 3 (Campaign 3).

On the other hand, the user journey of user groups with strongly engaged behavior is illustrated in Fig. 10. These groups of users were identified from Cluster 3 in Campaign 2 and Cluster 4 in Campaign 3. Both clusters represented the most favourable user performance. Users in these clusters were highly engaged. These small groups of users accounted for only 0.07% of the total users in both campaigns respectively. These users showed a moderate level of *Attention*, highest *Interest*, highest *Desire*, and finally high *Action* as displayed by the journey. With their moderate level of frequency exposed to the campaign, these users exhibited the highest level of *Interest* and *Desire*. This implied a strong resonance with the campaign, indicating that the advertised content was particularly appealing and aligning well with their preferences or needs. Importantly, users in these clusters showcased a high propensity to take action. Their journey involved actively seeking out and engaging with the campaign, driven by genuine *Interest* and *Desire*, which ultimately results in meaningful *Action*. In essence,

TABLE 5. **The characteristics of each cluster in both campaigns.**

Campaign	Cluster	Total Users	Characteristics
2	0	225701 (67.72%)	Users with low *Attention*, low *Interest*, highest *Desire*, and low *Action*
	1	6574 (1.97%)	Users with low *Attention*, moderate *Interest*, highest *Desire*, and low *Action*
	2	**99304 (29.80%)**	**Users with low *Attention*, low *Interest*, low *Desire*, and low *Action***
	3	**240 (0.07%)**	**Users with moderate *Attention*, highest *Interest*, highest *Desire*, and high *Action***
	4	1451 (0.44%)	Users with moderate *Attention*, highest *Interest*, highest *Desire*, and low *Action*
3	0	226608 (68.83%)	Users with low *Attention*, low *Interest*, highest *Desire*, and low *Action*
	1	1423 (0.43%)	Users with moderate *Attention*, highest *Interest*, highest *Desire*, and low *Action*
	2	6468 (1.94%)	Users with low *Attention*, moderate *Interest*, highest *Desire*, and low *Action*
	3	**99328 (29.73%)**	**Users with low *Attention*, low *Interest*, low *Desire*, and low *Action***
	4	**238 (0.07%)**	**Users with moderate *Attention*, highest *Interest*, highest *Desire*, and high *Action***

* Blue signifies strongly engaged behavior, while red indicates weakly engaged behavior.

these users not only showed interest but were proactive in the process, showing a higher likelihood of converting their enthusiasm into tangible actions.

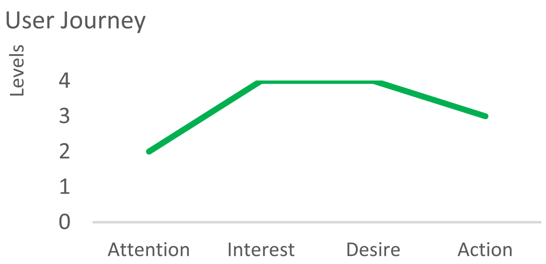

FIGURE 10. **User journey of cluster 3 (Campaign 2) & cluster 4 (Campaign 3).**

Subsequently, Fig. 11, 12, and 13 illustrate the user journeys for the remaining clusters, with each depicting distinct paths. Notably, the variations among these clusters primarily lied in the *Attention* and *Interest* stages. It is noteworthy that these user groups exhibited similar behaviors, where they all showed the highest level of *Desire* but a low level of *Action*, as observed in Cluster 0, Cluster 1, Cluster 4 in Campaign 2 and Cluster 0, Cluster 2, Cluster 1 in Campaign 3 respectively.

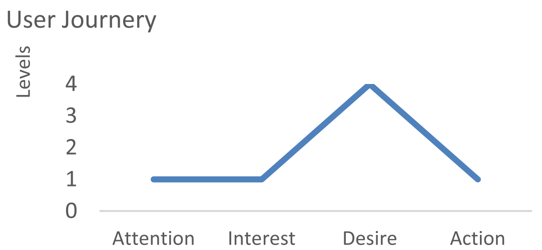

FIGURE 11. **User journey of cluster 0 in both campaigns.**

These behaviors are aligned with findings from [36], where the study used questionnaires to assess the effectiveness of e-tourism websites as a medium for tourism promotion by using the AIDA Model. According to the findings of the study, 76% of participants indicated a high level of *Desire*.

