

Machine Learning Assignment

Now Interact works with online data to understand visitors' intentions and optimize their purchase journey and maximize sales. In order to personalize and optimize the contact options on a website, Now Interact predicts many different parameters for each visitor. These predictions help us answer *who* to offer help to, *what* to offer, *where* on the site to offer it and *when* to make the offer. One parameter that is relevant for several of these questions is "Is this person going to accept an assistance offer at this time?"

Task

Together with this document, you have received a dataset containing actual online behaviors from a website with a number of different contact options or offers. Use your weapon of choice to create a model that predicts whether the displayed content resulted in a contact or not (i.e. predict the value in the Contact column).

This is a task that helps us get insight into how you would handle a typical assignment. We expect that this task will take 2-3 hours to complete. This is of course not at all enough time to build a very good predictor and the intention is not for you to build the ultimate algorithm, but to show how you approach a problem. The actual performance of your model is not part of the evaluation and for this reason we do not supply a test dataset.

Once you have finished building your model, please answer the following questions and send your answers to micael.ehn@nowinteract.com. You do not need to send us your code.

1. Briefly describe the steps you went through in your process to reach your final model
(For example: How did you first approach the problem? Did you run into any particular difficulties?)
2. Describe your final model
(For example: Which algorithm(s) did you use, features, interesting insights?)
3. How did you evaluate the results of your model? What was the resulting score?
(Any particular reason you went with this approach?)
4. What improvements would you make to your model if you had more time?
(For example: Did you have ideas left that you didn't explore? Did you identify any weaknesses that could be mitigated with more time?)
5. Do you think that your model would have a relevant business impact? Why or why not?
What could you do to further evaluate this?
(We know that this is a simple test that you have not spent significant amount of time on, expecting a business impact is not required. We are merely interested in your reasoning)
6. How much time did you spend working on this? Do you have any other comments?

Data Description

The data consists of “contact offers” on a website. An offer may be something as simple as a visible phone number on the page, a chat button, or it may be a popup that asks for the visitor’s phone number with a promise to call back. Each line in the dataset represents a contact offer and a user can receive several contact offers during a visit to the website.

This is a real world dataset and is therefore a little bit messy, please treat it as such and try to work around data quality issues you may run into. Missing or not applicable values are coded as NA.

Variable	Description
UID	Unique ID for the user
LandingPage	The first page viewed by the user in this session (anonymized URL)
TrafficSource	Where the visitor came from when this session started. Can be one of the following: Search: This visitor used a search engine and clicked on a normal link SEM: This visitor used a search engine and clicked on an advertisement Affiliate: This visitor clicked an advertisement on a site which is not a search engine
FirstSessionStartTime	The first time that this visitor was observed on the website
TotalVisits	The total number of times the visitor has visited the website
PreviousSessionStartTime	The last time the visitor visited the website. Will be NA if there was no previous visit (when TotalVisits = 0)
CurrentSessionStartTime	The time that this visit started
DisplayTime	The time at which the contact offer that this row represents was made
ClicksBeforeThisDisplay	The number of clicks (pageviews) that this visitor made during this visit prior to receiving the contact offer
PreviousDisplaysThisSession	The number of previous contact offers during this session
PreviousDisplayTime	The timestamp of the previous contact offer. NA if there were no previous contact offers during the session (when PreviousDisplaysThisSession = 0)
CurrentURL	The (anonymized) URL of the page the contact offer was made on
Country	The country that this visitor comes from. The country names have been replaced to anonymize the data
MessageType	An ID that identifies which type of content that was used for the offer. For example a large popup or a small phone number in the footer
ResponsePromise	Some offers have a response time “promise” such as “we will call you within 1 minute” or “we will call you later today”.
DisplayReason	An ID corresponds to the reason that the offer was made. This may be because of user behavior during the session or it may just be that certain pages always have a contact offer.
CallCenterStatus	Is the call center open or closed? This depends mostly on opening hours – some visitors will try to call outside normal opening hours. The call center may also show up as closed if the number of current callers is very high.
DeviceType	Type of device used by the visitor: mobile/tablet/desktop
Contact	This is the column to predict in this task. It indicates whether the visitor chose to accept the contact offer