International Conference on Business Research
University of Moratuwa, Moratuwa, Sri Lanka
October 27, 2020, 233-252.

BUSINESS RESEARCH UNIT
FACULTY OF BUSINESS
UNIVERSITY OF MORATUWA

TRAVEL DEMAND ANALYTICS BASED CUSTOMER-SERVICE DECISION MODEL

Abilashini Jeyaratnam[1], I. Mahakalanda[2], and Tiloka de Silva[3]

[1,2,3]*Department of Decision Sciences, University of Moratuwa*

Emails: [1]Abijeya@live.com; [2] indram@uom.lk; [3] tilokad@uom.lk

Abstract

The tourism industry adds high value to the economy in Sri Lanka by attracting people from all around the world. Within the tourism industry, a large amount of data is collected with regards to user demographics and their purchases on a daily basis. Though less used in the tourism sector of developing countries like Sri Lanka, data analytics techniques can be utilized to understand the customer better and thereby add additional value to tourism organizations. This research uses customer records from a tourism operator to develop a model to predict the complexity of a customer's requirements as well as group customers according to the complexity. While K-means clustering is used to group the customers as easy, moderate and complex customers, an ordinal logistic model is used to build the predictive model. The outputs obtained through the data analytic models proposed in this study will support the study organization to manage their customers efficiently and provide better attention to their needs.

Keywords — Business Analytics, Tourism Industry, Customer segmentation

1. Introduction

In the modern day and age of competitive business, it has become vital for companies across all industries to look into new ways of improving business efficiency and profitability through the understanding of the various factors that play a role on how each industry behaves. Data analytics plays a major role in providing better overall visibility on business progress which in turn is useful to make the necessary adjustments to improve continuously. The tourism industry is one area where continuous improvement to their products and services is required to remain competitive.

Tourism industry is one of the largest revenues generating industries in Sri Lanka, which adds high value to the economy by attracting people from all around the world. A tourism product or package may be conceived as an amalgamation of tangible and intangible elements involving natural, historical, socio cultural factors, accessibility, infrastructural facilities, recreational and shopping facilities, etc. (Das & Mukherjee, 2008). Tourism products and services vastly vary based on each individual customer's preferences and background, making it a difficult task for companies to identify and address each requirement efficiently. A product should cover the complete experience of a tourist starting from when they leave their home until the time they return. It is highly important for organizations in the tourism industry to understand such varying requirements of each customer proactively, in order to retain and attract more customers by providing them with solutions catered towards each type of individual. Providing optimum tourism products/packages catered towards each specific customer and market effectively and efficiently would allow companies to eventually attain better profitability and market position.

One of the main problems that a tourism operator faces is the time and resources needed to understand the complex requirements of different customers, in order to provide appropriate packages. For instance, in some cases the final package that a customer would purchase varies widely from their initial requirement. This poses a challenge to an organization in serving their customers efficiently as they work on multiple customer requests with limited resources. It would therefore be beneficial to such an organization to be able to categorize different types of customers and better understand the products that they are likely to purchase.

This paper analyzes customer data from a selected tourism operator and converts the data into meaningful information in order to categorize customers into groups and predict their purchase behavior.

To achieve this, a prediction model based on ordinal logistic regression is used to predict the products that a customer would purchase as well as the extent of the deviations from initial requirements. Along with this, unsupervised machine learning would be performed using K-means clustering to categorize users into 3 different categories [easy, moderate and complex]. Once the model is in place, inputs from the new users could be used to identify which category they fall under so that their requests can be processed accordingly. This can then be used to provide faster and customized services catered towards each customer by using the proposed data analytics techniques.

While the study will focus on a single company in a specific industry, the understanding obtained through such research could be easily adapted into different domains allowing it to be used across multiple applications. This would be highly significant in Sri Lanka where there are a multitude of industries at a developing stage. The use of analytics techniques studied in this research could therefore help companies in different industries to improve their business by following similar practices.

