

## CASE STUDY - ECOMMERCE WEB ANALYTICS



Website: [www.analytixlabs.co.in](http://www.analytixlabs.co.in)

Email: [info@analytixlabs.co.in](mailto:info@analytixlabs.co.in)

**Disclaimer:** This material is protected under copyright act AnalytixLabs©, 2011-2018. Unauthorized use and/ or duplication of this material or any part of this material including data, in any form without explicit and written permission from AnalytixLabs is strictly prohibited. Any violation of this copyright will attract legal actions.

**BUSINESS PROBLEM:** Client wants to predict successful all “unique Ids” from unique Id column who have high chance of getting 1 as “Target” column value.

**Input Two Datasets:**

- a) Training – train.csv
- b) Testing – test.csv

**Train/Test Data Schema (Data Description):**

- a) 'Target' Column in the training set is Class Label. “-1” signifies missing value in the columns.
- b) TOP: Time on Page (in seconds)
- c) Exits: Number of times unique Id has exited the page
- d) Binary\_Var: contains binary values
- e) Metric\_Var: contains continuous values
- f) Unique Id: The primary key/unique identifier
- g) Page1\_Visited: Whether unique id has visited the page or not

**Submissions:**

1. Detailed code including Complete Preprocessing & Feature Engineering Steps
2. Output file in .csv format which consists of 2 columns “Unique Id” & “Predict\_Target”
3. You need to share the following evaluation metrics
  - a) Confusion Matrix
  - b) F1 Score
  - c) Accuracy Score
  - d) AUC\_ROC\_Score
4. The code must be clearly commented. Visualizations are optional

*PS: The column names have been changed & some column values have been hashed for data confidentiality reasons*