

ITM885 Case Study 1: Consumer Conversion

We have data "conversion.csv" about users who hit a website: whether they converted or not as well as some of their characteristics such as their country, the marketing channel, their age, whether they are repeat users and the number of pages visited during that session (as a proxy for site activity/time spent on site). We want to do some data mining, predict the conversion of users by a binary classification model and interpret it. Below is the description of the data columns.

Predictors

country : user country based on the IP address

age : user age. Self-reported at sign-in step

new_user : whether the user created the account during this session or had already an account and simply came back to the site

source : marketing channel source (categorical variable)

 Ads: came to the site by clicking on an advertisement

 Seo: came to the site by clicking on search results

 Direct: came to the site by directly typing the URL on the browser

total_pages_visited: number of total pages visited during the session. This is a proxy for time spent on site and engagement during the session.

Label:

converted: 1 means they converted within the session, 0 means they left without buying anything.