The rest of the paper is organized as follows: Section 2 briefly reviews relevant literature. Section 3 describes the methodology. Section 4 presents analysis and discussion. Finally, Section 5 concludes.

2. Literature Review

This literature survey is in two parts: literature on the tourism industry and travel packages, and literature on the applications of data analytics techniques in the tourism industry.

A. Tourism Industry

Tourism is a major business area that represents 7% of total exports worldwide. It is identified to be a continuously growing market, especially with people being more aware of new travel destinations through technology and cheap travel options (Holloway, 2006; Hall, 2008; Jafari & Xiao, 2016].

There are two main types of tourists: the ones who travel on business and the tourists who travel for personal reasons such as leisure, visiting friends and relatives, education; religious trips; sporting activities. In the case of business travel the choice of when and where to travel is out of the traveller's control and the intention of business travel is generally not focused towards enjoying the facilities and experiences at the travel destination. Personal travel is arranged with the intention of experiencing different attractions and activities that a tourist destination has to offer. Usually, a personal tour is planned well in advance keeping weather, tourist seasons and various other factors in mind (Swarbrooke & Horner , 2009).

A tourist destination is a location that consists of attractions required to entertain tourists and meet their expectations (Djurica & Djurica, 2010). Destinations contain tangible characteristics and possess several physical attributes, such as attractions, amenities, buildings, landscapes, etc. However, the perception of a tourist which is another major factor is not tangible. Factors including hospitality of the hosts, the environment created by a specific event and many other emotions could be triggered by a certain tourist destination (Murphy, Pritchard & Smith, 2000; Hall 2008; Camilleri, 2018). Tourist destinations could be categorized into different sections such as adventure, culinary, cultural, ecotourism and urban tourism destinations.

A product in tourism is a combination of physical and emotional fulfilment experienced by tourists throughout their journey to and from their travel location and duration of the stay. A tourism product focuses on facilities and services designed to meet the needs of the tourist. An overall tourism product consists of a combination of all facilities that a tourist would need during their visit. (E.g accommodation, transport, attractions, amenities and other facilities such as food and activities). While the tourist product should be distinguished from the destination, the destination is not the end product (Koutoulas, 2004).

How a tour package is defined varies significantly based on the literature that one would refer to. A common definition of a tour package is that it is a method of travel arranged by tour organizers, where the tour package includes a combination of different products (e.g. accommodation, transportation, tours, activities, etc.) In other contexts, it is also referred to as a package that combines tours, activities, transport and lodging, pre-organized by a tour operator (Jafari & Xiao (2016).

Sri Lanka has become an increasingly popular tourist destination specially in the last decade. Sri Lanka was recognized by Lonely Planet as the top destination in the world for tourists in 2013 and 2019. In 2015, Forbes magazine declared Sri Lanka among the "top ten coolest countries" to visit. In 2016 Lonely Planet, Rough Guides, The Guardian, The New York Times have identified Sri Lanka as a top location to visit [Ministry of Tourism Development and Christian Religious Affairs, 2017).

Due to this highly suitable background for tourism, Sri Lanka can be considered as an ideal sample for the study carried out in this research. The amount of options and activities available along with the different types of tourists that visit the county from all around the world, creates an ideal data set with high diversity which helps with training an accurate model.

B. Data analytics in Tourism Industry

Data Analytics has become an increasingly popular topic in many industries where it is used to make an impact on the business. Data analytics and machine learning techniques are slowly being considered in the tourism industry where it could help with the challenges faced by tourism organization as well as with growing and expanding their businesses efficiently (Alcántara-Pilar, Del Barrio-García, Crespo, & Porcu, 2017). By reorganizing how the available data is used for the betterment of the business, companies would be able to achieve more with higher efficiency.

Today, many organizations consider their existing data to be a valuable asset. While many industries related to technology are commonly utilizing data analytics to add value to their business, this concern is still valid for many industries such as tourism where the use of technology has not picked up considerably. As a result, when it comes to the implementation of such innovative data analytics and machine learning models for value creation in tourism industry, the available research is mostly limited to theory and a few trials.