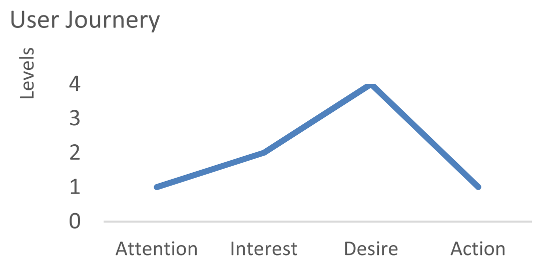

FIGURE 12. **User journey of cluster 1 (Campaign 2) & cluster 2 (Campaign 3).**

However, none of them ultimately took significant *Action*, while 72% of respondents displayed a low level of *Action*. This alignment in findings underscores a consistent trend across the studies, emphasizing the gap between desire and actualized actions.

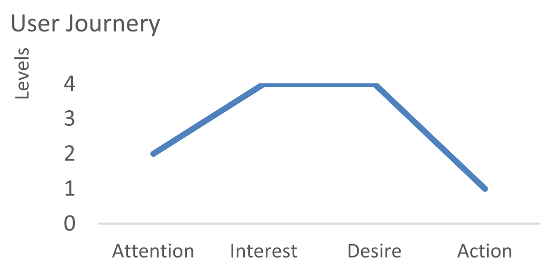

FIGURE 13. **User journey of cluster 4 (Campaign 2) & cluster 1 (Campaign 3).**

The wider implications of these findings for businesses or marketers are significant. It offers a better insight into how different user segments interact with campaigns and how they can refine their marketing strategies to better meet the needs and behaviors of each segment. For example, the user journey of the weakly engaged user groups in Fig. 9 shows that users have low engagement across all the AIDA stages. In this case, marketers can increase user engagement by redesigning the advertising campaign to be more engaging and visually appealing, especially at the entry point to catch the attention from the users. Another example in Fig. 13

demonstrates user groups with moderate *Attention*, highest *Interest* and *Desire* but their *Action is* low. Thus, marketers can refine their marketing strategies to enhance conversion rates. They may highlight exclusive offers or additional content in the campaign to motivate users to take *Action*. With these user journeys, marketers can gain deeper insights into each user group and leverage their characteristics to design and implement more effective marketing strategies, thereby boosting the performance of their advertising campaigns.

VI. CONCLUSION

This paper focused on performing customer segmentation using a user journey analysis approach with the integration of the AIDA Model and clustering algorithms. Various clustering methods were explored, including K-Means, BIRCH, and GMM, with K-Means identified as the most effective algorithm in the tested dataset, yielding five optimal clusters.

The findings of this paper underscore the usefulness of using the AIDA-based segmentation approach to comprehend the diverse user journeys within the Wi-Fi advertising system. Through clustering and subsequent analysis, the AIDA metrics provide an understanding of user progression through the AIDA stages along their user journey. The metric effectively links each stage of the AIDA Model, with the *attention* quantified by the frequency of user views for a campaign; *interest* captured through the total extra view duration; *desire* is then measured as the ratio of instances where users watched beyond the mandatory view time to their attention value, showcasing the transition from interest to desire. Finally, *action* is represented by the number of user visits, reflecting the final action of user progression with the advertising content. These insights not only validate the AIDA Model as an effective way for assessing user journey, they also suggest the application of the AIDA Model across various Wi-Fi advertising attributes. By understanding the specific stages where customers might drop off or show heightened interest, it enables businesses to tailor their marketing strategies to meet the specific preferences and needs of the identified user segments. This may navigate users through AIDA stages, driving desired actions effectively.

