




[Center for Machine Learning and Intelligent Systems](#)

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

☒ Repository ☐ Web 

[View ALL Data Sets](#)

# Online Shoppers Purchasing Intention Dataset Data Set

*Download:* [Data Folder](#), [Data Set Description](#)

**Abstract:** Of the 12,330 sessions in the dataset, 84.5% (10,422) were negative class samples that did not end with shopping, and the rest (1908) were positive class samples ending with shopping.

<b>Data Set Characteristics:</b>	Multivariate	<b>Number of Instances:</b>	12330	<b>Area:</b>	Business
<b>Attribute Characteristics:</b>	Integer, Real	<b>Number of Attributes:</b>	18	<b>Date Donated</b>	2018-08-31
<b>Associated Tasks:</b>	Classification, Clustering	<b>Missing Values?</b>	N/A	<b>Number of Web Hits:</b>	26192

## Source:

Source

1. C. Okan Sakar  
Department of Computer Engineering, Faculty of  
Engineering and Natural Sciences, Bahcesehir University,  
34349 Besiktas, Istanbul, Turkey

2. Yomi Kastro  
Inveon Information Technologies Consultancy and Trade,  
34335 Istanbul, Turkey

## Data Set Information:

The dataset consists of feature vectors belonging to 12,330 sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

## Attribute Information:

The dataset consists of 10 numerical and 8 categorical attributes. The 'Revenue' attribute can be used as the class label.

"Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration" represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories. The values of these features are derived from the URL information of the pages visited by the user and updated in real time when a user takes an action, e.g. moving from one page to another. The

"Bounce Rate", "Exit Rate" and "Page Value" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site. The value of "Bounce Rate" feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session. The value of "Exit Rate" feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session. The "Page Value" feature represents the average value for a web page that a user visited before completing an e-commerce transaction. The "Special Day" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction. The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentine's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8. The dataset also includes operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.

## Relevant Papers:

Sakar, C.O., Polat, S.O., Katircioglu, M. et al. Neural Comput & Applic (2018). [[Web Link](#)]

## Citation Request:

If you use this dataset, please cite:

Sakar, C.O., Polat, S.O., Katircioglu, M. et al. Neural Comput & Applic (2018). [[Web Link](#)]

---

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)