When it comes to tourism, refining and processing of the available data using data analytics to create meaningful information would help add value to an organization in the tourism industry. Using data analytics to the available data can lead to different outcomes such as; customer segmentation, customer sentiment analysis, recommendation engines and predictive analysis (Activewizards, 2019). The use of data analytics and machine learning techniques could provide organizations in the tourism industry with multiple benefits such as improving customer service, better decision-making power due to intelligence and visibility, efficient identification of viable products/services along with the risks involved.

Machine Learning is widely considered as a discipline which is a special branch of artificial intelligence. In machine learning, a model is trained to perform a certain task on a new or unseen set of data, based on the exposure that the model has gained through the learning data set (Upadhyay, 2018; Das, Dey, Pal & Roy 2015). Machine learning tasks are classified into two main categories; supervised (the goal is to map the input data to the set of output data set) and unsupervised learning (the goal is to label the unlabelled data) (Simon, Deo, Selvam & Babu, 2016; Praveena & Jaiganesh, 2017).

Integration of machine learning techniques is changing the tourism industry and the use of its application is increasing globally. Few major applications of machine learning are recommendation engines, flight fare and hotel price forecasting, intelligent travel assistants, optimized disruption management and etc. (Ray, 2019). Specific examples include a recommender system that has been designed to assist customers by matching their preferences with the services and options which are available among the Indian tourism operators (Muthukumaran, 2020), and a system that was designed for segmenting tourists and detect tourism hotspots based on tourists' geo-located blogging behaviour using machine learning techniques (Kaufmann, Siegfried, Huck. & Stettler, 2019).

Research has also been done in tourism demand forecasting using machine learning techniques. A study on tourism in Hong Kong has been conducted based on different machine learning techniques and found that regression, radial basis function, k-nearest neighbour regression and Gaussian process models can provide better performance in predicting tourism demand (Kamel, Atiya El Gayar & El-Shinshiny, 2007).

3. Methodology

This paper uses quantitative methods and secondary data to meet the research objectives.

A. Research Design

The research design used in this paper is given in Figure 1. The first phase of the study was understanding the business where the project objectives were derived by consulting the study organization. In the next phase of data understanding, raw data was collected and reviewed in order to perform the data preparation. During the third phase, the data was then cleaned and turned into meaningful data with the selected variables. the two models selected were executed independently. Ordinal Logistic Regression technique was implemented as a prediction model based on initial preferences while K-Means clustering model was implemented to cluster the customers under 3 categories namely; easy, moderate and complex. As the final stage of the study, the models were evaluated, and the results have been interpreted to understand how the outputs of each model could be used to achieve the objectives of the study.

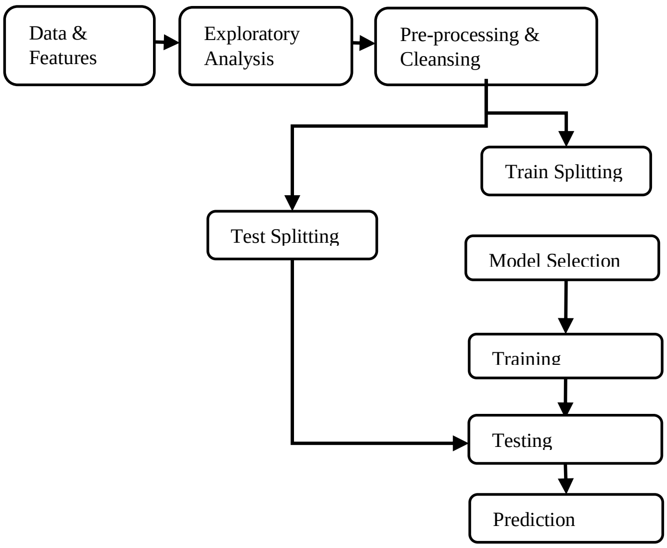


Figure 1: Research Design

B. Data

In order to collect the required data, secondary sources have been used which will finally help to achieve the objectives of this research. Secondary sources used in this study include the selected organization's databases and archives, company information, websites, online research papers/journals related to data mining techniques used in tourism industry, tourism packages, attributes, products and services.