Furthermore, the findings of this study also highlight practical applications for real-world Wi-Fi advertising systems. By utilizing the AIDA-based customer segmentation along with user journey analysis, businesses can achieve a more nuanced understanding of user behavior within each user segment. This insight allows for the implementation of tailored marketing strategies based on different user journeys through the AIDA stages. Consequently, businesses can create highly targeted and personalized advertising campaigns, which can improve user experience and increase marketing effectiveness.

VII. LIMITATIONS AND FUTURE WORKS

The main limitation of this research lies in the specificity of the developed AIDA metric to the Wi-Fi advertising dataset used. The AIDA Model outlines a multi-stage formula of user engagement, quantifying the levels that each user demonstrates towards each campaign. However, the metric for each stage was calculated based on the details present in the Wi-Fi advertising system. Without such specific data, the model is less generalizable and limiting its applicability to other datasets that do not share these exact features. Therefore, future research could generalize the AIDA Model by utilizing more widely available features in diverse advertising contexts, enhancing its adaptability across different contexts. Besides, machine learning techniques could be employed to identify relevant features for each AIDA stage, allowing the model to adjust the calculations based on the available data. This forward-looking approach aims to overcome current limitations and contributes to a more flexible AIDA framework applicable to various advertising datasets, fostering adaptability and broader applicability.

REFERENCES

[1] M. Carnein and H. Trautmann, "Customer segmentation based on transactional data using stream clustering," *Proc. 23rd Pacific–Asia Conf.*, 2119, pp. 280–292, doi: 10.1007/978-3-030-16148-4_22.

[2] Z. Y. Ünlü, "Analysis of online and in-store clothes shopping experiences taking a user journey approach," M.S. thesis, Dept. Ind. Des., Middle East Tech. Univ. Ankara, Turkey, 2022. [Online]. Available: https://hdl.handle.net/11511/96205

[3] S. Hassan, S. Z. A. Nadzim, and N. Shiratuddin, "Strategic use of social media for small business based on the AIDA model," *Proc.-Social Behav. Sci.*, vol. 172, pp. 262–269, Jan. 2015, doi: 10.1016/j.sbspro.2015.01.363.

[4] Y. I. Lee and P. R. J. Trim, "Enhancing marketing provision through increased online safety that imbues consumer confidence: Coupling AI and ML with the AIDA Model," *Big Data Cognit. Comput.*, vol. 6, no. 3, p. 78, 2022, doi: 10.3390/bdcc6030078.

[5] T. T. A. Nguyen, C. P. Chang, and S. M. Yuan, "A Wi-Fi union mechanism for internet advertising reciprocal platform in microenterprises," *Sensors*, vol. 17, no. 7, p. 1617, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/7/1617

[6] AIMS Data Centre. (Oct. 2019). *Digi WiFi—Uses TM's Free Wifi Network, Cost Up to RM3*. [Online]. Available: https://www.malaysianwireless.com/2019/10/digi-wifi-internet-service/

[7] WiMedia. (2024). *WiFi Advertising Platform*. [Online]. Available: https://wimedia.am/about

[8] A. Alghamdi, "A hybrid method for customer segmentation in Saudi Arabia restaurants using clustering, neural networks and optimization learning techniques," *Arabian J. Sci. Eng.*, vol. 48, no. 2, pp. 2021–2039, Feb. 2023, doi: 10.1007/s13369-022-07091-y.

[9] C. Wang, "Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach," *Inf. Process. Manage.*, vol. 59, no. 6, Nov. 2022, Art. no. 103085, doi: 10.1016/j.ipm.2022.103085.

[10] V. A. Nasir, A. C. Keserel, O. E. Surgit, and M. Nalbant, "Segmenting consumers based on social media advertising perceptions: How does purchase intention differ across segments?" *Telematics Informat.*, vol. 64, Nov. 2021, Art. no. 101687, doi: 10.1016/j.tele.2021.101687.

