Machine Learning Application in Analyzing Online Customer Journey

*** Subhabaha Pal**
**** Sampa Pal**

Abstract

This is the era of internet which has given rise to multiple opportunities for online users to purchase from digital marketplaces. Understanding the consumer journey of users is important and it is also important to understand their behaviour in the online space. This is an attempt to analyze the behaviour of online users and to predict whether they will be purchasing or opting for a service based on the last touch-point they have accessed (whether it is a customer-initiated touchpoint or a company-initiated touchpoint).

Keywords **: CIC, customer journey, e-commerce, FIC, machine learning**

I. INTRODUCTION

This is the era of the internet. We live in times where the web world is growing more rapidly each and everyday. It doubles in size every 12 months. As per estimates, the internet will occupy about 44 zettabytes of space by 2020 [4]. The advent of internet has given rise to immense opportunities for internet users to be effective online consumers. Online consumers have gained access to a variety of offerings and a number of options before making a desired purchase [8]. The internet users have a lot of alternatives to consider. They have the option to decide from a number of channels and media. They can pick what is best suited for them [5]. Studying the process of translating these options into communication channels, it is possible to infer about the consumer's path-to-purchase [2]. A consumer purchase journey describes a part of a general customer experience which consists of distinct contacts (touchpoints) among the customer and the firm. Thus, the purchase journey can be defined as "the process a customer goes through all stages and touchpoints" [5].

Businesses can derive insights from navigational paths to understand and may predict online consumer behaviour [3]. A touchpoint refers to "a customer contact point, or a medium through which the firm and the customer interact" [10]. Various researchers, for example, [11], [5], and [12] have divided the touchpoints into two basic categories: (1) The customer-initiated or customer-owned contacts (CIC); and (2) the firm-initiated or brand-owned contacts (FIC). A customer-initiated contact, by definition, is a contact that is initiated by a customer. This can be either a generic/ more specific search in a search engine machine, a type in the company's website or an action made in an application ([5], [6], [12], [13]). While a firm-initiated contact is any touch point that results from a company's initiative and usually, is also managed by it. Channels such as display, retargeting, affiliate, and e-mail advertising have been classified as contacts initiated by the firm ([13],[5], [6], [1]). An interesting work is understanding which touchpoints ultimately lead to conversion. One can draw the inference that consumers are using multiple channels to evaluate alternatives, search information, and make conversions ([6], [7]). Thus, consumers' interactions with various channels can lead to purchases from a company [6] or make consumers buy from competitors.

Consumers initiating contacts by themselves are expected to have more conversions in the firm's website ([9],[12],[14]).However, the nature of consumer

Manuscript received March 22, 2019; revised April 25, 2019; accepted April 28, 2019. Date of publication June 6, 2019.
*S. Pal is Senior Faculty (Data Science and Machine Learning) with Manipal ProLearn (Manipal Academy of Higher Education – South Bangalore Campus), 3rd Floor, Salarpuria Symphony, 7, Service Road, Pragathi Nagar, Electronic City Post, Bengaluru, India -560 100. (email: subhabaha.pal@manipalglobal.com)
**S. Pal is Consultant with Arminus Software Private Limited, HSR Layout, Bengaluru, India - 560 102. (email:sampadas1508@gmail.com)
DOI : 10.17010/ijcs/2019/v4/i3/146161

purchase journey involves more than CICs; customers are also coming across firm-initiated contacts which can influence their purchase decisions [6].

Yet not all the consumers express the same preferences when navigating online. It is important to consider consumer heterogeneity across options and to not consider their behavior as a whole. Demographics and behavioural patterns differ among various segments and play a role throughout the path-to-purchase [8].

However, to the researcher's knowledge, no research has still addressed the effect of consumer heterogeneity on the various touchpoints and its effect on conversions.

The basic idea behind this paper is giving an overview on how machine learning can be effectively used to analyze customer journey and get insights from the touch-points visited by customers. It is also tries to predict based on the first touch-point visited, whether the transaction will be converted into purchase or not.