This study used 300 customer records from the organization in order to perform the study with data comprising of the variables mentioned in the operationalization table. Collected data included demographic factors (e.g gender, age, nationality, marital status), traveller type, indicated period of stay, destinations and activities along with the respective finally agreed period of stay, destinations and activities.

C. Analytical methods

The study is focused on developing a prediction system to identify and segment easy, moderate and hard customers in order to reduce the operational time required to identify and process requests of each customer type. Here, easy, moderate or hard is measured by how much the customer deviates from their initial requirement – customers who deviate less are categorized as easy and those deviating more are categorized as hard. The analysis was supported mainly using Machine Learning algorithms through platforms such as Azure Machine Learning Studio and Python. IBM SPSS statistical software has been used for descriptive analysis and Microsoft Excel Software packages were used to clean and process the data in order to conduct the analysis.

1) Complexity prediction model

An ordinal logistic regression model was developed to predict deviations from the initially specified length of the trip, and number of destinations and activities. The conceptual framework developed from the given data of traveller's destination preferences and actual purchase of destinations and activities along with the relevant variables are presented in Figure 2.

According to the conceptual framework, the key dependent variables were the deviations in the number of indicated period of stay (planned no. of days) from the finally agreed period stay (actual no. of days) for the trip and the number of planned destinations and activities from the actual number of destinations and activities in the consumed package. Given the ordinal nature of the dependent variables, ordinal logistic regression was selected as the most appropriate model for prediction.

Performance of the model can be obtained by calculating precision, recall and F1 score of the output variables. The following measures are used to calculate the performance:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)}$$

Precision is the ratio between correctly predicted positive values and the total predicted positive values. Whereas, recall is the ratio obtained by considering the correctly predicted positive values among all the actual values. F1 score is the weighted average of recall and precision values by considering both positive and negative values. F1 score can provide a better understanding of prediction model's accuracy.

In order to train the model, 80% of the input data set was defined as training data while the remaining 20% was used as a test data set for the purpose of testing the model.

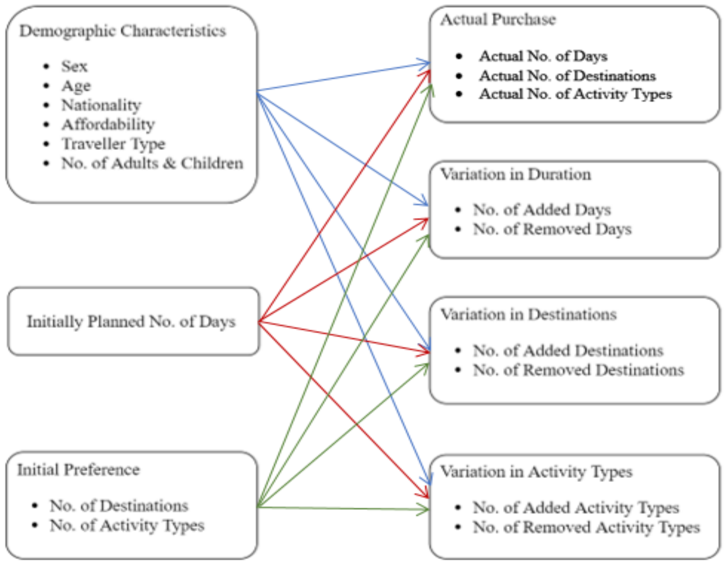


Figure. 2. Conceptual framework for predictive model

For the purpose of implementing ordinal logistic regression, Bevel library has been used as it provides a detailed statistical summary that allows the identification of the significant relationships between the independent variables and dependent variables. Additional libraries and modules used for the study include; Pandas, Scipy, Numpy, Requests and Seaborn.

