WILEY | Hindawi

Research Article
Predicting Online Consumer Transaction from Big Data: Influential Factors and Strategic Planning

Chiang-Yu Cheng [ID],[1] **Ming-Ying Lu** [ID],[1] **and Han-Ping Tsen** [ID][2]

[1]*School of big Data Management, Soochow University, Taiwan*
[2]*Institute of Information Management, National Yang Ming Chiao Tung University, Hsinchu, Taiwan*

Correspondence should be addressed to Han-Ping Tsen; hanping311111@gmail.com

Received 30 September 2020; Accepted 24 August 2021; Published 25 September 2021

Academic Editor: Feng-Jang Hwang

Online transaction has recently benefited from coronavirus; however, the sales of e-commerce in some areas are substantially on the decline. The current study proposes a theoretically constructed and empirically viable way for predicting the relevant factors that may detract or foster e-commerce success. We apply web analytics (one of the big data techniques) to simultaneously, generalizably, and objectively measure the influential factors of e-commerce success. The findings indicate that (1) pageviews is an important key for consumers to make transactions. (2) Bounce rate of the website should not be a member factor of e-commerce success. (3) Adhesion strategy and repeatability strategy can be used to induce consumer online transaction. Several theoretical contributions and practical implications are also provided.

1. Introduction

The new type of coronavirus has invaded the world and has reduced people's willingness to go outside; however, this harmful impact fits quite well with the characteristic of e-commerce, because people do not need to actually patronize the storefronts but can still make their purchases at home. Even though online transaction may be benefited from the outbreak of such virus pandemic, the sales of e-commerce in some areas are substantially on the decline. According to the survey reported by eMarketer [1] that global retail e-commerce sales decelerated to a 16.5% growth rate in 2020 (down from 20.2% in 2019), this tendency is particularly obvious in North America and Asia-Pacific. Due to the fact that e-commerce sales in the US and China are accounted for the majority of overall global sales metrics, it is necessary to figure out the factors that can be used to boost online transactions for both countries. Previous studies have adopted some ways to explore the influential factors of e-commerce success. For example, Wang et al. [2] utilized structural equation modeling to validate the proposed e-commerce success model. Kshetri [3] applied case study to clarify the barriers of e-commerce in the developing country. Liu et al. [4] conducted an expert interview to measure the performance of e-commerce. Although these methods are applicable to understand the influential factors of e-commerce success, there exist several disadvantages that should be further considered. *First*, in most cases, structural equation modeling is subject to validate the relationship between independent and dependent variables [5], but it is rarely used to verify the relationship between independent variables. In fact, the success of e-commerce is definitely determined by multiple independent variables. Thus, structural equation modeling in this scenario may not be suitable. *Second*, case study is one of qualitative research methods which is useful for the theory building, especially when existing theories are inapplicable [6]. Many researchers have paid attention to the relevant theories in the field of e-commerce, and many related theories have also been published, even though some e-commerce proprietors are still at a financial loss. In addition, the results of case study method are only applicable to specific subjects but not for the overall e-commerce industry in general. Therefore, when the research goal is to propose a generalizable method for the whole

industry, the applicability of case study is in doubt. *Finally*, expert interview can help researchers obtain valuable references especially for those within the company. For instance, an interviewee points out that the main reason for the decline in e-commerce profit is the market plunder coming from the competitors. However, in some cases, the opinions of experts may be too subjective; meaning, this revenue decreasement may be caused by internal factors rather than external ones, such as shopping cart abandonment on the website.

To improve the measurement of e-commerce success, the current study proposes a theoretically constructed and empirically viable way for predicting the relevant factors that may detract or foster e-commerce success. We apply web analytics to simultaneously, generalizably, and objectively measure the influential factors of e-commerce success. It should be noted that e-commerce success in our study is reasonably defined as an actual purchasing behavior performed by the online consumers, because consumer consumption should be one of necessary conditions for the success of e-commerce [7]. After our data analytics, an insightful strategic planning is provided to guide the improvement of e-commerce revenue.

2. Literature Review

2.1. E-Commerce Behavioral Process. Previous studies have divided consumer behavioral process into several steps. From the perspective of advertising, it is suggested that consumers' response to the receiving advertising conforms to the three steps, including thinking, feeling, and doing [8]. Thinking refers to consumers' cognitive aspects with regard to the receiving advertisements, such as beliefs, thoughts, or knowledge, while feeling pertains to their affectively experiences or emotions toward the brand or the content of the advertisement. As for doing, it describes consumers' willingness to proactively perform a certain behavior (e.g., intention to purchase). In the context of e-commerce, Kim et al. [9] proposed a three phase model that depicts consumer online purchase process, including prepurchase, purchase, and postpurchase. The scenario of these three phases can be that consumers often receive advertising over the Internet, and then they will determine whether the received message is trustworthy. If that message is worthy to be trusted, they are very likely to click the advertising URL that directs them into the shopping website (prepurchase phase). When the consumers are brought to the shopping website, they get involved in the next phase. In this purchase phase, consumers spend times on the website and try to scrutinize the products page by page. After surfing the product pages, some of consumers may immediately decide whether to purchase the product, and the others may take several days to consider their purchase decision. Regardless of immediate or hesitant consumers, there are only two results at this phase that is to buy or not to buy. Once they have actually purchased the product, the final phase is said to be emerged. In the postpurchase stage, consumers have converted their role from visitor to customer, and product quality, after-sales service, or product warranty should be emphasized by them. If customers are satisfied with these concerns, they are very likely to conduct their consumption again and share their positive shopping experiences with others (e.g., friends, family members, and colleagues).