II. DATASET

The data used for this study are collected and provided by GfK (Growth for Knowledge), a German market research company which is among the four biggest worldwide. The data is event-based and is gathered by GfKCrossmedia link. Crossmedia link is an innovative technology which tracks and measures consumers' behavior across different media based on a single source cross-media panel (www.gfk.com).

The observation period is from June 1, 2015 to September 32, 2016 and contains in general, 12,252 consumer journeys, 3,674 total bookings of which 192 are the focal's brand. The data set accommodates information both about the journey and demographic characteristics of the consumers. Such information includes type of touchpoints visited, whether a journey has led to conversions, type of device used to assess a contact, and consumer characteristics such as age, income, occupation, social class etc.

The data is available at https://drive.google.com/open?id=1lYmf2aMUm6c3m0nYh7OfBu3sRHIezhy6

III. DATA DESCRIPTION

The data is on purchase journey of customers in a travel website. The detailed description of the fields of the dataset is given in the file codebook. The data provides useful information about the customer journey. We give a basic overview here on how this data can be used for getting useful information about the customer journey.

IV. ANALYSIS OVERVIEW

The data gives details of customer journeys before making a purchase. Each row in the file TravelDemo.csv gives details of the touch-points in each of the transactions. The last touch-point is of considerable importance and it mainly decides whether anything would be purchased or not. We can take the last touch points and build a logistic based classification model to predict whether a transaction will be converted to a purchase based on the last touch point (whether it is customer-initiated or farm-initiated). As a preliminary analysis, we worked on whether customer initiated contact or farm initiated contact in the end tends to convert the transaction into purchase. We have fitted a logistic regression model and the results of the analysis is as follows:

```
Deviance Residuals:
   Min      1Q  Median      3Q     Max
-0.2377  -0.1004  -0.1004  -0.1004   3.3453

Coefficients:
              Estimate  Std. Error  z value   Pr(>|z|)
(Intercept)   -3.5526     0.5856    -6.066   1.31e-09 ***
dummy_CIT1    -1.7353   0.6052  -2.867      0.00414 **
mobile          -0.3038    0.3688  -0.824  0.41010
—
Significance codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 655.58 on 10481 degrees of freedom
Residual deviance: 649.32 on 10479 degrees of freedom
AIC: 655.32

Number of Fisher Scoring iterations: 8

Confusion Matrix and Statistics

pred_4    0      1
       0 3443  16
       1   34    2

Accuracy : 0.9857
          95%  CI : (0.9812, 0.9894)
   No Information Rate : 0.9948
p-value [Acc> NIR] : 1.00000
```

Kappa : 0.0677

Mcnemar's Test pValue : 0.01621

Sensitivity : 0.99022
Specificity : 0.11111
PosPredValue : 0.99537
NegPredValue : 0.05556
Prevalence : 0.99485
Detection Rate : 0.98512
Detection Prevalence : 0.98970
Balanced Accuracy : 0.55067

'Positive' Class : 0

The accuracy (0.9857) is very good for the model (we have taken threshold as 0.05). Hence, we can conclude that using this model we can predict whether initial customer touchpoints or firm initiated touchpoints will be able to get the customer to opt for the service (or purchase the service).

V. CONCLUSION

This paper shows some outline of how machine learning techniques can be used to analyze data on customer journey and get useful insights from it. The model discussed in the paper explains how to predict whether some service will be purchased or not based on the last touch-point the user has chosen to access. This paper can be further enhanced to analyse the effectiveness of each touch point in details.

REFERENCES

[1] E. Anderl, I. Becker, F. V.Wangenheim, and J. H. Schumann, “Mapping the customer journey: Lessons learned from graph-based online attribution modelling,” *International Journal of Research in Marketing*, vol. *33*, no. *3*, pp. 457-474, 2015.doi: http://dx.doi.org/10.2139/ssrn.2685167

[2] P. D. Berger, R. N. Bolton, D. Bowman, E. Briggs, V. Kumar, A. Parasuraman, and C. Terry,“Marketing actions and the value of customer assets: A framework for customer asset management,”*Journal of Service Research*, vol. *5,* no.*1*, pp. 39-54,2002.