2) *Customer Segmentation*

The method used to design the model does not require a conceptual framework. The entire data set will be used; no test data is required.

A K-Means clustering model was used to categorize the customers. The variables used for the clusters are the preferred, variation between indicated period of stay at the time of initial contact and finally agreed period of stay, destinations and activities. The number of clusters required were already defined by the organization (k = 3) and the centroids were calculated using K++ means initialization method, which is one of the latest methods, allowing for better accuracy in clustering.

4. Analysis and Discussion

A. Descriptive Statistics

The customer records from the dataset include demographic characteristics such as region, gender and age group of booking customer, preferred budget and type of traveler (solo, couple, family or group).

Most of the bookings were for travelers from Northern Europe (48.3%), Oceania (14.3%) and South Asia (10.3%). The majority of customers making the booking were in their thirties (44.3%) and forties (22%). The majority (45%) of customers requested mid-range packages (between $200 and $400 per person per day), while 23% requested high-range packages. Finally, more than half of travellers were travelling as couples, followed by 25% travelling with their families.

The remaining information in the database was related to the indicated period of stay and finally agreed period of stay denoted by "planned no. of days" and "actual days" respectively, destinations and activities specified initially as well as finalized. The descriptive statistics of these variables are given in Table 1.

Table 1: Summary statistics

Variable	Min	Max	Mean	Std. Dev.
Planned No. of Days	1	16	7.46	4.165
Preferred Destination Count	0	8	2.05	1.852

Preferred Activity type Count	0	7	1.47	1.315
Actual Days	1	14	7.1	4.01
Actual Destination Count	1	8	3.41	1.688
Actual Activity type Count	1	9	4.1	1.629
No. of Added Days	0	3	0.14	0.494
No. of Removed Days	0	6	0.51	1.08
No. of Added Destinations	0	6	1.5	1.681
No. of Removed Destinations	0	4	0.2	0.584
No. of Added Activity Types	0	9	2.75	1.798
No. of Removed Activity Types	0	3	0.17	0.447

Source: Author developed

As shown in Table 1, the average initial length of the trip was 7.46 with 2.05 destinations and 1.47 activities, on average. Note that some customers did not specify any destinations or activities at the onset and so the planned counts for these variables is zero. The key dependent variables of interest are the deviations between planned and actual number of days agreed, destinations and activities. As the summary statistics show, the average number of removed days exceeds the average number of added days but the number of added destinations and activities exceeds the number removed, indicating that on average, trip duration is reduced but locations and activities are added on.

B. Results from Ordinal Logistic Regression

In this section we present the results obtained from the predictive model. As mentioned previously, an ordinal logistic regression model was estimated with 80% (N=24) of the dataset used to train the model and the remaining 20% to test it. Results of the training model estimated for each of the dependent variables (measures of deviation) are given in Table 2.

The results in Table 2 can be used to identify which of the explanatory variables have a significant impact on how much a client deviates from their initial specification. While none of the variables are significant in predicting the number of added days, all of the other models have two or more significant predictors. The variables with a significantly positive effect provide an indication of the type of customer that requires more attention. For instance, the number of initially specified days is positively associated with upward and downward deviations of trip duration, destinations and activities. This means that customers who initially plan longer trips tend to make more adjustments to their package before finalizing and so, require more customization. The tourists initially specifying a larger number of destinations are more likely to remove destinations but less likely to remove activities or period of stay. Customers starting out with a larger number of activities are also more likely to remove destinations and activities. It seems reasonable that customers opting for larger packages be given more attention as this type of customer bring in more revenue to the company.

