

INSY 6500
Information Systems for Operations

Data Set 1 - Online Shoppers

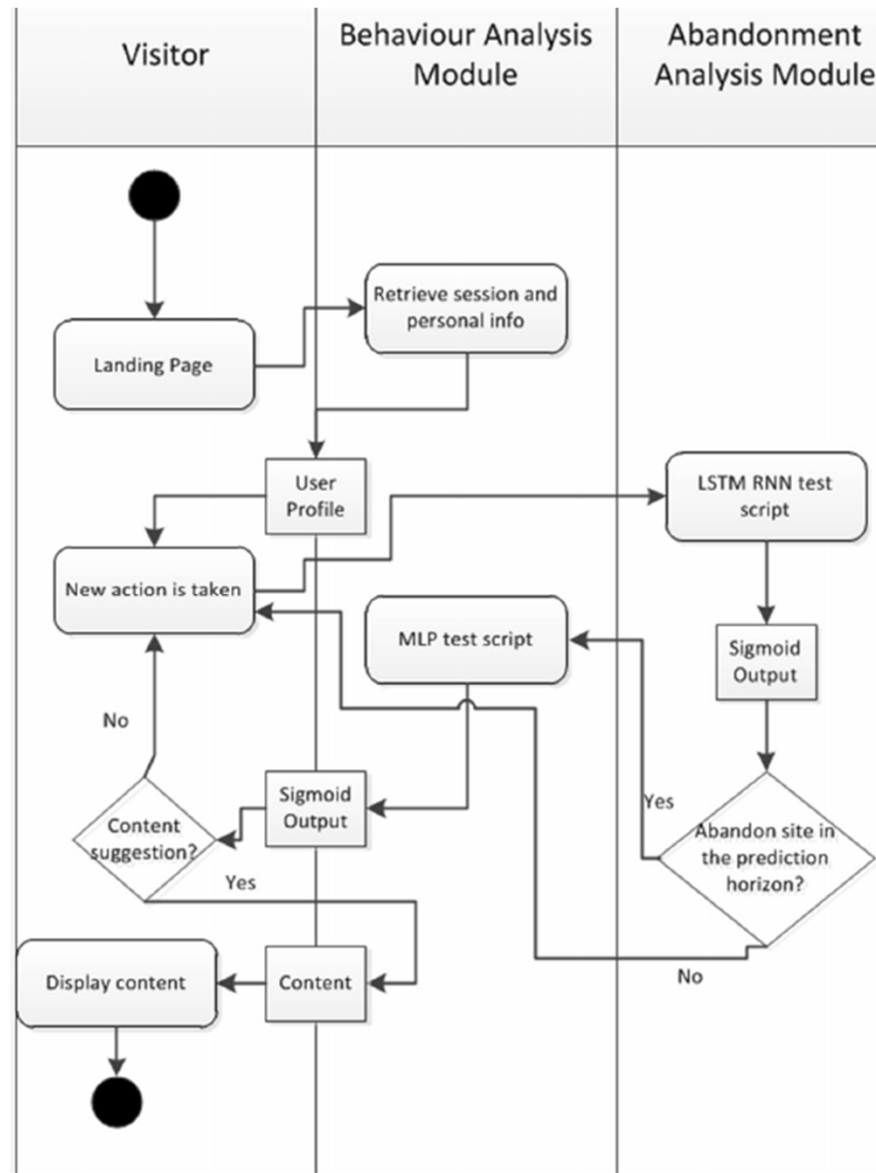
Fall 2019

Jeffrey S. Smith

Industrial and Systems Engineering Department
Auburn University
jsmith@auburn.edu

References

- Sakar, C.O., Polat, S.O., Katircioglu, M. et al., “Real-time Prediction of Online Shoppers’ Purchasing Intention Using Multilayer Perceptron and LSTM Recurrent Neural Networks,” *Neural Computing and Applications* (2018), DOI: <https://doi.org/10.1007/s00521-018-3523-0>. (The PDF of the paper is on Canvas)
- Dataset:
<https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset>
- The dataset and associated Jupyter notebooks are on our GitHub site
 - This is the dataset that you used for HW2.



*Sakar, C.O., Polat, S.O.,
Katircioglu, M. et al., 2018*

Table 1 Numerical features used in the user behavior analysis model

Feature name	Feature description	Min. value	Max. value	SD
Administrative	Number of pages visited by the visitor about account management	0	27	3.32
Administrative duration	Total amount of time (in seconds) spent by the visitor on account management related pages	0	3398	176.70
Informational	Number of pages visited by the visitor about Web site, communication and address information of the shopping site	0	24	1.26
Informational duration	Total amount of time (in seconds) spent by the visitor on informational pages	0	2549	140.64
Product related	Number of pages visited by visitor about product related pages	0	705	44.45
Product related duration	Total amount of time (in seconds) spent by the visitor on product related pages	0	63,973	1912.25
Bounce rate	Average bounce rate value of the pages visited by the visitor	0	0.2	0.04
Exit rate	Average exit rate value of the pages visited by the visitor	0	0.2	0.05
Page value	Average page value of the pages visited by the visitor	0	361	18.55
Special day	Closeness of the site visiting time to a special day	0	1.0	0.19

Table 2 Categorical features used in the user behavior analysis model

Feature name	Feature description	Number of categorical values
OperatingSystems	Operating system of the visitor	8
Browser	Browser of the visitor	13
Region	Geographic region from which the session has been started by the visitor	9
TrafficType	Traffic source by which the visitor has arrived at the Web site (e.g., banner, SMS, direct)	20
VisitorType	Visitor type as “New Visitor,” “Returning Visitor,” and “Other”	3
Weekend	Boolean value indicating whether the date of the visit is weekend	2
Month	Month value of the visit date	12
Revenue	Class label indicating whether the visit has been finalized with a transaction	2

*Sakar, C.O., Polat, S.O.,
Katircioglu, M. et al., 2018*