[11] S.-T. Lim, L.-Y. Ong, and M.-C. Leow, "New RFI model for behavioral audience segmentation in Wi-Fi advertising system," *Future Internet*, vol. 15, no. 11, p. 351, Oct. 2023, doi: 10.3390/fi15110351.

[12] A. Pantouvakis and A. Gerou, "The theoretical and practical evolution of customer journey and its significance in services sustainability," *Sustainability*, vol. 14, no. 15, p. 9610, 2022, doi: 10.3390/su14159610.

[13] A. Terragni and M. Hassani, "Analyzing customer journey with process mining: From discovery to recommendations," in *Proc. IEEE 6th Int. Conf. Future Internet Things Cloud (FiCloud)*, Aug. 2018, pp. 224–229, doi: 10.1109/FICLOUD.2018.00040.

[14] Y. Tueanrat, S. Papagiannidis, and E. Alamanos, "A conceptual framework of the antecedents of customer journey satisfaction in omnichannel retailing," *J. Retailing Consum. Services*, vol. 61, Jul. 2021, Art. no. 102550, doi: 10.1016/j.jretconser.2021.102550.

[15] D. Herhausen, K. Kleinlercher, P. C. Verhoef, O. Emrich, and T. Rudolph, "Loyalty formation for different customer journey segments," *J. Retailing*, vol. 95, no. 3, pp. 9–29, Sep. 2019, doi: 10.1016/j.jretai.2019.05.001.

[16] S. Shen, M. Sotiriadis, and Y. Zhang, "The influence of smart technologies on customer journey in tourist attractions within the smart tourism management framework," *Sustainability*, vol. 12, no. 10, p. 4157, May 2020, doi: 10.3390/su12104157.

[17] H. Song, W. J. Ruan, and Y. J. J. Jeon, "An integrated approach to the purchase decision making process of food-delivery apps: Focusing on the TAM and AIDA models," *Int. J. Hospitality Manage.*, vol. 95, May 2021, Art. no. 102943, doi: 10.1016/j.ijhm.2021.102943.

[18] M. Prathapan, D. S. Sahadevan, and K. A. Zakkariya, "Effectiveness of digital marketing: Tourism websites comparative analytics based on AIDA model," *Int. J. Innov. Res. Stud.*, vol. 8, no. 4, pp. 262–273, 2018. [Online]. Available: https://www.pramanaresearch.org/gallery/jun-p191.pdf

[19] R. Mumtaz, "Awareness and perspectives social media as new strategic marketing approach in minor industries; Notion grounded on AIDA model," *J. Content, Community Commun.*, pp. 12–19, Dec. 2019, doi: 10.31620/JCCC.12.19/22.

[20] M. C. Ikhsan, V. S. Astuti, A. A. Wijaya, F. Finuliyah, and A. M. Qulub, "Does paid promote influences user frequency to purchase? An analysis using AIDA dimension," *Airlangga J. Innov. Manage.*, vol. 3, no. 2, pp. 114–122, Nov. 2022, doi: 10.20473/ajim.v3i2.20460.

[21] X. L. Polk, "Marketing: The key to successful teaching and learning," *J. Marketing Dev. Competit.*, vol. 12, no. 2, pp. 49–57, 2018, doi: 10.33423/jmdc.v12i2.1257.

[22] S. Dridi, "Unsupervised learning—A systematic literature review," Center Open Sci., OSF Preprints, Charlottesville, VA, USA, Tech. Rep., pp. 1–20, 2021, doi: 10.31219/osf.io/mpkht.

[23] A. Nagar and J. Sharma, "A comparative evaluation of supervised learning and unsupervised learning," *Resmilitaris*, vol. 12, no. 5, pp. 1498–1507, 2022. [Online]. Available: https://resmilitaris.net/menu-script/index.php/resmilitaris/article/view/2285

[24] I. Ibrahim and A. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 10–19, Mar. 2021, doi: 10.38094/jastt20179.