Table 2. Results from ordinal logistic regression

	Days		Destinations		Activities	
Explanatory variables	**Added**	**Removed**	**Added**	**Removed**	**Added**	**Removed**
Male	0.269	-0.025	0.497	-0.083	-0.048	0.118
Region (base=Africa+Middle East)						
Europe	13.168	-1.531*	-0.559	-1.120	-0.839	-0.473
America	14.037	-1.315	0.247	-1.476*	-0.778	0.979
Asia	13.994	-1.368	-0.105	-1.314	-1.116	-1.317
Oceania	13.967	-0.650	-0.751	-2.305***	-1.669	0.042
Age group (base=Sixties)						
Twenties	0.140	-0.579	0.995	0.044	-0.303	-0.232
Thirties	0.693	-0.763	0.872	0.347	-0.019	-0.381
Forties	1.195	-0.574	1.326**	0.991*	0.644	0.377
Fifties	0.619	0.043	1.263*	0.528	0.551	0.955
Specified budget (base=High)						
Low	0.294	-0.114	0.784**	0.299	0.115	-0.441
Mid range	-0.313	-0.395	0.719**	0.231	0.109	0.028
Traveller type (base=Solo)						
Couple	0.520	-0.253	0.168	0.062	-0.116	-0.577
Family	-0.817	-0.407	-0.667	-1.246**	-1.138*	-0.334
Group	-1.414	-0.133	-0.565	-0.307	-0.331	0.126
Group size						
No. of Adults on trip	0.253	0.002	0.067	0.159	0.133	0.011
No. of Children on trip	0.335	0.389	0.398	0.457	0.418	0.219
Initial specification						
No. of days	-0.043	0.269***	0.483***	0.592***	0.544***	0.067
No. of destinations	0.037	-0.261**	-1.415***	0.133*	0.003	-0.328**
No. of activities	0.302	-0.098	0.159	0.195*	-1.498***	0.687***

Source: Author developed

Note: Significance * 10% ** 5% *** 1%

From among the demographic characteristics, it can be seen that tourists from Europe have significantly lower number of removed days than tourists from the base region of Africa and the Middle East while American and Australian tourists are significantly less likely to remove destinations. Moreover, tourists who request budget or mid-range packages are significantly more likely to add more destinations to their package later than tourists requesting luxury packages. Among the age groups, it can be seen that tourists in the forties are more likely to add or remove destinations compared to tourists in their sixties. There are no significant differences in the deviations by traveller type except for family travellers who are significantly less likely than solo tourists to add activities or remove destinations.

Table 3 illustrates the precision, recall and F1 scores of the prediction model for the different dependent variables. F1 score can be interpreted as the weighted average of both precision and recall values in machine learning. Considering the F1 scores of the output variables, it can be said that the model is able to predict the output variables correctly, except in the case of number of added destinations and number of added activity types, since F1 scores for these variables are closer to 0.

However, it should be noted that since the test data used in this model is small, the reliability of predictions may not be adequate. Increasing the amount of data is essential for improving this model further and obtaining more reliable performance metrics.

Table 3. Prediction Metrics

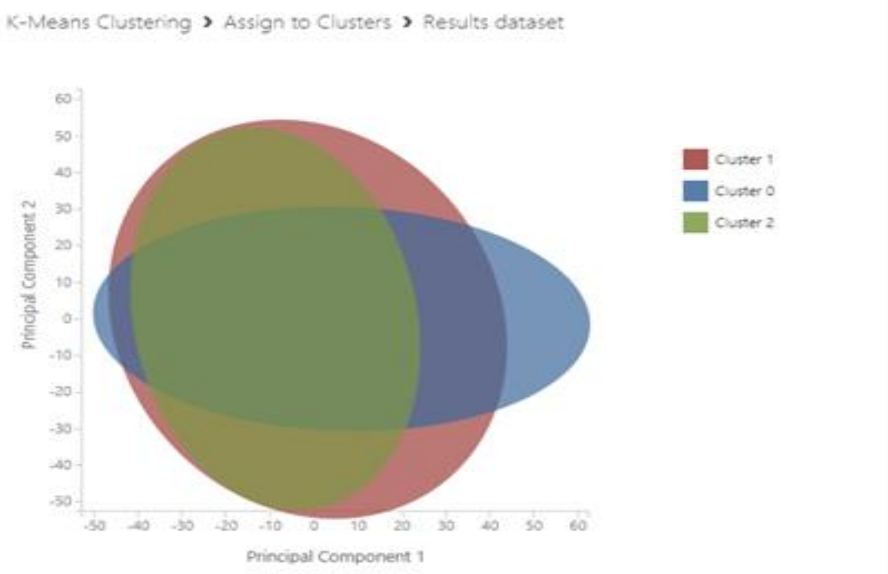
Prediction	**Precision**	**Recall**	**F1-score**
No. of added days	0.72	0.85	0.78
No. of removed days	0.69	0.82	0.75
No. of added destinations	0.55	0.5	0.49
No. of removed destinations	0.87	0.92	0.89
No. of added activities	0.4	0.35	0.36