[3] R. Bucklin and C. Sismeiro, “Click here for internet insight: Advances in clickstream data analysis in marketing,” *Journal of Interactive Marketing*, vol. *23*, no. *1*, pp. 35-48, 2009.doi: https://doi.org/10.1016/j.intmar.2008.10.004

[4] J. Dold and J. Groopman,“The future of geospatial intelligence,” *Geo-Spatial Information Science*, vol. *20*, no. *2*, pp. 151-162, 2017.

[5] K. N. Lemon and P. C. Verhoef, “Understanding customer experience throughout the customer journey,” *Journal of Marketing*, vol. *80*, no. *6*, pp. 69-96, 2016.doi: https://doi.org/10.1509/jm.15.0420

[6] H. Li and P. K. Kannan, “Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment,” *Journal of Marketing Research*, vol. *51*, no. *1*, pp. 40-56, 2014. doi: https://doi.org/10.1509/jmr.13.0050

[7] S. Neslin and V. Shankar, “Key issues in multichannel customer management: Current knowledge and future directions,”*Journal of Interactive Marketing*, vol. *23*, no. *1*, pp. 70-81, 2009.doi: https://doi.org/10.1016/j.intmar.2008.10.005

[8] B. Scheibehenne, R. Greifeneder, and P. M. Todd, “Can there ever be too many options?A meta-analytic review of choice overload,” *Journal of Consumer Research*, vol. *37,* no. *3*, pp. 409-425, 2010.[Online]. Available: http://scheibehenne.com/ScheibehenneGreifenederTodd2010.pdf

[9] V. Shankar and E. C. Malthouse, “The growth of interactions and dialogs in interactive marketing,”*Journal of Interactive Marketing*, vol. *21*, no. *2*, pp. 2-4, 2007.Doi: https://doi.org/10.1002/dir.20080

[10] S. A. Neslin, D. Grewal, R. Leghorn,V. Shankar, M. L. Teerling, J. S. Thomas, and P. C. Verhoef, “Challenges and opportunities in multichannel customer management,” *Journal of Service Research,* vol. *9*, no. *2*, pp. 95-112, 2006.doi: https://doi.org/10.1177/1094670506293559

[11] D.Bowman and D.Narayandas, “Managing customer-initiated contacts with manufacturers: The impact on share of category requirements and word-of-mouth behaviour,” *Journal of Marketing Research,* vol. *38*, no. *3*, pp. 281-297, 2001.

[12] E. Andler, I. Becker, F. V. Wangenheim, and J. Schumann, “Mapping the customer journey: Lessons learned from graph-based online attribution modelling,” *International Journal of Research in Marketing*, vol. *33*, no. *3*, pp. 457-474, 2016.

[13] P. Berger, R. Bolton, D. Bowman, E. Briggs, V. Kumar, A. Parasuraman, and C. Terry,“Marketing actions and the value of customer assets: A framework for customer asset management,”*Journal of Service Research*, vol. *5*, no. *1*, pp. 39-54, 2002.

[14] T. Wiesel, K. Pauwels, and J. Arts, “Marketing's

profit impact: Quantifying online and off-line funnel progression," *Marketing Science,* vol. *30*, no. *4*, pp. 604-611, 2011. Retrieved from http://www.jstor.org/stable/23012014

About the Authors

Dr. Subhabaha Pal is currently attached to Manipal Academy of Higher Education Data Science programs as Senior Faculty. He had been named among the top 20 Data Science and Machine Learning Academicians in India by Analytics India Magazine in 2018 and was also honoured as 'Best Supportive Mentor' by Data Science Society, Bulgaria in 2017. He has authored 4 books on Statistics that have been published by reputed publishers like New Age International (Previously, Wiley Eastern Private Limited), CRC Press (Taylor & Francis) and Manipal Press. He has authored 36 research papers in reputed national and international journals. He has done his Ph.D. on non-life insurance analytics from the University of Calcutta, India.

Sampa Pal is working as a Consultant with Arminus Software Pvt.Ltd.. She has worked in the IT and KPO sector for more than three years. She has done M.Sc. in Electronics from the Vidyasagar University, West Bengal, India. She is a certified Big Data Professional from AnalytixLab, Bangalore, India.