Apparently, consumer online behavior can be concretely observed. We therefore theorize the journey of online transaction to be the three phases as well, including *acquisition*, *behavior*, and *conversion*. From the perspective of e-commerce proprietors, the proposed three phases here are relative to the phases mentioned earlier (see Table 1). For example, consumers performing certain online behaviors in the prepurchase phase can be attributed from the effectiveness of consumer acquisition strategy. We believe that this portraying can serve as a navigator to the guidance of e-commerce success.

2.2. Definitions of Big Data. Big data has been inconsistently defined by the previous studies, but most researchers believe that it consists of at least three characteristics, including volume, velocity, and variety [10]. Volume refers to the amount of data, while velocity denotes whether the data is collected in a real time manner and variety is the diversity of data. Although this 3Vs definition is well accepted, some researchers raise different views. For example, Mayer-Schonberger and Cukier [11] argued that big data denote exhaustively data collection rather than data sampling. Boyd and Crawford [12] defined big data as relationality data that contains common fields to enable the conjoining of a brand new dataset. Uddin and Gupta [13], on the other hand, considered that big data can be meaningfully shifting in relation to the context in which they are collected. Gandomi and Haider [14], however, defined big data as value-oriented data interpretation that many meaningful data insights can be extracted. In the current study, we postulate that the record of consumer online behavior is a kind of big data, because (1) as long as consumers visit the website, they generate footprints all the time until they leave the site. Thus, if the continuously generated data is collected, the amount of data should be very large (i.e., volume). (2) Consumers' footprints can be comprehensively collected as long as they happen on the website; meaning, the data is entirely captured rather than being sampled (i.e., exhaustivity). (3) Data value can be obtained through the interpretation of footprint data (i.e., value-oriented data). For example, consumers who stay on the website for more than five minutes showed a 10% purchase rate. Even though e-commerce proprietors possess consumer footprint data, most of them do not realize how to convert valuable information on hand into data insight [15], and we therefore introduce a feasible way in which consumer online behavior can be largely, exhaustively, and valuably collected, especially for those related to online transaction.

2.3. Big Data Analytics. Web analytics is one of big data techniques which can be used to observe the visitor's online journey and elicit the visitor's browsing and purchasing patterns from the website traffic [16]. Generally speaking, web analytics has two types: offsite web analytics and onsite web analytics. The former can be used regardless of the ownership of a website, while the latter can only be used if the

TABLE 1: The correspondence of the two process models.

Behavioral process model	Phase 1	Phase 2	Phase 3
Online transaction journey (managerial guidance)	Acquisition	Behavior	Conversion
Consumer shopping process (behavior observation)	Prepurchase	Purchase	Postpurchase

analyst owns a website or has the permission to access it. Onsite web analytics measures a visitor's behavior performed on the specific website especially for the measures that are unable to be captured by offsite web analytics (e.g., the number of a website's visitors who conduct a specific action to go beyond a casual content view or simply website visit). As the goal of our study is to predict online transaction, onsite web analytics is more applicable than offsite web analytics, because (1) setting a self-maintained e-commerce website is pervasive for almost all proprietors. This means that proprietors themselves are able to install tracking code into their websites without the need of access permission. (2) E-commerce website itself can be seen as a means by which proprietors apply to the persuasion of potential consumers, and onsite web analytics to this end helps proprietors to estimate consumer interaction with the website. For example, it can reconnoiter what pages potential consumers have visited, where they have been referred from, how much time they have spent on a website, and even clicking behaviors in their journeys of the website. (3) Due to the fact that consumers may withdraw a shopping website from their consideration sets if the product information provided by that website is not insufficient, with the aid of on-site web analytics, proprietors can understand the actual online behavior of potential consumers and accordingly conduct an improvement that highlights consumers' needs of product information. For instance, onsite web analytics can calculate the bounce rate (i.e., the percentage of visitors who enter the website and leave that website immediately) for a shopping website. If a high bounce rate is detected, this means that the landing page of the website is not capable of guiding potential consumers to continue browsing the website in depth. Accordingly, the current study applies onsite web analytics to collect behavioral big data on the e-commerce website which is essential to the prediction of online transaction.