No. of removed activities	0.78	0.87	0.82

Source: Author developed

C.Results from K-means clustering

In this section, we present the results from the clustering analysis used to segment customers into easy, moderate and complex. When configuring the K-Means clustering module, the number of centroids were set to three while the initialization method chosen was K-Means++ with an iteration of 100. Since the customers should be segmented considering the variation between initial preferences and actual purchases, the variables selected for the analysis included indicated period of stay, finally agreed period of stay, destinations, and activities, destinations and activities, and added/removed no. of days, destinations and activities. Then the model needs to be assigned into clusters against the entire data set which has been accomplished using "assign to clusters" module. (Illustration in Figure 4.).



Clustering visualization

As Figure 4 indicates, though three clusters were specified, it appears that clusters 1 and 2 are similar whereas cluster 0 is different, though here too there is significant overlap. To map the level of complexity of the customer requirements with the clusters, we now examine a summary of the variation between planned and actual for each cluster as given in Table 3.

Variations that occurred in cluster 0 are minimal when compared to cluster 1 and 2. Overall, when comparing cluster 1 and 2, it can be said that cluster 2 has higher variation than cluster 1. Therefore, customers in cluster 0 have been labelled as Easy, cluster 1 as Moderate and Cluster 2 as Complex customers. Depending on the customer group, the study organization can decide their approach in attending each customers' needs. The trained model can be tested using the predictive environment created by ML Studio where data can be filled into a form where the results can be obtained to check the accuracy of the model.

Variation	Cluster 0	Cluster 1	Cluster 2
Range of Added Days	0 - 1	0 - 3	0 - 2
Range of Removed Days	0 - 3	0 - 6	0 - 4
Range of Added Destinations	0 - 4	0 - 6	0 - 6
Range of Removed Destinations	0 - 3	0 - 3	0 - 2 & 4
Range of Added Activities	0 - 4	0 - 5	0 - 9
Range of Removed Activities	0 - 2	0 - 2	0 - 3

Source: Author developed

5. CONCLUSION

In this paper, a model was derived to predict varying purchase patterns of different customers. Ordinal logistic regression has been used to train a model using previous customer data. This developed model will help the company to identify the customers who will vary significantly from their initial preference by predicting their purchase patterns. Based on the level of variation predicted, the study organization can identify the amount of attention/resources that should be allocated to each customer. This would allow the organization to improve efficiency and effectiveness of their customer service. The generated model was able to predict the output with a moderate accuracy level, though an important caveat to reliability of the predictions is the small size of the dataset.

K-means clustering was also used to train a model using the existing user data to categorize users in to 3 groups [easy, moderate and complex]. By using this model, new users could be segregated into the 3 different groups by providing the required input fields. This in turn helps the organization identify the customers that would require special attention, so that they can be catered based on their requirements.

6. LIMITATIONS & FURURE RESEARCH

There are several limitations identified in this analysis. The key limitation is related to sample size, which affects the reliability of the predictions generated by the model. By increasing the number of customer records, the test data can be increased and the performance of the model can be enhanced further. The second limitation is the limited scope of the available input data. For instance, data on areas such as transportation and accommodation preferences could help improve the scope of the prediction model, thereby providing the organization with more details and visibility. In addition, the available data set is from a single organization in a specific market. To provide a generalized model, inputs from multiple organizations from different market areas would be required.