[25] M. C. Nwadiugwu, "Gene-based clustering algorithms: Comparison between denclue, fuzzy-C, and BIRCH," *Bioinf. Biol. Insights*, vol. 14, Jan. 2020, Art. no. 117793222090985, doi: 10.1177/1177932220909851.

[26] A. Lang and E. Schubert, "BETULA: Numerically stable CF-trees for BIRCH clustering," in *Proc. Int. Conf. Similarity Search Appl.*, Cham, Switzerland, 2020, pp. 281–296, doi: 10.1007/978-3-030-60936-8_22.

[27] E. Patel and D. S. Kushwaha, "Clustering cloud workloads: K-means vs Gaussian mixture model," *Proc. Comput. Sci.*, vol. 171, pp. 158–167, Jan. 2020, doi: 10.1016/j.procs.2020.04.017.

[28] A. Abdulhafedh, "Incorporating K-means, hierarchical clustering and PCA in customer segmentation," *J. City Dev.*, vol. 3, no. 1, pp. 12–30, 2021. [Online]. Available: https://pubs.sciepub.com/jcd/3/1/3/index.html

[29] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.

[30] X. Wang and Y. Xu, "An improved index for clustering validation based on silhouette index and calinski-harabasz index," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 569, no. 5, Jul. 2019, Art. no. 052024, doi: 10.1088/1757-899x/569/5/052024.

[31] C.-E. Ben Ncir, A. Hamza, and W. Bouaguel, "Parallel and scalable Dunn index for the validation of big data clusters," *Parallel Comput.*, vol. 102, May 2021, Art. no. 102751, doi: 10.1016/j.parco.2021.102751.

[32] H. Y. Kim, Y. Lee, E. Cho, and Y. J. Jung, "Digital atmosphere of fashion retail stores," *Fashion Textiles*, vol. 7, no. 1, pp. 1–17, Sep. 2020, doi: 10.1186/s40691-020-00217-6.

[33] A. Wind, E. Constantinides, and S. de Vries, "Marketing a transparent artificial intelligence (AI): A preliminary study on message design," in *Proc. 18th Int. Mark. Trends Conf.*, 2019, pp. 1–7. [Online]. Available: https://archives.marketing-trendscongress.com/2019/pages/PDF/91.pdf

[34] Z.-Y. Lim, L.-Y. Ong, and M.-C. Leow, "A review on clustering techniques: Creating better user experience for online roadshow," *Future Internet*, vol. 13, no. 9, p. 233, Sep. 2021, doi: 10.3390/fi13090233.

[35] Y. Zhang, M. Li, S. Wang, S. Dai, L. Luo, E. Zhu, H. Xu, X. Zhu, C. Yao, and H. Zhou, "Gaussian mixture model clustering with incomplete data," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 1, pp. 1–14, Jan. 2021, doi: 10.1145/3408318.

[36] N. Amanda, P. I. Santosa, and W. W. Winarno, "Effectiveness of e-tourism website as a tourism promotion media using AIDA model: A study in context of visitingjogja.com," in *Proc. Int. Conf. Creative Econ., Tourism Inf. Manage.*, 2019, pp. 133–138, doi: 10.5220/0009866001330138.

SHI-YEN WONG received the bachelor's degree (Hons.) in information technology (business intelligence and analytics) from Multimedia University, Malaysia. She is a recent graduate and eager to apply her knowledge and skills to real-world challenges and contribute to the advancement of the field. Her research interests include behavior and data analytics.

LEE-YENG ONG (Senior Member, IEEE) received the Ph.D. degree in computer vision from Multimedia University, Malaysia. She is currently an Assistant Professor with the Faculty of Information Science and Technology, Multimedia University. Her research interests include object tracking, computer vision, data science, and big data analytics.

MENG-CHEW LEOW (Senior Member, IEEE) received the Doctor of Philosophy degree from Multimedia University. His research interest includes game-based learning, specifically in role-playing game-based learning. He is also interested in system science, practical spirituality, and philosophy.