3. Research Design

The current study attempts to identify the features of online behavior that encourage consumers to conduct a transaction on the website and then to develop a prediction model for e-commerce success. Figure 1 illustrates the research process. First, a researched e-commerce website was selected and then variables that are related to online transaction were derived from web analytics. Traffic data of the researched website were used to create a decision tree that predicts consumer transaction on the website. Finally, the accuracy of the prediction model was evaluated and compared to the results of the other prediction algorithms, such as logistic regression, support vector machine, and random forest.
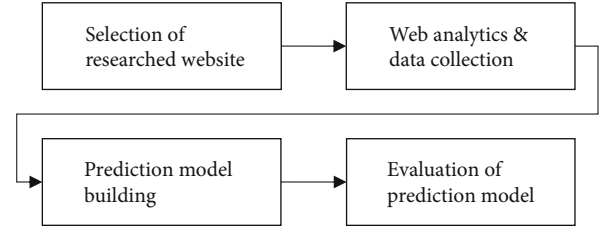

FIGURE 1: Research process.

3.1. Decision Tree. Decision tree is a supervised machine learning algorithm for the establishment of the discriminatory model [17]. It uses a tree-like model to extract valuable relationships from the information hidden in a data source and is frequently used for the purpose of classification or prediction [18]. There are two kinds of decision tree in general, including regression tree and classification tree. The former is developed when the goal factor is continuous, whereas latter is created when the goal factor is discrete. Regardless of which type of decision tree, they all go through a series of tests until the flow path reaches the terminal node. In the current study, we adopt classification tree rather than regression tree, because our goal is to predict whether or not consumers will conduct a transaction on the website which belongs to the concept of binary value.

3.2. Research Subject. For maintaining external validity, we collected the data from a real e-commerce website of Google (i.e., https://shop.googlemerchandisestore.com) which is a large-scale e-commerce website operated in the United States. Google e-commerce website sells many kinds of self-branded product, including backpack, apparel, electronics, and accessories. Collected data was elicited from the tool of web analytics developed by Google which is called Google Analytics (GA). Figure 2 outlines an overview of GA report. In default, this report includes data columns about users, new users, sessions, number of sessions per user, pageviews, pages/session, avg. session duration, and bounce rate.

3.3. Research Variables. In the context of online transaction, several variables are related to e-commerce success. Table 2 describes the variables that are used for the creation of decision tree. *Channel grouping, totals_bounces,* and *visitNumber* are the three factors relevant to the effectiveness of consumer acquisition (i.e., phase 1). Channel grouping (channelGrouping) is defined as the way of website traffic in which visitors are acquired, including organic search, direct, paid search, referral social, display, and affiliate. For example, if a visitor gets into the website by searching keywords on the search engine, this entrance traffic will be termed by GA as organic search. However, if a website visit is attributed by the search engine results that is the result of paid advertising, it can be termed as paid search. In addition, website
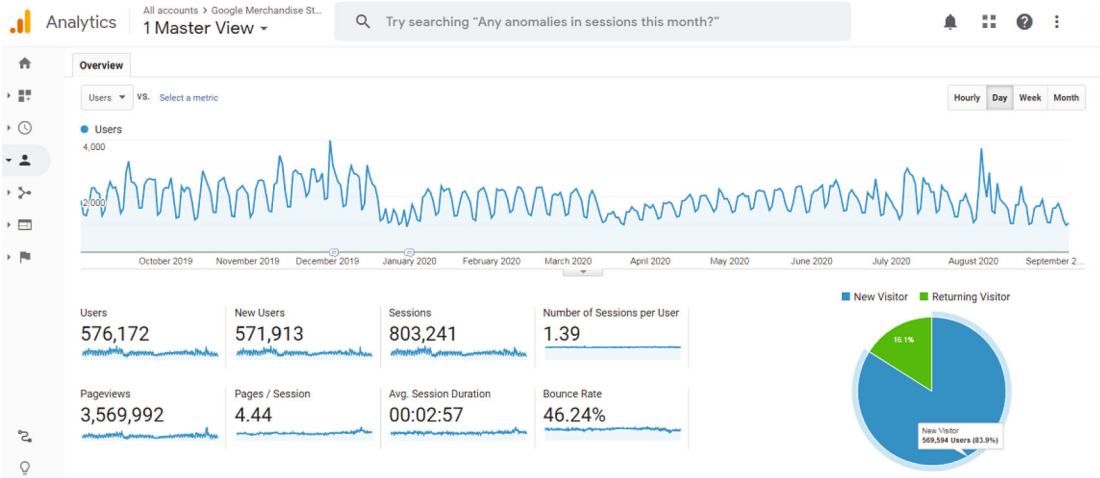
FIGURE 2: An overview of Google Analytics report.

proprietors may sometimes acquire wrong visitors by the wrong traffic channels. If this is true, the visitors who get into the website will very likely to leave that site without performing any further behaviors. Thus, the number of bounces (totals_bounces) can be used to monitor such a harmful behavior of the website. Furthermore, the number of times visitors have entered the website (visitNumber) plays as a necessary condition for the establishment of a complete shopping journey; as a result, it is impossible to make a purchase without the website visit. Due to the fact that website visitors attracted from different traffic channels exert differently on their website journeys, especially for those add-to-cart behaviors [19] and the appropriate traffic channel can effectively fix the problem of a landing page with high bounce rate [20], or even the number of website visits is highly related to website engagement of online consumers [21], the current study therefore includes these three measures into the decision tree and theorizes them as representative factors in the phase of consumer acquisition.