In extensions to this research it would be useful to take into account the type of destinations and activities that were added or removed rather than simply using counts of the deviations as there might be valuable insights that could be generated. It would also be beneficial to the organization if the results of the prediction and clustering models were combined to calculate a risk score for prospective clients.

References

Activewizards. (2019). Top 7 Data Science Use Cases in Travel, Retrieved from https://www.kdnuggets.com/2019/02/top-7-data-science-use-cases-travel.html.

Alcántara-Pilar,J., Del Barrio-García, S., Crespo E., & Porcu L.(2017). "Toward an understanding of online information processing in e-tourism: does national culture matter?" in Journal of Travel & Tourism Marketing, no. 15, 1128-1142.

Camilleri, M, (2018). *The Tourism Industry: An Overview In Travel Marketing, Tourism Economics and the Airline Product. Switzerland*: Springer.

Das D., & Mukherjee, K. (2008) "Development of an AHP-QFD framework for designing a tourism product," *International Journal of Services and Operations Management*, *4*(3).

Das, S., Dey A., Pal A., & Roy N.(2015). Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *International Journal of Computer Applications*, 115(9), 31-41.

Djurica M., & Djurica N.(2010). Tourism destination marketing management. *Tourism and Hospitality Management,* 890-901.

Hall, C.M. (2008) *Tourism planning: Policies, processes and relationships*, London: Pearson.

Holloway, J.C. (2006) *The business of tourism, London*: Pearson Education.

Jafari, J., & Xiao H. (2016) *Encyclopedia of Tourism, Switzerland*: Springer International Publishing.

Koutoulas, D.D (2004). Understanding the tourist product, Greece: Interim symposium of the Research Committee on International Tourism.

Kamel, N., Atiya A.F., El Gayar, N., & El-Shinshiny, H. (2007). Tourism Demand Foreacsting using Machine Learning Methods, unpublished.

Kaufmann, M., Siegfried, P., Huck., L., & Stettler, J. (2019). Analysis of Tourism Hotspot Behaviour Based on Geolocated Travel Blog Data: The Case of Qyer*, Internation Journal of Geo-Information, 8* (11), pp. 493.

Ministry of Tourism Development and Christian Religious Affairs. (2017) Sri Lanka Tourism Strategic Plan 2017 - 2020. Colombo: Ministry of Tourism Development and Christian Religious Affairs.

Murphy, P., Pritchard P.M., & Smith, B. (2000). The destination product and its impact on traveller perceptions*. Tourism Management*, *21*(2), 43-52.

Muthukumaran, B. (2020). Machine Learning Changing the Dynamics of Tourism Industry in India, Retrieved from Silliconindia: https://industry.siliconindia.com/viewpoint/cxoinsights/machine-learning-changing-the-dynamics-of-tourism-industry-in-india-nwid-10142.html

Praveena M., & Jaiganesh, V. (2017). A Literature Review on Supervised Machine Learning Algorithms and Boosting Process. *International Journal of Computer Applications*, *169* (8), pp. 32-35.

Ray, T. (2019). Machine Learning, AI and Big Data Analytics in the Travel & Hospitality Industry, Retrieved from https://www.stoodnt.com/blog/machine-learning-ai-big-data-analytics-travel-hospitality-applications-scopes-job-market/

Simon, A., Deo, M.S., Selvam V., & Babu, R. (2016). An Overview of Machine Learning and its Applications. *International Journal of Electrical Sciences & Engineering* (IJESE), 1, 22-24.

Swarbrooke J., & Horner, S.(2001) *Business Travel and Tourism*, Oxford: Butterworth-Heinemann.

Syratt G., & Archer J.(2003). *Manual of travel agency practice*. Oxford: Buttweworth-Heinemann.

Upadhyay, N. (2018). *Big Data: Management and Analytics,* India: Cengage.