On the other hand, once visitors have actually entered the website, the number of pages they have viewed on the website (totals_pageviews) and the length of time the visitors have spent on the website (totals_timeOnSite) should be the other two important factors need to be addressed. Indeed, the more the webpages visitors have viewed, the more likely they will stay longer on the website, specifically for the visitors whose purpose is exploratory visit rather than goal-oriented visit [22]. Although it is difficult to measure the real success of e-commerce, technical measures here can serve as an alternative to indirectly validate the so-called success, because consumer website stickiness should be a consequence of high pageviews and long staying duration [23] which in turn affects consumers' intention to purchase online [24]. It is well known that purchase intention of a consumer does not necessarily correlated to her/his actual purchase behavior [25], for example, a consumer puts a product into the shopping cart but consequently regrets this temporary collection, and cart abandonment is said to have occurred. Thus, the number of transactions made by visitors on the website (totals_transactions) is particularly important

to e-commerce success. In general, totals_pageviews and totals_timeOnSite are theorized in phase 2, while totals_transactions in phase 3 is characterized as the destination of online transaction journey.

4. Data Analysis

4.1. Summary of Research Data. In this study, the rationality of traffic data is regarded as an important data cleaning criterion, because unreasonable traffic will result in bias which is detrimental to our prediction model. In the context of web analytics, the so-called unreasonable traffic can be a webpage with very short time on site. If a website visitor stays on the page for very short period of time (e.g., 3 seconds), the possibility of completing a purchase in such short period of time is unlikely. Accordingly, webpages with time of site less than 10 minutes will be removed from our data set. Anesbury et al. [26] also reported that most consumers spend more than 10 minutes to complete their online transactions. After data cleaning, a total of 1,692 website traffic data were obtained in the current study. Table 3 summarizes the outcome of descriptive statistics for all the research variables, including three continuous variables and two categorical variables. For the continuous variables, which show that the average number of visits to Google Merchandise Store is 3.07 (visitNumber), the average pageviews received by this website is 14.20 pages (totals_pageviews), and the average time visitors spend on the website is 642.46 seconds (totals_timeOnSite). For the categorical factors, which show that 25 percent of data enter the website and bounce the websites, 51 percent of data has transacted. Referring to bounce rate, 25% of consumers left the website without additional interaction (totals_bounces), while nearly 51% of consumers conducted their transactions on the website (totals_transactions). It should be noted that totals_bounces and totals_transactions are categorical factors; thus, the values of them are extremely distributed.

4.2. Decision Tree Development. A decision tree to predict e-commerce success is developed by taking consumers' online

TABLE 2: Research variables.

Variables and corresponding phase	Description
Acquisition (phase 1)	
channelGrouping	The origins of the website traffic that commonly includes direct, referral, organic search, paid search, display, affiliate, and social.
totals_bounces	The number the visitors get into the website but immediately leave the site without additional activities.
visitNumber	The number of times visitors have entered the website.
Behavior (phase 2)	
totals_pageviews	The number of pages viewed by visitors on the website.
totals_timeOnSite	The length of time visitors spend on the website.
Conversion (phase 3)	
totals_transactions	The number of transactions made by visitors on the website.

TABLE 3: Summary of research data (sample size = 1692).

Research variables	Minimum	Maximum	Mean	Standard deviation
visitNumber	1	376	3.07	13.281
totals_pageviews	1	165	14.20	15.320
totals_timeOnSite*	600	6865	642.46	746.500
totals_bounces	0	1	0.25	0.434
totals_transactions	0	1	0.51	0.500

*The metric of totals_timeOnSite is expressed in seconds by default.

transactions as the dependent variable (i.e., to buy or not to buy), and the four characteristics related to the phases 1 and 2 of online transaction journey are treated as the independent variables. Out of 1692 valid website traffic data, a total of 1184 were categorized as training data set (nearly 70% of all data), and the remaining of 508 was served as testing data set (nearly 30% of all data). Testing data set was used to validate the effectiveness of the proposed training model. SPSS Statistics with the Exhaustive CHAID (Chi Squared Automatic Interaction Detection) algorithm was applied to the tree construction. This algorithm is a modification of CHAID which evaluates all the possible splits for each independent variable. In addition, Exhaustive CHAID consists of three steps, including merging, splitting, and stopping. These three steps are repeatedly operationalized on each node from the top of tree until no additional nodes can be produced. In other words, if an independent variable has the strongest association with the root variable, it will be treated as the first branch in the tree with a leaf for each category that is significantly different from its parent variable. This process is continuous until no significant parent variables exist. It should be noted that the max_depth of the decision tree is not presupposed in our study, because we want to expand all the nodes in order to obtain pure leaves. This is the reason why we choose a modification of CHAID in SPSS so that all possible splits for each predictor can be measured. Yang and Shami [27] also adopted the same strategy to choose the best split node for the impure node. Accordingly, the hyperparameter of max_depth was determined to be three as it reveals the best accuracy for the prediction model and not the most complicated.

Figure 3 illustrates the outcome of decision tree development. The most important factor related to totals_transactions is totals_pageviews. Compared with consumers who are unwilling to perform purchase behavior and have website visit for less than 4 pages (nonpurchase: 99.6%, $n = 458$ vs. purchase: 0.4%, $n = 2$), consumers who visit the website for 4 to 11 pages are more likely to show purchase behavior (nonpurchase: 66.7%, $n = 93$ vs. purchase: 33.3%, $n = 47$). Moreover, when consumers visit the website for more than 11 pages, they exhibit higher number of transactions (nonpurchase: 5.4%, $n = 32$ vs. purchase: 94.6%, $n = 552$). This phenomenon is consistent with the finding of previous studies [28] that the higher the degree of website involvement, the more likely consumers are able to show their purchase intentions. Due to the fact that totals_pageviews plays a key role in purchase decision, it is necessary to clarify its facilitating factors. totals_timeOnSite and visitNumber function as main factors for the inducement of high totals_pageviews via respective traffic channels. For totals_timeOnSite (green squares in the figure), consumers who stayed on the website for over 1,461 seconds can perform higher transaction rate through organic and social channels in comparison with those who stayed on the website for less than 661 seconds.

This implies that consumers with higher stickiness are willing to spend more time on the website, because they get into the website voluntarily (i.e., channelGrouping: organic search) or they come to the website through the recommendation of their social friends (i.e., channelGrouping: social). Accordingly, organic search and social in channelGrouping are ingredients of customer acquisition, while totals_timeOnSite is prerequisite of customer retention, either the former or the later is essential to the success of customer value management [29]. As for visitNumber (blue squares in the figure), consumers who visit the website for over one time exhibit higher transaction rate through different channels (i.e., Direct, Referral, Paid Search, Display) than those who visit the website for merely one time. This finding confirms the importance of website visit that the more visits to the website, the more likely the purchase will
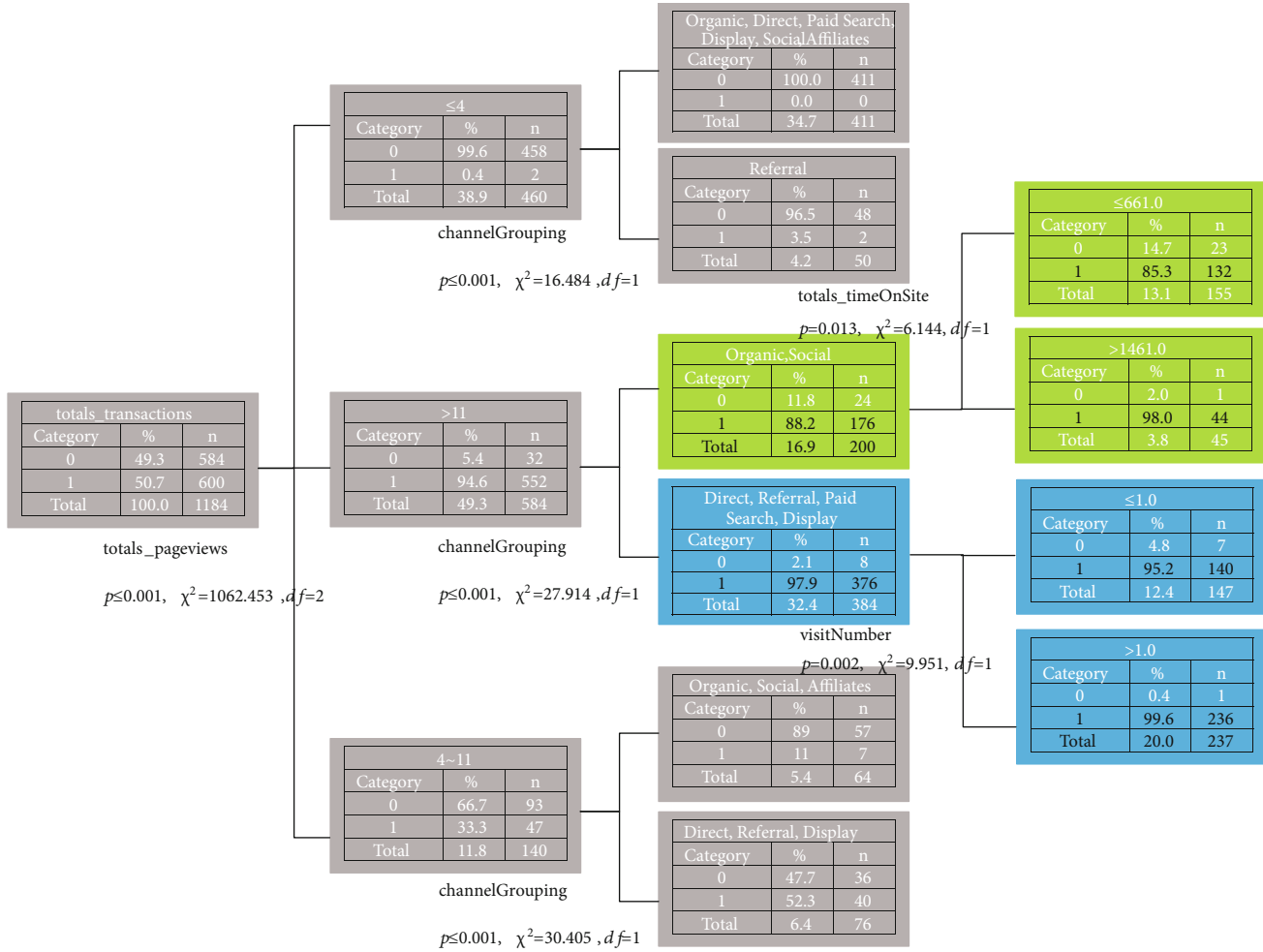
Figure 3: Decision tree.

be conducted. Despite a large number of website visits still end without any purchase, it can induce high pageviews when these traffics are triggered by the right channels which in turn derives purchase behavior. Generally speaking, if a visitor stays long enough on the website or if she/he frequently visits the website, a sufficient number of pageviews will be elicited, and the visitors should be inclined to have a purchase.

4.3. Model Evaluation. Table 4 outlines the evaluation results of the proposed mode. The prediction accuracy of the two data groups (i.e., non-purchase vs. purchase) is 93.9% and 99.2%; thus, the overall accuracy is 96.7%. Apparently, the prediction accuracy of the purchase group outperforms the nonpurchase group.

To further evaluate the effectiveness of the proposed model, we use different algorithms to develop other prediction models for the purpose of model comparison, including logistic regression, SVM, and random forest. Table 5 summarizes the results. Regardless of the overall prediction accuracy or the prediction accuracy of the purchase group, the performance of the proposed model outperforms the other three control models.

Table 4: Prediction accuracy of the proposed model (decision tree).

Observed	Predicted		
	Nonpurchase	Purchase	Percent correct
Nonpurchase	229	15	93.9%
Purchase	2	262	99.2%
Overall percentage	45.5%	54.5%	96.7%

Table 6 reveals the results of five-fold crossvalidation. The average accuracy and standard deviation of them are 94.4% and 0.009, respectively, confirming the stability of the model.

In summary, the decision tree algorithm in our study has better predictive ability than the other three algorithms, and it is therefore suitable for predicting the likelihood of consumer purchase and even to elicit under what circumstances consumers are willing to engage in online transaction [30].

5. Discussion

The current study found that pageviews is an important key for consumers to make transactions. In other words, as long as the number of pageviews reaches the threshold detected

TABLE 5: Prediction accuracy of the comparative models.

Observed	Predicted		
	Nonpurchase	Purchase	Percent correct
Logistic regression			
Nonpurchase	227	17	93.4%
Purchase	4	260	97.3%
Overall percentage	45.5%	54.5%	95.5%
SVM			
Nonpurchase	228	16	86.9%
Purchase	7	257	98.1%
Overall percentage	46.3%	53.7%	92.7%
Random forest			
Nonpurchase	227	17	93.0%
Purchase	4	260	98.5%
Overall percentage	45.5%	54.5%	95.9%

TABLE 6: The five-fold crossvalidation.

	1st fold	2nd fold	3rd fold	4th fold	5th fold
Accuracy	94.7%	94.4%	92.9%	94.7%	95.6%

Mean: 94.4% Std: 0.009

from the web analytics, consumers are likely to engage in purchase behavior. Based on our findings, there can be two strategies to effectively increase the number of pageviews on the website, and they are *adhesion strategy* and *repeatability strategy*. The focus of adhesion strategy is to enable consumers increasing their stay on the website, especially when their website entrances are through voluntary search engine queries (i.e., organic search) or recommendations from friends (i.e., social). In fact, people can put more effort into voluntary things [31] or things recommended by friends [32]; hence, overly forcing consumers or the wrong source of information may reduce their website stickiness. As for repeatability strategy, a repeated visit to a website would be devoted to seek for information that was not covered by the first time visit [33], because not all consumers are willing to complete a purchase within one visit to the website. As a result, consumers may need to visit the website multiple times before they actually engaging in the purchase. The association between website repeated visit and actual consumption behavior can also be validated by our findings that consumers who have entered the website for more than once through various channels (i.e., direct, referral, paid search, and display) are more able to demonstrate actual purchase behavior. It is worth noting that the two strategies here may have been mentioned in literature, but they are frequently mentioned with other e-commerce success factors. However, too many factors are mentioned simultaneously will cause the operators have no idea about the priority of these influencing factors. Thus, the above two strategies specifically inform large-scale e-commerce companies to effectively improve online transactions. Furthermore, previous studies frequently adopt experimental design or questionnaire to be the way of data collection. However, these two

methods have inherent limitations of research bias. In our study, we use web analytics to collect comprehensive data of online behavior that it has the potential to eliminate a certain degree of research bias. In other words, the two proposed strategies are based on the support of a large amount of behavior data. In sum, the current study is one of few empirical analytics that deal the prediction of website traffic with a managerial perspective of online transaction. It is also the preliminary study which merges the theoretical underpinning of online transaction journey into the decision tree model and therefore generates meaningful insights into online shopping characteristics that increase the likelihood of e-commerce success.

6. Conclusion

The current study focuses on the prediction of online purchase which is a prerequisite for e-commerce proprietors to succeed in the marketplace. We proposed a decision tree model for predicting whether or not consumers will actually conduct their online transactions. Variables in the proposed model were derived from e-commerce behavioral process, and the value for each variable is collected from web analytics which is real data performed by online consumers over the website. The prediction accuracy of the proposed model is higher than the other three control models which brings us to conclude that the decision tree model is a quite stable approach for the purpose of data classification, especially when the goal of analysis is to explore the data insights hidden in the dataset.

6.1. Academic Contribution. Several theoretical contributions can be drawn from our findings. *First*, unlike technical studies in general, the current study tries to merge theoretical underpinning into the decision tree under which the root node of the tree denotes conversion phase of online transaction journey, while the leaf nodes of the tree represent acquisition and behavior phases of that journey. We believe that such a theoretical mapping can shed light on data insight, especially when the research focus is to understand the uncertainty of consumer online transaction. As far as we know, this research plays a pioneering role in rationalizing the necessity of theory and technique integration. *Second*, we found that the bounce rate should not be a member factor of acquisition phase. Many previous studies have focused their research point on the decreasement of the bounce rate [34–36]; however, when the goal of the research is to clarify what causes the purchase rather than the unwillingness to purchase, paying attention of the bounce rate is somewhat unhelpful. Thus, the current study theoretically points out what causes consumers to engage in online purchase. *Finally*, online transaction journey is a theoretically inseparable concept even though it can be separately measured in practice. In other words, any phase in the journey should have certain impact on each other; therefore, it is unreasonable to treat them separately when the goal is to build a prediction model of e-commerce success. The decision tree model in our study is theoretically supported by online

transaction journey which has the potential to present the overall shopping behavior found on the website.

6.2. Managerial Implications. The current study proposes a prediction model that addresses the situations in which consumers would like to conduct online transactions. It is confirmed that the number of pageviews is a prerequisite to the success of e-commerce website. Proprietors can therefore clarify some ways that have the potential to promote consumer involvement of the website. For example, proprietors can invite their consumers to visit the website for more than once in order to exchange pageviews with them. Proprietors can also use channel identification to understand in what channels are able to guide valid traffic with both the longest stay and the highest pageviews on the website. All these strategic planning are relevant to online transaction which can be derived from the proposed prediction model. In other words, when there is any variance of independent variable found in the tree, proprietors can therefore predict its possible impact on consumer online transaction.

6.3. Limitations. Similar to most studies, the current study has several limitations without exception. *First*, our study applies the decision tree model to predict website traffic; however, there can be other classification methods that may produce different research findings. Future studies are recommended to used different classification algorithms and revalidate our research findings. *Second*, the research data in our study is limited to a specific date interval. It is suggested that future studies can try to expand the date range of the research data in order to get a complete picture of website traffic. *Finally*, the prediction model of our training data is based on the large-scale e-commerce website; thus, it may not be applicable to small- and medium-sized websites with insufficient traffic. Therefore, we suggest that future studies can apply our prediction model to similar scale e-commerce websites for the purpose of research revalidation.

Data Availability

The web traffic data used to support the findings of this study were supplied by researched websites, under license and so cannot be made freely available. Requests for access to these data should be made to Han-Ping Tsen (hanping311111@gmail.com).

Conflicts of Interest

This paper does not contain any conflict of interest.

References

[1] eMarketer, "Global Ecommerce 2020: E-commerce decelerate amid global retail contraction but remains a bright spot," 2020, 2020, https://www.emarketer.com/content/global-ecommerce-2020.

[2] W. T. Wang, Y. S. Wang, and E. R. Liu, "The stickiness intention of group-buying websites: the integration of the commitment-trust theory and e-commerce success model," *Information & Management*, vol. 53, no. 5, pp. 625–642, 2016.

[3] N. Kshetri, "Barriers to e-commerce and competitive business models in developing countries: a case study," *Electronic Commerce Research and Applications*, vol. 6, no. 4, pp. 443–452, 2007.

[4] H. Liu, V. V. Krasnoproshin, and S. Zhang, "Alogarithms for evaluating and selecting e-commerce websites," *Journal of Computational Optimization in Economics and Finance*, vol. 4, no. 2-3, pp. 135–148, 2012.

[5] W. W. Chin, "How to write up and report PLS analyses," in *Handbook of partial least squares*, pp. 655–690, Springer, Berlin, Heidelberg, 2010.

[6] W. J. Tibben, "Theory building for ICT4D: systemizing case study research using theory triangulation," *Information Technology for Development*, vol. 21, no. 4, pp. 628–652, 2015.

[7] A. Bodhani, "Getting a purchase on AR," *Engineering & Technology*, vol. 8, no. 4, pp. 46–49, 2013.

[8] T. E. Barry and D. J. Howard, "A review and critique of the hierarchy of effects in advertising," *International Journal of Advertising*, vol. 9, no. 2, pp. 121–135, 1990.

[9] D. J. Kim, D. L. Ferrin, and H. R. Rao, "Trust and satisfaction, two stepping stones for successful e-commerce relationships: a longitudinal exploration," *Information Systems Research*, vol. 20, no. 2, pp. 237–257, 2009.

[10] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of big data based on its essential features," *Library Review*, vol. 65, no. 3, pp. 122–135, 2016.

[11] V. Mayer-Schonberger and K. Cukier, *Big Data: A Revolution That Will Transform how we Live, Work, and Think*, Houghton Mifflin Harcourt, London, 2013.

[12] D. Boyd and K. Crawford, "Critical questions for big DATA," *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679, 2012.

[13] M. F. Uddin and N. Gupta, "Seven V's of big data understanding big data to extract value," in *Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education*, pp. 1–5, Bridgeport, CT, USA, 2014.

[14] A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.

[15] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," *MIT Sloan Management Review*, vol. 52, no. 2, pp. 21–32, 2011.

[16] F. Shirazi and M. Mohammadi, "A big data analytics model for customer churn prediction in the retiree segment," *International Journal of Information Management*, vol. 48, pp. 238–253, 2019.

[17] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: from big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.

[18] J. P. Singh, S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. Kumar Roy, "Predicting the "helpfulness" of online consumer reviews," *Journal of Business Research*, vol. 70, pp. 346–355, 2017.

[19] D. Drossos, M. Zacharioudakis, and G. Dionysiou, "Online traffic sources and persuasion techniques: how to change consumer behavior," in *Proceedings of the 2019 3rd International Conference on E-commerce, E-Business and E-Government*, pp. 80–84, New York, NY, USA, 2019.

[20] S. J. Turner, "Website statistics 2.0: using Google Analytics to measure library website effectiveness," *Technical Services Quarterly*, vol. 27, no. 3, pp. 261–278, 2010.

[21] Q. Jiang, C. H. Tan, C. W. Phang, J. Sutanto, and K. K. Wei, "Understanding Chinese online users and their visits to websites: application of Zipf's law," *International Journal of Information Management*, vol. 33, no. 5, pp. 752–763, 2013.

[22] J. Li, A. Abbasi, A. Cheema, and L. B. Abraham, "Path to purpose? How online customer journeys differ for hedonic versus utilitarian purchases," *Journal of Marketing*, vol. 84, no. 4, pp. 127–146, 2020.

[23] C. Tangmanee, "Comparisons of website visit behavior between purchase outcomes and product categories," *IJRBS*, vol. 6, no. 4, pp. 1–10, 2017.

[24] J. C. C. Lin, "Online stickiness: its antecedents and effect on purchasing intention," *Behaviour & Information Technology*, vol. 26, no. 6, pp. 507–516, 2007.

[25] V. G. Morwitz, "Methods for forecasting from intentions data," in *Principles of forecasting*, pp. 33–56, Springer, Boston, MA, 2001.

[26] Z. Anesbury, M. Nenycz-Thiel, J. Dawes, and R. Kennedy, "How do shoppers behave online? An observational study of online grocery shopping," *Journal of Consumer Behaviour*, vol. 15, no. 3, pp. 261–270, 2016.

[27] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[28] A. H. C. Hwang, J. Oh, and A. C. Scheinbaum, "Interactive music for multisensory e-commerce: the moderating role of online consumer involvement in experiential value, cognitive value, and purchase intention," *Psychology & Marketing*, vol. 37, no. 8, pp. 1031–1056, 2020.

[29] P. C. Verhoef and K. N. Lemon, "Successful customer value management: key lessons and emerging trends," *European Management Journal*, vol. 31, no. 1, pp. 1–15, 2013.

[30] D. Westreich, J. Lessler, and M. J. Funk, "Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression," *Journal of Clinical Epidemiology*, vol. 63, no. 8, pp. 826–833, 2010.

[31] P. Dekker, "Tocqueville did not write about soccer clubs: participation in voluntary associations and political involvement," in *Modernizing democracy*, pp. 45–47, Springer, New York, 2014.

[32] M. Zhao and J. Xie, "Effects of social and temporal distance on consumers' responses to peer recommendations," *Journal of Marketing Research*, vol. 48, no. 3, pp. 486–496, 2011.

[33] C. W. Choo, B. Detlor, and D. Turnbull, "Information seeking on the web–an integrated model of browsing and searching," in *Proceedings of the Annual Meeting of the American Society for Information Science (ASIS)*, pp. 3–16, Washington, 1999.

[34] A. Gofman, "Consumer driven multivariate landing page optimization: overview, issues and outlook," *The IPSI BgD Transactions on Internet Research*, vol. 3, no. 2, pp. 7–9, 2007.

[35] D. Schellong, J. Kemper, and M. Brettel, "Generating consumer insights from big data clickstream information and the link with transaction-related shopping behavior," in *Twenty-Fifth European Conference on Information Systems (ECIS)*, pp. 1–15, Guimaraes, 2017.

[36] J. Xun, "Revisiting the two-stage choice model: an empirical study of consumer choice on brand website visits," *Behaviour & Information Technology*, vol. 33, no. 11, pp. 1192–1207, 2014.